A rRNA hybridization-based approach for rapid and accurate identification of diverse fungal pathogens

Elizabeth A. Yee[1*], Brooks J. Burt[2*], Poppy C. S. Sephton-Clark[2,3], M. Lauren Donnelly-Morrell[1,4], Isaac H. Solomon[5], Evan Mojica[7], Christina A. Cuomo[1, 6#], Roby P. Bhattacharyya[1, 8#]

[1]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge MA 02142 USA
[2]Genomic Center for Infectious Disease, Broad Institute of MIT and Harvard, Cambridge MA 02142 USA
[3]Valo Health, 75 Hayden Avenue Suite 200, Lexington, MA 02421 USA
[4]Department of Pathology, Massachusetts General Brigham, Harvard Medical School, Boston, MA USA
[5]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School Boston, MA 02115 USA
[6]Department of Molecular Microbiology and Immunology, Brown University, Providence RI 02912 USA
[7]Microbiology Laboratory, Department of Pathology, Massachusetts General Hospital, Boston, MA 02114, USA
[8]Infectious Disease Division, Department of Medicine, Massachusetts General Hospital, Boston MA 02114 USA
*Denotes Co-First authorship EAY and BJB contributed equally.
[#]To whom correspondence should be addressed

Abstract

Invasive fungal infections are a global threat; early diagnosis is critical for patient outcomes, but current diagnostic measures remain notoriously slow. Here we extend our multiplexed, hybridization-based rRNA-targeted strategy for rapid, sensitive pathogen identification, previously designed for diverse bacteria and *Candida* species, to identify diverse fungal pathogens. We created a set of 91 probes targeting 86 medically relevant fungal species, designed to recognize regions of differential conservation across taxonomic groupings, from class- to species-specific probes. We assessed assay performance across a Training Set of 93 clinical isolates spanning 32 species of common fungal pathogens across 18 genera, with Pearson correlations of probeset reactivity profiles (PSRPs) identifying the pathogen at the species, genus, and family level with 83%, 94%, and 95% accuracy, respectively, in a leave-one-out analysis. We developed a more sophisticated classifier on this Training Set, using taxonomic categories to select progressively more informative probes at each taxonomic level. After optimization, we assessed performance on an independent Validation Set of 54 clinical isolates spanning the same species as the Training Set, with 91%, 94%, and 98% at the species, genus, and family levels, respectively. We piloted our assay on formalin-fixed paraffin-embedded (FFPE) tissue, demonstrating rapid, culture-independent fungal identification from this high-value clinical sample type, often the sole specimen available. The assay requires <30 minutes hands-on time (or <65 minutes from FFPE tissue), returning results in <8 hours from cultured specimen or FFPE tissue to answer on an RNA detection platform available in clinical laboratories.

Importance

Timely identification of fungal pathogens is critical for patient outcomes. A multiplexed hybridization-based assay targeting rRNA enables accurate identification of >50 species of pathogenic fungi from crude lysates of cultured clinical isolates.

Introduction

Invasive fungal infections (IFIs) are a growing threat to human health, accounting for 1.6 million global deaths annually.[1] Advances in immunosuppressive drugs and chemotherapy have resulted in larger immunocompromised populations vulnerable to opportunistic pathogens. [2–4] Early diagnosis of fungal infections is critical for patient outcome, but is notoriously slow, relying heavily on expert clinical mycologists identifying morphology from cultured or fixed patient specimens.[1,5,6] As such, healthcare systems are in need of rapid and accurate diagnostic tools to identify fungal infections.[3-5]

Progress has been made in fungal diagnostics using several different modalities. Serological testing can help diagnose IFIs by detecting fungal antigens, or host antibodies reflecting exposure, but many assays have cross reactivity.[7,8] Serology platforms often have a quick turnaround time and no culturing requirement, but are limited by a narrow species range or high false positivity rate for more general fungal markers such as 1,3-β-D-glucan.[3,7,9,10] Widely used for bacterial identification, matrix assisted laser desorption/ionization time of flight (MALDI-TOF) mass spectrