Host-Exclusive Primer Design Enables Effective Metabarcoding via 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: Prof. Dr. Kristina Haslinger, k.haslinger@rug.nl, telephone +31-(0)631921108, fax +31 (0)50 363 3000

Running title: Host-Free Primer Design for Metabarcoding

Keywords: Amplicon sequencing, host-associated microbiome, Oxford Nanopore Technology, discriminative primer design, metabarcoding; marker genes

The type of article: Method

References: 39

Abstract: 149 words

Main text: 5371 words

5 color figures and 1 table

Supporting Information: one pdf with supporting figures S1-S3, supporting tables S1-5, Note S1 and S2; and one tabulated data file with supporting datasets 1-8.

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 Availability

The mbc-prime tool is available on GitHub (https://github.com/thackl/mbc-prime) and on Zenodo (https://doi.org/10.5281/zenodo.14218237). Sequencing reads are available on the European Nucleotide Archive (ENA) under the accession PRJEB82852. Other research data are available on Zenodo (https://doi.org/10.5281/zenodo.16411378) and in the supplementary files.

Abstract

Metabarcoding is a powerful tool to simultaneously identify multiple taxa within a habitat. However, its application to host-associated microbiomes is challenged by 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 design of primers for long-read metabarcoding of endophytic fungi in *Vinca minor*. *Mbc-prime* streamlines the design of fungus-specific primers, enabling efficient and plant-free amplification of fungal rDNA from mixed DNA samples. Our workflow can be used to study the composition of complex host-associated microbiomes. It should 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.

Introduction

High-throughput sequencing of taxonomic marker genes (metabarcoding) has become a widely adopted approach for studying biodiversity and microbial ecology across diverse ecosystems (Abdelfattah *et al.* 2018). Among the critical steps in any metabarcoding workflow, the choice and design of suitable primers is paramount. An optimal primer set must balance broad taxonomic coverage of target organisms while minimizing amplification of non-target taxa (Tedersoo *et al.* 2022).

This issue of non-target amplification extends beyond microbial profiling and is encountered across a wide range of ecological and evolutionary studies. For instance, ancient DNA analyses are often compromised by high levels of microbial or environmental DNA that obscure the target sequences (Slon *et al.* 2017). Marine eDNA studies employing broad-range primers are susceptible to amplification biases, whereby highly abundant eukaryotic taxa, such as dominant zooplankton orders, are preferentially amplified, potentially limiting the detection sensitivity for less abundant taxa within the community (Needham & Fuhrman 2016; Zhan *et al.* 2014).

This complication is particularly pronounced when studying host-associated target taxa, for instance in dietary metabarcoding, where the DNA of the predator or herbivore can dominate over that of the consumed prey or plant material, leading to biased interpretations (De Barba *et al.* 2014); or in parasitological diagnostics where scientists struggle aiming to detect in their hosts frequently struggle with the low abundance of target DNA of helminths or protozoan parasites amidst the overwhelming host background (Avramenko *et al.* 2015). In bacterial profiling of the human microbiome, a common challenge arises from the co-amplification of host organellar DNA that overwhelms the microbial signals, especially in low-biomass niches such as skin, blood, or mucosal surfaces (Karstens *et al.* 2019; Marotz *et al.* 2018). In plant-associated microbiomes, chloroplast and mitochondrial sequences often overshadow bacterial 16S rRNA gene amplification, obscuring true microbial diversity (Beckers *et al.* 2016; Lundberg *et al.* 2013). Fungal profiling, using eukaryotic marker genes such as ITS, 18S, or 28S rRNA, introduces additional complexity. These markers often carry regions that are conserved with host ribosomal DNA, making it difficult to avoid co-amplification, particularly in plant tissues such as roots

73 or leaves (Gdanetz *et al.* 2017; Hadziavdic *et al.* 2014). This can mask the presence of less abundant
74 fungal taxa and complicate downstream ecological interpretation. These examples collectively highlight
75 the widespread impact of non-target DNA interference in amplicon sequencing, emphasizing the need
76 for robust primer design strategies that improve target specificity and minimize the amplification of host
77 DNA to ensure accurate and efficient microbial community profiling.

78 To overcome these issues, a wide array of bioinformatics tools has been developed to assist in primer
79 design. General-purpose tools like Primer3 (Rozen & Skaletsky 2000), Primer-BLAST (Ye *et al.* 2012),
80 and PrimerQuest™ offer design capabilities based on melting temperature, GC content, and structural
81 properties. However, they are limited in their ability to assess primer suitability for complex,
82 taxonomically diverse samples. Taxon-aware tools such as PrimerProspector (Walters *et al.* 2011) and
83 DegePrime (Hugerth *et al.* 2014) allow for broader coverage assessments and degeneracy control,
84 making them more suitable for environmental applications. Still, these tools generally lack features
85 required for discriminative primer design, such as the ability to target specific taxonomic groups while
86 excluding the host. Other tools, such as BarCrawl (Frank 2009), support barcode design but not primer
87 development; Primer Validator provides taxonomic validation without de novo primer design or
88 advanced scoring systems; and tools like RDP's Probe Match (Cole *et al.* 2005) or Greengenes' probe
89 search (DeSantis *et al.* 2006) are limited to database-matching probes without flexible input or
90 customization. Moreover, small-scale tools like Primrose (Ashelford *et al.* 2002) is not scalable to
91 modern large-scale datasets and lack features like 3′-end scoring, which is important for predicting
92 primer success.

93 Despite this extensive toolkit, most existing programs are not designed to avoid amplification of host
94 DNA, a key requirement for studying host-associated microbiomes. Therefore, there remains a strong
95 need for specialized computational solutions that can design host-exclusive primers, ensuring that
96 sequencing reads derive primarily from microbial taxa rather than host contaminants (Yap *et al.* 2020).
97 Such primers are essential for increasing classification sensitivity, reducing bias, and improving the
98 cost-effectiveness of microbial community profiling in host-rich samples.

99 Here, we present an effective metabarcoding workflow tailored for host-associated microbiomes,
100 beginning with the design of highly discriminative primers using our new computational tool, *mbc-*
101 *prime*. This tool automatically identifies and ranks candidate primers based on microbial specificity and
102 host exclusivity. We demonstrate the utility of this approach by profiling the community of endophytic
103 fungi associated with the evergreen plant *Vinca minor* (lesser periwinkle), using long-read nanopore
104 sequencing. Our method provides a robust solution for the accurate detection of microbial and other
105 organismal diversity in challenging, host-associated environments.

106 Materials and Methods

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

108 For the sampling, 10 - 15 individual plants of *Vinca minor* were collected from Italy (45°26'3'' N,
109 8°49'53''W), immediately placed in an ice box, and transported to the lab. After surface sterilization
110 (He *et al.* 2023), the plant material was quickly frozen with liquid nitrogen, ground into fine powder,
111 and stored at – 70 °C until DNA extraction.

112 Holobiont genomic DNA was extracted using NucleoSpin® Microbial DNA kit (BIOKÉ, Leiden, The
113 Netherlands) in accordance with the protocols provided by the manufacturer with adjustment. Briefly,
114 0.2 g ground plant material was suspended in 100 µL elution buffer BE, 40 µL buffer MG and 10 µL
115 liquid proteinase K. The plant cells were disrupted by vortexing at 1000 rpm for 3 mins. After adding
116 600 µL buffer MG, samples were vortexed for another 1 min at 1000 rpm, followed by 2 min
117 centrifugation at 13,000 x g. The supernatant was transferred onto a NucleoSpin® Microbial column,
118 washed twice with buffer BW and B5 buffer, and eluted with 30 µL elution buffer BE. The DNA was
119 quantified with NanoDrop N-100 (ThermoFisher, Waltham, MA, USA) and stored in – 20 °C until
120 further use.

121 Mock community construction

122 To generate a synthetic mock community for initial testing of the new primers, we used six species
123 isolated from healthy-looking, surface-sterilized leaves of *Vinca minor,* collected in Groningen (The

124 Netherlands) in November 2021. Fungal isolation, cultivation, and genomic DNA isolation were
125 executed as described in our previous work (He *et al.* 2023). The SR1R-Fw/ LR12-R primer pair
126 (D'Andreano *et al.* 2021) was used to amplify the entire fungal rDNA locus. Reactions were run in
127 volumes of 25 µL, containing 1 µL each of the forward (10 µM) and reverse primer (10 µM), 12.5 µL
128 LongAmp Hot Start Taq 2x Master Mix (New England Biolab, MA, USA), 1 µL of DNA template (40
129 fmol), and 9.5 µL of nuclease-free water. PCR amplification was carried out with the following protocol:
130 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
131 for 5.5 min, followed by a final 65°C extension for 10 min. The PCR products were purified using
132 AMPure XP magnetic beads (Beckman Coulter, California, USA) and the concentrations were
133 determined using Qubit 3.0 (Invitrogen, Waltham, MA, USA), and the Qubit dsDNA HS Assay Kit.
134 These six fungal amplicons were mixed in equimolar ratios to construct the mock community DNA
135 sample with a concentration of 55 ng/ µL.

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

137 To analyze endophytic fungi in *Vinca minor*, all related sequences were downloaded from the SILVA
138 database (Supplementary File 2, Datasets S1 and S2). For the SSU region, 46 sequences from the
139 Gentianales order were extracted from the downloaded database file
140 (SILVA_138.1_SSURef_NR99_tax_silva.fasta.gz.). In the same way, a dataset of 118 fungal sequences
141 was collected, which represents all major fungal divisions (ascomycota, basidiomycota, chytridiomycota,
142 cryptomycota, zoopagomycota, microsporidia, mucoromycota, neocallimastigomycota and
143 blastocladiomycota). These 164 sequences were aligned using the MAFFT multiple sequence alignment
144 software (ouput format: fasta format, input order; strategy: L-INS-i; arguments: --ep)(Katoh & Standley
145 2013). For the LSU region, 62 fungal sequences (Supplementary File 2, Dataset S3) and 6 available
146 Gentianales sequences (Supplementary File 2, Dataset S4) were extracted from the downloaded database
147 file (SILVA_138.1_LSURef_NR99_tax_silva.fasta.gz). Alignment was performed among these 68
148 sequences in the same way as for the SSU.

149 Both alignments were analyzed with *mbc-prime* version 0.6.0 (-s = 0.5) (see Results section for
150 implementation). We obtained 147 candidate discriminative primers across the SSU (Supplementary
151 File 2, Dataset S5) and 371 across the LSU (Supplementary File 2, Dataset S6). Four candidates for
152 forward primers on the SSU and two reverse primers on the LSU were selected from the initial set based
153 on the following criteria: 1) at least 80% coverage of fungi at zero or one mismatch (*in silico*); 2) at least
154 one mismatch at the 3' end between primer compared to the plant sequences; 3) a GC content between
155 40 and 60%; 4) approx. 20 base pairs in length. In addition, we manually optimized the length of
156 candidate primer sequences to lower the difference in melting temperatures and the likelihood of
157 heterodimer formation in primer combinations as well as incorporated degenerate positions to maximize
158 the coverage of the target group. Finally, all selected primers were evaluated *in silico* using SILVA
159 TestProbe 3.0 (https://www.arb-silva.de/search/testprobe/) for the individual primers and
160 OligoAnalyzer™ (Version 3.1, Integrated DNA Technologies,
161 https://eu.idtdna.com/pages/tools/oligoanalyzer) for primer pairs before being tested on the DNA
162 samples of a fungal mock community and plant material. A step-by-step tutorial for using *mbc-prime* is
163 listed in the Supplementary Note S1.

164 Amplicon generation for Oxford Nanopore sequencing

165 The strategy employed in this study involves two steps of PCR. The first PCR was performed with the
166 gene-specific rDNA primers including overhangs that serve as handles for PCR barcoding in the second
167 step (5' TTTCTGTTGGTGCTGATATTGC-[forward primer sequence] 3'; 5'
168 ACTTGCCTGTCGCTCTATCTTC-[reverse primer sequence] 3'). All primer pairs used in this study
169 are shown in Table 2 and sequences of all primers used in this study are listed in Supplementary File 2,
170 Dataset S7. Each PCR reaction mix consisted of 0.4 µM each of the forward and reverse primer, 1 x
171 LongAmp Hot Start Taq 2x Master Mix (New England Biolab, MA, USA), and 40 ng of template DNA,
172 combined in a 25 µL reaction. The PCR cycling conditions were as follows: initial denaturation of 94°C
173 for 30 s, followed by 30 cycles of a 10 s hold at 94°C, a 30 s hold at the respective annealing temperature
174 shown in Table 1, and an elongation period specific for the length of the amplicon as shown in table 1

175 at 65°C, followed by a final 65°C extension for 10 min. The PCR amplicons were purified using AMPure
176 XP magnetic beads (Beckman Coulter, California, USA) and quantified by Qubit. The second PCR was
177 used to incorporate the Oxford Nanopore barcode sequences with the PCR Barcoding Expansion 1-12
178 (EXP-PBC001, ONT, Oxford, UK). Each amplicon PCR product was used as a template (100-200 fmol)
179 in the second PCR reaction with 1 µL PCR barcoding primer mix (one of BC1- BC12, at 10 µM), and
180 25 µL LongAmp Hot Start Taq 2X Master Mix (New England Biolab, MA, USA) in a total of 50 µL.
181 The PCR cycling conditions were as follows: initial denaturation of 95°C for 3 min, followed by 15
182 cycles of 95°C for 15 s, 62°C for 15 s, and an elongation period specific for the length of the amplicon
183 as shown in table 1 at 65°C, followed by a final 65°C extension for 10 min. The barcoded PCR products
184 were further purified with 0.4x volume of AMPure XP Beads (Beckman Coulter, California, USA). All
185 barcoded PCR products were pooled in equal amounts to prepare 1 µg of DNA in 50 µL nuclease-free
186 water for library preparation. The metadata of samples as input for the library preparation is listed in the
187 Supplementary File 2, Dataset S8.

188 Library preparation

189 The library preparation was performed with the ligation sequencing kit V14 (ONT, SQK-LSK114)
190 following the manufacturer's instructions
191 (https://community.nanoporetech.com/docs/prepare/library_prep_protocols/ligation-sequencing-gdna-
192 pcr-barcoding-sqk-lsk114-with-exp/v/pbc_9182_v114_revh_07mar2023). First, the pooled barcoded
193 amplicons underwent end repair with the NEBNext® Ultra II End Repair / dA-tailing Module (NEB,
194 E7546). This reaction consisted of 50 µL DNA (1 µg), 7 µL Ultra II End-prep Reaction Buffer and 3
195 µL Ultra II End-prep Enzyme Mix, and was incubated at 20°C for 5 minutes and 65°C for 5 minutes.
196 The end-repaired amplicons were further purified using AMPure XP magnetic beads and quantified by
197 Qubit. For the adapter ligation step, 60 µL of the end-repaired and purified amplicons, 25 µL Ligation
198 Buffer (LNB), 10 µL NEBNext Quick T4 DNA Ligase, and 5 µL Ligation Adapter (LA) were combined
199 in a reaction tube and incubated for 10 min at room temperature. After ligation, the samples were mixed
200 with 0.4 x volume of AMPure XP magnetic beads and then washed twice with 250 µL Long Fragment

201 Buffer provided in the ligation sequencing kit. The library was finally eluted with 15 µL Elution Buffer
202 (EB), incubated at 37 °C for 10 minutes, and quantified by Qubit. 35-50 fmol of this final prepared
203 library were loaded for sequencing.MinION sequencing data analysis

204 Sequencing was performed on a MinION device with a R10.4.1 flow cell running with MinKNOW
205 software v24.02.10 . Basecalling and demultiplexing were performed with Dorado v7.3.9 in super
206 accurate mode. We trimmed reads using NanoFilt v2.8.0 and assessed the read quality using NanoPlot
207 v1.42.0. Subsequently, the FASTQ files were uploaded to EPI2ME Labs v4.1.4, and the "wf-
208 metagenomics" workflow was performed with the following settings: classifier: Kraken2; Kraken2
209 memory mapping: true; dataset: ncbi_16s_18s_28s_ITS or SILVA_138_1; N taxa barplot: 9; bracken
210 length: 2000; min percent identity: 90; min len: 2000; max len: 6000. The mock community samples
211 were analyzed with a custom database built with Kraken2. Relative abundances were further visualized
212 with GraphPad Prism 10.

213 Results

214 Rationale and workflow for effective host-exclusive primer design

215 Metabarcoding relies on the targeted PCR amplification of suitable genomic markers that comprise
216 conserved regions serving as the primer binding sites and variable regions with enough differences to
217 distinguish individual species. Herein, it is crucial that the primers amplify DNA equally well across a
218 broad taxonomic range. Different marker genes are used for different biological groups: for animals, the
219 mitochondrial gene cytochrome c oxidase I (COI) is widely applied (Hebert *et al.* 2003); in plants,
220 chloroplast genes such as *matK* and *rbcL* are commonly used (Hollingsworth *et al.* 2009); for bacteria
221 and archaea, the 16S rRNA gene is the standard marker, while fungi and protists are typically identified
222 using the internal transcribed spacer (ITS) regions, along with the 18S and 28S rRNA genes (Klindworth
223 *et al.* 2013).

224 Expanding on this principle, the rationale for creating host-exclusive, discriminative primers is simple:
225 identify a variable-region-enclosing primer pair that is conserved among all target group members but

226 differs in sequence in all unwanted taxa. Such locations can be discerned in one or more combined
227 multiple-sequence alignments (MSAs) of representative sequences of the target and exclusion group.
228 This rationale allowed us to devise an effective workflow for host-exclusive primer design for holobiont
229 metabarcoding (Figure 1). Since this workflow requires identifying discriminative primer sites in MSAs,
230 which is slow and tedious if done by hand, we also devised the program *mbc-prime* (described in detail
231 in the next section) to help automate this crucial step. The list of discriminative, candidate primers
232 returned by *mbc-prime* can then be further inspected for desired properties of the primer, for instance
233 the GC content and melting temperature of the primer and the position of the non-target mismatches
234 within the primer sequence. With the most discriminative primers in hand, additional optimization of
235 the primers can be performed by introducing degenerate bases into the sequence to optimize the
236 coverage among the target sequences. Lastly, the candidate primers can be evaluated for their specificity
237 for target group and efficiency of excluding non-target group.

238 Detection of discriminative primer binding sites with *mbc-prime*

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

242 *mbc-prime* identifies discriminative primers based on a user-supplied MSA in FASTA format,
243 comprising DNA sequences of the chosen genomic locus in the target and exclusion sets. In a first step,
244 the MSA is imported and preprocessed. Terminal gaps, introduced by partial sequences frequently
245 present in amplicon marker databases, are masked to distinguish them from genuine gaps. Columns with
246 more than 20% (default value) gaps or ambiguous bases (non-ATGCU) are trimmed from the alignment.
247 The trimmed MSA is then analyzed via a sliding window with a size corresponding to the intended
248 primer length (default: 20 bp). For each window and sequence set (target and exclusion), the abundance
249 of each sequence variant is counted, and the most abundant sequence variant of the target set is chosen
250 as primer candidate. Next, for each remaining sequence variant from both sets, the number of
251 mismatches compared to the primer candidate is computed, and the overall percentage of sequence

252 variants with 0, 1, 2, 3 or 4-or-more mismatches are determined. The underlying edit distances are
253 computed via Python bindings of the Edlib library (Šošić & Šikić 2017). This results in two mismatch
254 profiles per primer candidate, one for the target set and one for the exclusion set. An example of
255 promising profiles for a discriminative primer would be the following: 90% and 10% of the target
256 sequences align with 0 and 1 mismatch, respectively, while 10%, 20% and 70% of the exclusion
257 sequences align with 2, 3, and 4-or-more mismatches, respectively.

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

263
$$\text{score} = \frac{1}{mm} \sum_{k=1}^{mm} \left(\sum_{i=1}^{k} \text{target}[i] - \sum_{i=1}^{k} \text{exclude}[i] \right)$$

264 Finally, candidate primers in the MSA are filtered for a maximum number of gaps, ambiguous bases
265 and a minimum score (default 0.5), and overlapping primers are grouped together. The filtered list of
266 above-threshold primer candidates is reported together with the positional information, score, mismatch
267 profiles, and sequence in a tab-separated table (Figure 2).

268 Case study

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

270 To demonstrate the applicability of our metabarcoding primer design workflow, we targeted the
271 endophytic mycobiome of *Vinca minor* (lesser periwinkle) with long-read sequencing of the rDNA locus.
272 *V. minor* belongs to the order of Gentianales in the Apocynaceae family and is a popular plant for its
273 ornamental and medicinal properties. To date, the microbiome of *V. minor* remains largely unexplored.
274 In our initial amplicon sequencing experiments with published primers targeting the rDNA of fungi for
275 long-read sequencing, we observed that 99 % of reads were derived from the *V. minor* rDNA locus (data

276 not shown). Therefore, we considered this the ideal test case for the performance of our holobiont
277 metabarcoding workflow.

278 Since we wanted to target the entire rDNA locus, we downloaded fungal (target) and plant (exclusion)
279 sequences spanning the SSU and LSU regions from the SILVA database. We selected 118 fungal SSU
280 sequences from different phyla (Ascomycota, Basidiomycota, Chytridiomycota, Cryptomycota,
281 Zoopagomycota, Microsporidia, Mucoromycota, Neocallimastigomycota, and Blastocladiomycota) to
282 represent all major fungal divisions (Supplementary File 2, Dataset S1), and 46 sequences from the order
283 Gentianales (Supplementary File 2, Dataset S2). For the LSU, we included 62 fungal sequences
284 (Supplementary File 2, Dataset S3) and 6 sequences from Gentianales (Supplementary File 2, Dataset
285 S4). We created two MSAs for all SSU and LSU sequences, respectively, and scanned them for
286 discriminative primer binding sites. Using *mbc-prime* with default settings, we obtained 147 candidate
287 primers across the SSU (Supplementary File 2, Dataset S5) and 371 across the LSU (Supplementary
288 File 2, Dataset S6). From the list, we selected primers that had scores of 0.6 and higher and carried at
289 least one mismatch at the 3' end when compared to plant (exclusion) sequences. With these criteria we
290 obtained four forward primers in different regions of the SSU and two reverse primers in different
291 regions of the LSU (Figure 3A). Following careful examination of the GC content, the predicted melting
292 temperatures, and the respective primer binding sites within the MSA, we manually optimized the primer
293 sequences to establish the final sequences (Figure 3B, Supplementary File 2, Dataset S7). Primer mbc-
294 1220-F was extended by two nucleobases at the '5 end, mbc-2148-R was shortened by one base at the
295 '5 end, and in mbc-2479-F and mbc-2148-R we incorporated degenerate positions to maximize the
296 coverage of the target group (Fig. S1).

297 To assess the performance of the final primers *in silico*, we used the SILVA TestProbe 3.0 tool with
298 default settings and "fungi" and "Gentianales" or "Chloroplastida" as test sets. For completeness, we
299 also included several primers from literature (Nilsson *et al.* 2019; Tedersoo *et al.* 2008, 2015;
300 Wurzbacher *et al.* 2019) Similar to the literature primers, all of the new primers show a wide coverage
301 across fungi. In terms of coverage across the Chloroplastida, our newly designed primers, similar to the

302 literature primers LR5-Fung and RCA95m, exhibited low coverage within this group (Table 1). We
303 included LR5-Fung, which is also identified by our *mbc-prim*e tool, for further experimental analysis.

304 With these primer candidates in hand, we performed an *in silico* analysis of all possible primer
305 combinations to identify those with favorable parameters, such as a low difference in melting
306 temperatures and a low likelihood of heterodimer formation. The 10 primer pairs that met our quality
307 standards (pairs 1-10; Table S1, Table S2) and one primer pair from literature ("3.5 kb") (D'Andreano
308 *et al.* 2021) were carried forward for experimental validation with a fungal mock community and
309 genomic DNA extracted from the surface-sterilized *V. minor* holobiont.

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

311 We generated a standardized DNA sample by mixing equimolar amounts of PCR-amplified rDNA from
312 six fungal isolates of our in-house strain collection (Table S3). The strains were chosen to cover a wide
313 taxonomic range and different levels of GC content in order to assemble an authentic mock community
314 for our primer testing. Next, we optimized the annealing temperatures of the PCR for each primer pair
315 to generate high-quality amplicons based on visual inspection of the bands after agarose
316 gelelectrophoresis (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 S4).

323 To classify the trimmed and filtered reads, we applied the *wf-metagenomics* workflow with Kraken2.
324 Since not all target regions of the mock community members were available in public databases, we
325 generated a custom database comprising the whole genome sequences of the six isolates with Kraken2
326 (Supplementary Note S2). Using this custom database, 99.80% of reads across all samples were
327 classified, indicating a high sample quality with minimal contamination or sample preparation artefacts,

328 and a high sequencing accuracy. With regards to the relative abundances of species within the mock
329 community, all samples showed a similar composition although among the amplicons generated with
330 primer pair 8, the basidiomycete *Trametes versicolor* seemed to be underrepresented. This indicates that
331 all primer pairs except for primer pair 8, successfully amplify rDNA regions across diverse fungal
332 isolates, producing amplicons that closely match the expected composition of the mock community
333 (Figure 4B).

334 Experimental evaluation of the designed fungus-specific primers using genomic holobiont 335 DNA of *V. minor*

336 Having confirmed the amplification efficacy of the new primer pairs on the mock community sample,
337 we next sought to assess their performance in a real-world holobiont scenario. We used the primer pairs
338 to specifically target the endophytic fungi in the holobiont DNA from *V. minor*. However, while
339 preparing for the experiment and performing the first amplification step with the holobiont DNA (Figure
340 S3), we noticed that primer pair 8 did not give a high-quality PCR product with a distinct band of the
341 correct size. The individual primers performed well in the other primer combinations thus we conclude
342 that the issue specifically arises from the combination of these primers in the context of the holobiont
343 sample, where the target fungal DNA is rare. Since we could not overcome this problem despite various
344 optimization attempts (annealing temperature, additives, polymerase), we abandoned primer pair 8.

345 Proceeding with the remaining three discriminative primer pairs 7, 9, and 10, and the non-discriminative
346 literature-based primer pair 3.5kb, we generated, a total of 599,335.274 reads with a Qscore greater than
347 18 in a single multiplex sequencing run (Table S5). When using the same classification approach as
348 before (*wf-metagenomics* workflow, Kraken2), this time with a custom database comprising the NCBI
349 RefSeq Targeted Loci database (ncbi_16s_18s_28s_ITS) and the rDNA locus of a closely-related Vinca
350 species (NCBI taxon ID 141628), over 92% of all reads were successfully classified (Figure 5). As
351 expected, a large proportion of the reads generated with the 3.5 kB primer set is classified as host (*Vinca*
352 sp.), while the new primer pairs did not generate such amplicons. In terms of species composition, the
353 samples generated with the new primer pairs show similar abundances of the most abundant taxa with

354 only slight variations. This demonstrates that our newly-designed primer pairs are indeed able to
355 efficiently discriminate between the rDNA locus of fungi and the host plant and that they provide a
356 similar reflection of the community composition.

357 For completeness, we also analyzed the data using the SILVA database, which includes a broader range
358 of eukaryotes. For amplicons generated with primer pair 7 97% of the reads were classified as fungi,
359 with primer pair 9 95%, and with primer pair 10 97%. Overall, none of the amplicons generated with
360 the new primer pairs were classified as plant. Interestingly, a subset of amplicons was assigned to protist
361 taxa, with the highest proportion, approximately 4.5%, observed in primer set 9. This indicates that
362 while all primer pairs can successfully exclude the plant rDNA locus, primer pair 9 may be also good at
363 targeting other microeukaryotes, which requires further investigation (Figure S3).

364 Discussion

365 In molecular biology, the presence of non-target DNA (e.g., host DNA) can interfere with the detection
366 and analysis of target DNA (e.g., microbial, pathogen, or environmental DNA). This is particularly
367 problematic in fields like microbiome research, clinical diagnostics, and environmental DNA (eDNA)
368 studies, where target DNA is often present in low abundance relative to other DNA. To address this
369 issue, various blocking methods have been developed to selectively suppress the amplification or
370 sequencing of non-target DNA. These techniques involve the use of chemically modified
371 oligonucleotides, such as C3 spacers, locked nucleic acids (LNAs), or peptide nucleic acids (PNAs),
372 which are designed to bind conserved regions of non-target DNA. The goal is to either direct the binding
373 of blocking oligos to these modified sequences or to physically block polymerase elongation during
374 PCR amplification. However, their effectiveness is often constrained by suboptimal binding affinity of
375 the blocking oligos to non-target templates (Vestheim & Jarman 2008), the requirement for complex
376 experimental optimization (Piñol *et al.* 2015), and limited cost-effectiveness (Azadnia *et al.* 2023).
377 Given these limitations, the design of highly specific, discriminative primers could be a more practical
378 and scalable alternative for minimizing non-target amplification.

379 In this study, we developed fungus-specific primers using a host-exclusive primer design workflow for
380 the generation of long amplicons that almost entirely exclude plant DNA as a cost-effective alternative
381 to the blocking methods. High-throughput sequencing of taxonomic marker genes (metabarcoding) is a
382 widely used and powerful technique for biodiversity and microbial ecology. However, many taxa
383 recovered from natural environments remain unidentifiable at the species level or even higher taxonomic
384 levels when working with short reads (Tedersoo *et al.* 2020). In this case, long-read sequencing has the
385 potential to overcome this problem. For instance, several of the six fungal isolates used for our mock
386 community cannot be unambiguously identified solely based on ITS sequences (amplified with ITS1F
387 and ITS4; data not shown) because they exhibit 100% sequence similarity with multiple taxa. For
388 instance for the *Thyridium oculorum* isolate, ITS-based identification returned matches to both
389 *Phialemonium* sp. and *Thyridium* sp., thus preventing a definitive taxonomic assignment. However,
390 long-read sequencing results from this study showed that 92.15% ~ 99.17% of all reads derived from
391 this isolate were classified as *Thyridium oculorum*, which is consistent with the identification based on
392 multi-locus classification. This highlights that long-read sequencing can provide superior classification
393 results in metabarcoding studies.

394 There are, however, still common pitfalls that need to be tackled in the future. First, reducing bias in the
395 relative abundance of sequence variants should be a primary objective in enhancing amplicon
396 sequencing methods. It is well established that even PCR primers designed for broad coverage tend to
397 amplify some phylogenetic subgroups more effectively than others. For instance in this study, the
398 inferred abundance of *Trametes versicolor* in the mock community was consistently lower than the
399 expected ratio, which might be attributed to amplification efficiencies of the used primer sets. In the
400 experiments with holobiont DNA, we found that the overall community profiles were fairly consistent
401 across primer pairs, though some minor differences were observed in the dominant taxa. This highlights
402 the importance of evaluating primer performance in both controlled and real-world contexts when
403 selecting appropriate primers for a given study.

404 Second, the accuracy of taxonomic classification is highly dependent on the reference database
405 employed. Currently, these databases mainly rely on parts of the rDNA locus. The SILVA database
406 includes near full-length sequences of the small and large subunits of the rRNA operon for both
407 prokaryotes and eukaryotes. However, these sequences are insufficient for amplicons spanning the ITS
408 region and generally allow identification only up to the genus level. Despite this limitation, the SILVA
409 database enables the identification of non-fungal sequences, such as those from Chloroplastida and
410 Metazoa (Quast *et al.* 2012). Conversely, for classification of fungi, the NCBI RefSeq Target Loci
411 database seems to outperform the SILVA database. This superior performance is due to NCBI's focus
412 on the RefSeq database, which includes ribosomal DNA loci (16S, 18S , 28S rDNA genes, and ITS
413 region) and allows for species-level identification, offering better accuracy and resolution in taxonomic
414 classification of microbes. The limitation of this database, however, is that it is focused on archaea,
415 bacteria and fungi, and does not include plants or other micro- and macro-eukaryotes. Therefore, reads
416 originating from such eukaryotes will either not be classified at all or incorrectly classified as fungi.
417 Furthermore, the fungal tree of life is still rather sparsely sampled for molecular barcodes deposited in
418 the RefSeq database. Thus, significant efforts are still needed to improve the database to enhance its
419 completeness and accuracy. This could involve developing databases tailored for eukaryotic full-length
420 rDNA operons (Krabberød *et al.* 2025) or customizing databases to meet specific research needs.

421 Conclusions

422 In the current study, we developed a host-exclusive primer design workflow, which is centered around
423 a new computation tool, *mbc-prime*, to design discriminative metabarcoding primers with the option to
424 exclude the DNA amplification of non-target taxa, for instance the host in host-associated microbiomes.
425 We demonstrated the use of *mbc-prime* when targeting endophytic fungi associated with the medicinal
426 plant *V. minor*. Initially, we evaluated our custom-designed primers *in silico*, demonstrating high
427 coverage of fungal lineages with low coverage to Chloroplastida. Subsequent evaluation using a mock
428 community revealed that the relative abundance of amplicons accurately reflected the community
429 composition of the mock community. Moreover, when applied to plant holobiont samples, our primers

successfully targeted a broad range of endophytic fungi while almost 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 or other biological systems 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 and XL are supported by scholarships 202006550001 and 202106550001 from the China Scholarship Council. Figure 1 was created in BioRender (He, T. (2024) https://BioRender.com/d86p488).

Competing interests

No competing interest is to be declared.

References

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

Ashelford, K.E., Weightman, A.J. & Fry, J.C. (2002). PRIMROSE: a computer program for generating and estimating the phylogenetic range of 16S rRNA oligonucleotide probes and primers in conjunction with the RDP-II database. *Nucleic Acids Res*, 30, 3481–9.

453 Avramenko, R.W., Redman, E.M., Lewis, R., Yazwinski, T.A., Wasmuth, J.D. & Gilleard, J.S.
454 (2015). Exploring the Gastrointestinal "Nemabiome": Deep Amplicon Sequencing to Quantify
455 the Species Composition of Parasitic Nematode Communities. *PLoS One*, 10, e0143559.

456 Azadnia, A., Mikryukov, V., Anslan, S., Hagh-Doust, N., Rahimlou, S., Tamm, H., *et al.* (2023).
457 Structure of plant-associated microeukaryotes in roots and leaves of aquatic and terrestrial
458 plants revealed by blocking peptide-nucleic acid (PNA) amplification. *FEMS Microbiol Ecol*,
459 99.

460 De Barba, M., Miquel, C., Boyer, F., Mercier, C., Rioux, D., Coissac, E., *et al.* (2014). DNA
461 metabarcoding multiplexing and validation of data accuracy for diet assessment: application to
462 omnivorous diet. *Mol Ecol Resour*, 14, 306–323.

463 Beckers, B., Op De Beeck, M., Thijs, S., Truyens, S., Weyens, N., Boerjan, W., *et al.* (2016).
464 Performance of 16s rDNA Primer Pairs in the Study of Rhizosphere and Endosphere Bacterial
465 Microbiomes in Metabarcoding Studies. *Front Microbiol*, 7.

466 Cole, J.R., Chai, B., Farris, R.J., Wang, Q., Kulam, S.A., McGarrell, D.M., *et al.* (2005). The
467 Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis.
468 *Nucleic Acids Res*, 33, D294-6.

469 D'Andreano, S., Cuscó, A. & Francino, O. (2021). Rapid and real-time identification of fungi up to
470 species level with long amplicon nanopore sequencing from clinical samples. *Biol Methods*
471 *Protoc*, 6, 1–6.

472 DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., *et al.* (2006).
473 Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with
474 ARB. *Appl Environ Microbiol*, 72, 5069–5072.

475 Frank, D.N. (2009). BARCRAWL and BARTAB: software tools for the design and implementation
476 of barcoded primers for highly multiplexed DNA sequencing. *BMC Bioinformatics*, 10, 362.

477 Gdanetz, K., Benucci, G.M.N., Vande Pol, N. & Bonito, G. (2017). CONSTAX: a tool for improved
478 taxonomic resolution of environmental fungal ITS sequences. *BMC Bioinformatics*, 18, 538.

479 Hadziavdic, K., Lekang, K., Lanzen, A., Jonassen, I., Thompson, E.M. & Troedsson, C. (2014).
480 Characterization of the 18S rRNA Gene for Designing Universal Eukaryote Specific Primers.
481 *PLoS One*, 9, e87624.

482 Hebert, P.D.N., Cywinska, A., Ball, S.L. & deWaard, J.R. (2003). Biological identifications through
483 DNA barcodes. *Proc R Soc Lond B Biol Sci*, 270, 313–321.

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

487 Hollingsworth, P.M., Forrest, L.L., Spouge, J.L., Hajibabaei, M., Ratnasingham, S., van der Bank,
488 M., *et al.* (2009). A DNA barcode for land plants. *Proceedings of the National Academy of*
489 *Sciences*, 106, 12794–12797.

490 Hugerth, L.W., Wefer, H.A., Lundin, S., Jakobsson, H.E., Lindberg, M., Rodin, S., *et al.* (2014).
491 DegePrime, a Program for Degenerate Primer Design for Broad-Taxonomic-Range PCR in
492 Microbial Ecology Studies. *Appl Environ Microbiol*, 80, 5116–5123.

493 Karstens, L., Asquith, M., Davin, S., Fair, D., Gregory, W.T., Wolfe, A.J., *et al.* (2019). Controlling
494 for Contaminants in Low-Biomass 16S rRNA Gene Sequencing Experiments. *mSystems*, 4.

495 Katoh, K. & Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7:
496 Improvements in performance and usability. *Mol Biol Evol*, 30, 772–780.

497 Klindworth, A., Pruesse, E., Schweer, T., Peplies, J., Quast, C., Horn, M., *et al.* (2013). Evaluation
498 of general 16S ribosomal RNA gene PCR primers for classical and next-generation
499 sequencing-based diversity studies. *Nucleic Acids Res*, 41, e1.

Krabberød, A.K., Stokke, E., Thoen, E., Skrede, I. & Kauserud, H. (2025). The Ribosomal Operon Database: A Full-Length rDNA Operon Database Derived From Genome Assemblies. *Mol Ecol Resour*, 25.

Lundberg, D.S., Yourstone, S., Mieczkowski, P., Jones, C.D. & Dangl, J.L. (2013). Practical innovations for high-throughput amplicon sequencing. *Nat Methods*, 10, 999–1002.

Marotz, C.A., Sanders, J.G., Zuniga, C., Zaramela, L.S., Knight, R. & Zengler, K. (2018). Improving saliva shotgun metagenomics by chemical host DNA depletion. *Microbiome*, 6, 42.

Needham, D.M. & Fuhrman, J.A. (2016). Pronounced daily succession of phytoplankton, archaea and bacteria following a spring bloom. *Nat Microbiol*, 1, 16005.

Nilsson, R.H., Anslan, S., Bahram, M., Wurzbacher, C., Baldrian, P. & Tedersoo, L. (2019). Mycobiome diversity: high-throughput sequencing and identification of fungi. *Nat Rev Microbiol*, 17, 95–109.

Piñol, J., Mir, G., Gomez-Polo, P. & Agustí, N. (2015). Universal and blocking primer mismatches limit the use of high-throughput DNA sequencing for the quantitative metabarcoding of arthropods. *Mol Ecol Resour*, 15, 819–830.

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

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

Slon, V., Hopfe, C., Weiß, C.L., Mafessoni, F., de la Rasilla, M., Lalueza-Fox, C., *et al.* (2017). Neandertal and Denisovan DNA from Pleistocene sediments. *Science (1979)*, 356, 605–608.

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

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

528 Tedersoo, L., Anslan, S., Bahram, M., Põlme, S., Riit, T., Liiv, I., *et al.* (2015). Shotgun
529 metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in
530 metabarcoding analyses of fungi. *MycoKeys*, 10, 1–43.

531 Tedersoo, L., Bahram, M., Zinger, L., Nilsson, R.H., Kennedy, P.G., Yang, T., *et al.* (2022). Best
532 practices in metabarcoding of fungi: From experimental design to results. *Mol Ecol*, 31, 2769–
533 2795.

534 Tedersoo, L., Jairus, T., Horton, B.M., Abarenkov, K., Suvi, T., Saar, I., *et al.* (2008). Strong host
535 preference of ectomycorrhizal fungi in a Tasmanian wet sclerophyll forest as revealed by DNA
536 barcoding and taxon-specific primers. *New Phytologist*, 180, 479–490.

537 Vestheim, H. & Jarman, S.N. (2008). Blocking primers to enhance PCR amplification of rare
538 sequences in mixed samples – a case study on prey DNA in Antarctic krill stomachs. *Front*
539 *Zool*, 5, 12.

540 Walters, W.A., Caporaso, J.G., Lauber, C.L., Berg-Lyons, D., Fierer, N. & Knight, R. (2011).
541 PrimerProspector: *de novo* design and taxonomic analysis of barcoded polymerase chain
542 reaction primers. *Bioinformatics*, 27, 1159–1161.

543 Wurzbacher, C., Larsson, E., Bengtsson-Palme, J., Van den Wyngaert, S., Svantesson, S.,
544 Kristiansson, E., *et al.* (2019). Introducing ribosomal tandem repeat barcoding for fungi. *Mol*
545 *Ecol Resour*, 19, 118–127.

546 Yap, M., Feehily, C., Walsh, C.J., Fenelon, M., Murphy, E.F., McAuliffe, F.M., *et al.* (2020).
547 Evaluation of methods for the reduction of contaminating host reads when performing shotgun
548 metagenomic sequencing of the milk microbiome. *Sci Rep*, 10.

549 Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S. & Madden, T.L. (2012). Primer-
550 BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC*
551 *Bioinformatics*, 13, 134.

552 Zhan, A., Bailey, S.A., Heath, D.D. & Macisaac, H.J. (2014). Performance comparison of genetic
553 markers for high-throughput sequencing-based biodiversity assessment in complex
554 communities. *Mol Ecol Resour*, 14, 1049–1059.

555 # Tables

556 **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]			Chloroplastida			Source
		0 mm	1 mm	2 mm	0 mm	1 mm	2 mm	0 mm	1 mm	2 mm	
Located on the SSU											
SSU-Fngs-F	-	0.2%	94.8%	96.4%	0%	80.0%	88.9%	0.5%	79.2%	86.2%	Tedersoo et al. 2015
mbc-2479-F	0.62	77.1%	87.8%	89.2%	0%	0%	0%	0.3%	0.5%	2.2%	This study
mbc-1840-F	0.62	66.1%	83.5%	90.2%	0%	0%	2.2%	0.2%	1.0%	8.6%	This study
mbc-136-F	0.78	80.1%	89.7%	92.3%	0%	0%	87.5%	25.2%	39.0%	81.3%	This study
mbc-1220-F	0.66	52.1%	82.3%	88.7%	0%	0%	11.1%	0.7%	24.1%	35.8%	This study
NS1short	-	60.05	69.5%	85.2%	45.0%	50.0%	65.0%	77.1%	87.0%	90.8%	Wurzbacher et al. 2018
Located on the LSU											
LR5-Fung	0.75	88.0%	91.8%	93.3%	0%	0%	100.0%	0.0%	0.0%	0.1%	Tedersoo et al. 2008
mbc-3159-R	0.80	87.0%	92.5%	95.1%	0%	0%	100.0%	11.8%	18.1%	84.7%	This study
mbc-2148-R	0.72	86.4%	92.0%	94.4%	0%	0%	0%	4.1%	8.8%	17.6%	This study
LR5-R	-	90.9%	90.3%	94.2%	100.0%	100.0%	100.0%	67.9%	80.2%	85.1%	Tedersoo et al. 2015
RCA95m	-	88.7%	91.3%	94.2%	0%	0%	0%	0.1%	0.1%	4.3%	Wurzbacher et al. 2018
Fun-rOP-R	-	80.9%	88.4%	90.8%	0%	100.0%	100.0%	3.5%	73.7%	84.4%	Wurzbacher et al. 2018

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

Figures

Figure 1

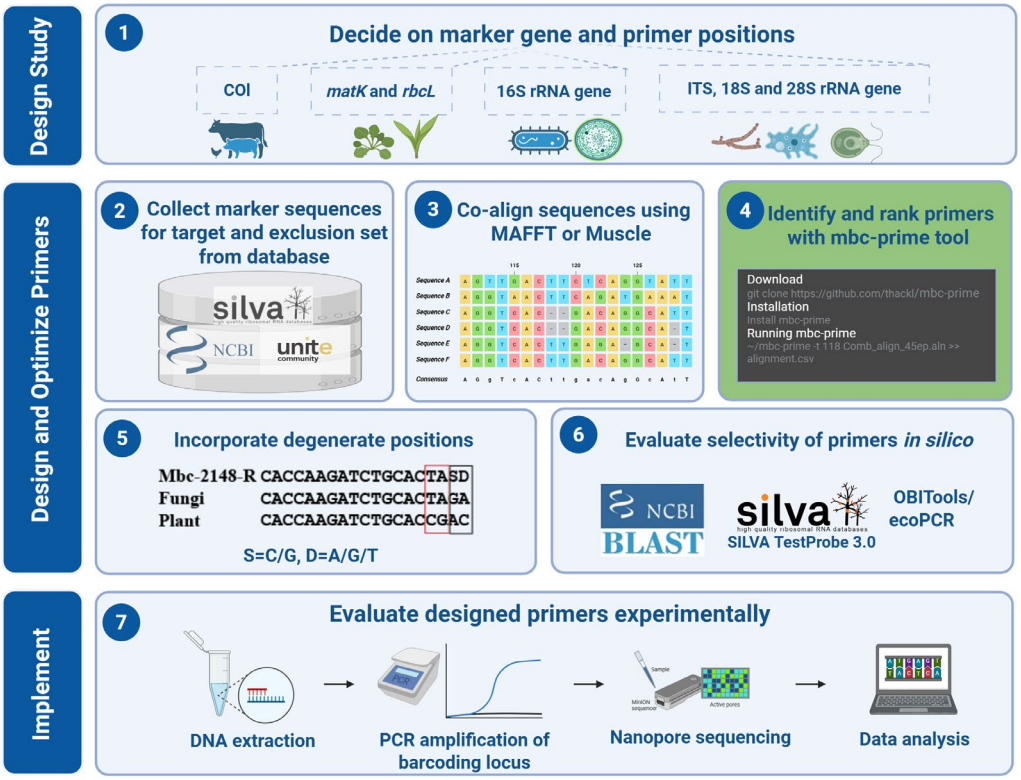


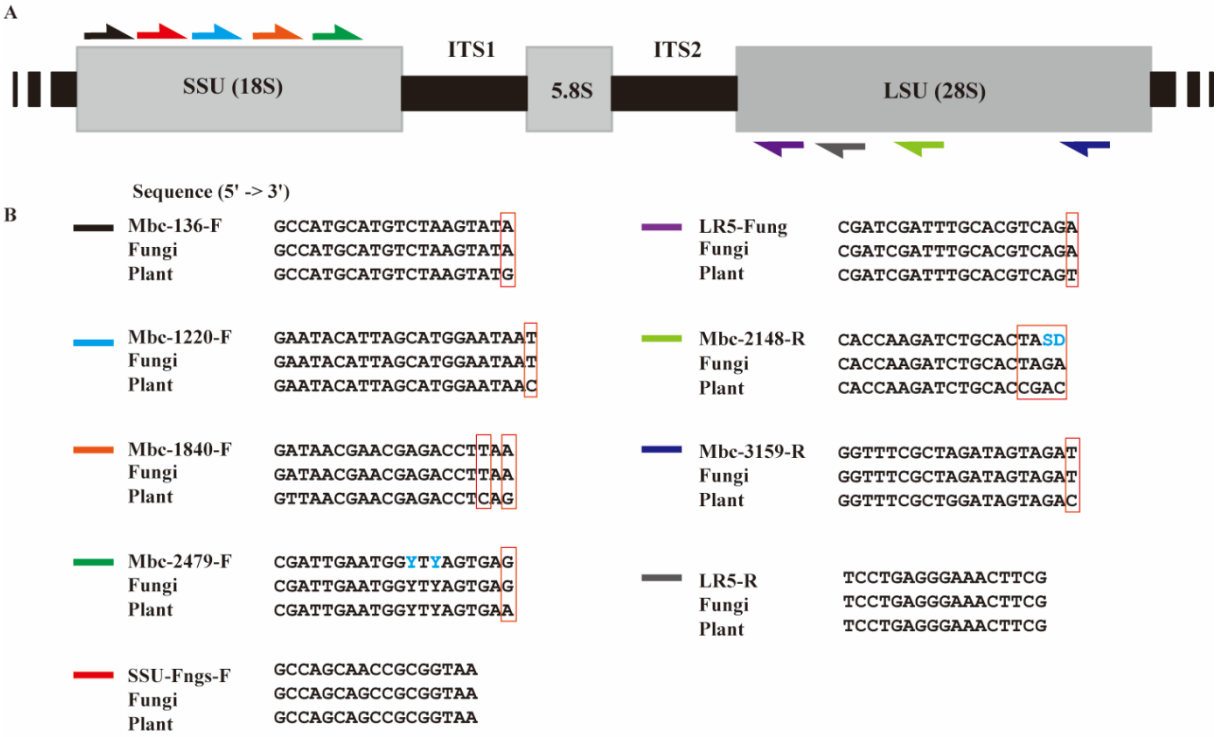
Figure 2

Positional Information | Score | Mismatch Profiles | Primer Sequences

A	B	C	D	E	F	G	H	I
group	pos	trimmed_pos	score	exc: 0,1,2,3,4+ mismatches	inc: 0,1,2,3,4+ mismatches	primer forward	primer reverse	inc-exc-matched
# locus 1								
1	136	46	0.78	[0.0, 0.0, 0.84, 0.88, 1.0]	[0.71, 0.86, 0.87, 0.87, 1.	gccaugcaugucuaaguaua	tatacttagacatgcatggc	||||||||||.|||||||.
1	137	47	0.76	[0.0, 0.04, 0.84, 0.88, 1.0]	[0.73, 0.82, 0.87, 0.87, 1.	ccaugcaugucuaaguauaa	ttatacttagacatgcatgg	||||||||||.|||||||.|
1	138	4	0.61	[0.0, 0.04, 0.84, 0.88, 1.0]	[0.5, 0.76, 0.85, 0.87, 1.6	caugcaugucuaaguauaaa	tttatacttagacatgcatg	|||||||||.|||||||.||
			0.56	[0.0, 0.04, 0.84, 0.88, 1.0]	[0.41, 0.76, 0.84, 0.85, 1.	augcaugucuaaguauaaac	gtttatacttagacatgcat	||||||||.|||||||.|||
#								
			0.5	[0.0, 0.0, 0.84, 0.91, 1.0]	[0.38, 0.61, 0.75, 0.86, 1.	uacuuggauaaccgugguaa	ttaccacggttatccaagta	||||.||||||||||.||||
			0.51	[0.0, 0.0, 0.84, 0.91, 1.0]	[0.41, 0.61, 0.83, 0.88, 1.	acuuggauaaccgugguaau	attaccacggttatccaagt	|||.||||||||||.|||||
			0.5	[0.0, 0.02, 0.84, 0.93, 1.0]	[0.41, 0.61, 0.83, 0.88, 1.	cuuggauaaccgugguaauu	aattaccacggttatccaag	||.||||||||||.||||||
2	262	138	0.55	[0.0, 0.02, 0.84, 0.93, 1.0]	[0.42, 0.7, 0.84, 0.88, 1.6	uuggauaaccgugguaauuc	gaattaccacggttatccaa	|.||||||||||.|||||||
2	263	139	0.7	[0.0, 0.02, 0.87, 0.93, 1.0]	[0.59, 0.84, 0.86, 0.88, 1.	uggauaaccgugguaauucu	agaattaccacggttatcca	.||||||||||.||||||||
# locus 3								
3	280	155	0.6	[0.0, 0.0, 0.91, 0.96, 1.0]	[0.43, 0.77, 0.82, 0.86, 1.	uucuagagcuaauacaugcu	agcatgtattagctctagaa	|||||||||||||||.|||
# locus 4								
4	337	199	0.55	[0.0, 0.0, 0.0, 0.38, 1.0]	[0.47, 0.62, 0.73, 0.8, 1.6	uguauuuauuagauaaaaaa	ttttttatctaataaataca	||.||||||||||.|||..
4	338	200	0.54	[0.0, 0.0, 0.0, 0.02, 1.0]	[0.46, 0.62, 0.74, 0.83, 1.	guauuuauuagauaaaaaac	gtttttttatctaataaatac	|.||||||||||.|||...
4	339	201	0.54	[0.0, 0.0, 0.0, 0.0, 1.0]	[0.46, 0.62, 0.74, 0.83, 1.	uauuuauuagauaaaaaacc	ggtttttttatctaataaata	.||||||||||.|||...|
4	340	202	0.54	[0.0, 0.0, 0.0, 0.02, 1.0]	[0.46, 0.62, 0.79, 0.86, 1.	auuuauuagauaaaaaacca	tggtttttttatctaataaat	||||||||||.|||...|.
4	341	203	0.53	[0.0, 0.0, 0.0, 0.02, 1.0]	[0.45, 0.61, 0.76, 0.83, 1.	uuuauuagauaaaaaaccaa	ttggtttttttatctaataaa	||||||||||.|||...|.|
# locus 5								
5	550	362	0.5	[0.0, 0.0, 0.04, 0.8, 1.0]	[0.34, 0.66, 0.88, 0.95, 1.	aacggguaacggggaauuag	ctaattccccgttacccgtt	.||||||.||||.|||||||
5	551	363	0.56	[0.0, 0.02, 0.78, 0.89, 1.0]	[0.41, 0.74, 0.91, 0.97, 1.	acgggguaacggggaauuagg	cctaattccccgttacccgt	||||||.||||.|||||||
5	552	364	0.56	[0.0, 0.02, 0.78, 0.89, 1.0]	[0.41, 0.74, 0.91, 0.97, 1.	cgggguaacggggaauuaggg	ccctaattccccgttacccg	|||||.||||.|||||||
5	553	365	0.56	[0.0, 0.02, 0.78, 0.89, 1.0]	[0.41, 0.74, 0.88, 0.95, 1.	gggguaacggggaauuagggu	accctaattccccgttaccc	||||.||||.|||||||||
5	554	366	0.57	[0.0, 0.02, 0.83, 0.93, 1.0]	[0.41, 0.74, 0.88, 0.91, 1.	gggguaacggggaauuaggguu	aaccctaattccccgttacc	|||.||||.|||||||||||
5	555	367	0.56	[0.0, 0.02, 0.83, 0.91, 1.0]	[0.41, 0.73, 0.84, 0.89, 1.	gggguaacggggaauuaggguuc	gaaccctaattccccgttac	||.||||.|||||||||||
5	556	368	0.54	[0.0, 0.02, 0.83, 0.91, 1.0]	[0.38, 0.72, 0.79, 0.86, 1.	uaacggggaauuaggguucg	cgaaccctaattccccgtta	|.||||.|||||||||||||
5	557	369	0.54	[0.0, 0.02, 0.83, 0.91, 1.0]	[0.38, 0.72, 0.79, 0.86, 1.	aacggggaauuaggguucga	tcgaaccctaattccccgtt	.||||.|||||||||||||

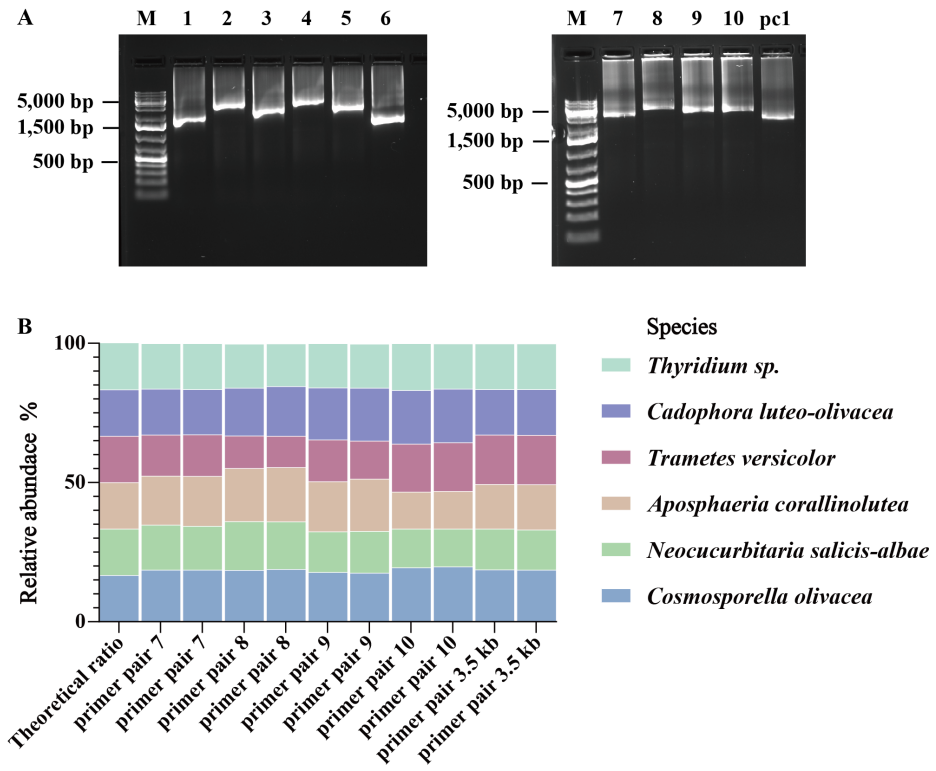
Primer used in this study: mbc-136-F

565 **Figure 3**



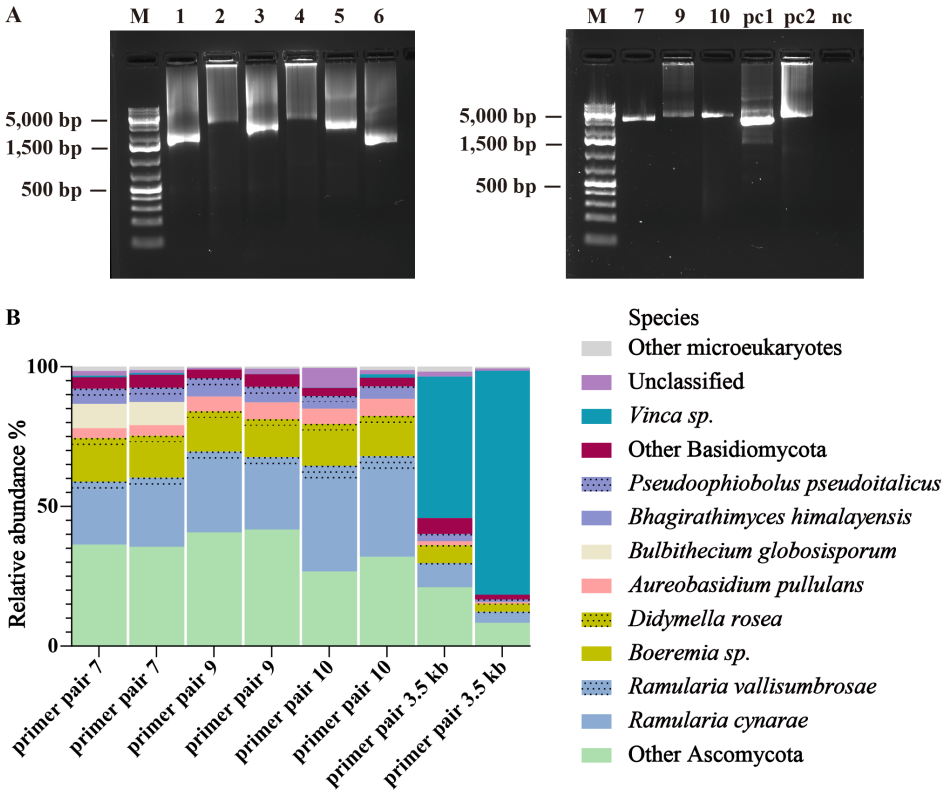
566

567 **Figure 4**



568

569 **Figure 5**



570

571

572 **Figure 1.** Workflow for effective host-exclusive primer design as described in this study. At the core of
573 the workflow stands the computational tool mbc-prime, which effectively identifies candidates for
574 discriminative primers based on multiple sequence alignments of the target and the exclusion sets (step
575 4).

576 **Figure 2.** Screen shot of an example output file of mbc-prime.

577 **Figure 3.** A. Illustration of primer locations in the rDNA locus consisting of SSU, ITS1, 5.8S, ITS2,
578 and LSU; B. Sequence alignments of primers with the consensus sequences of fungi (target set) and
579 plant (exclusion set). Degenerate bases (marked in blue): Y = C/T, S = C/G, D = A/G/T.

580 **Figure 4.** Metabarcoding of a fungal mock community. (A) Gel electrophoresis of rDNA amplicons
581 generated from gene specific primers including overhangs for PCR barcoding; Sample order: M: 1 kb
582 plus DNA ladder; 1-10: amplicons generated with primer pairs 1-10 respectively; pc1: positive control
583 generated with primer pair 3.5 kb. All primer pairs yielded products in the expected size range. (B)
584 Relative abundance of taxonomic groups classified based on the sequencing reads of amplicons
585 generated from the mock community with the different primer pairs (each primer pair has two replicates).
586 Species belonging to the same family are displayed in the same color with different fill patterns. All
587 primer pairs reproduce the overall structure of the mock community, albeit with different biases for
588 different groups.

589 **Figure 5.** Metabarcoding of holobiont DNA of *V. minor*. (A) Gel electrophoresis of rDNA amplicons
590 generated from gene specific primers including overhangs for PCR barcoding; Sample order in both
591 panels: M: 1 kb plus DNA ladder; 1-10: amplicons generated with primer pairs 1-10 respectively; pc1:
592 positive control generated with primer pair 3.5 kb; pc2: positive control generated with primer pair 10
593 with fungal mock community; nc: negative control (nuclease-free water). (B) Relative abundance of
594 taxonomic groups classified based on the sequencing reads of amplicons generated from the genomic
595 holobiont DNA of *V. minor* with the different primer pairs on species level (each primer pair has two
596 replicates).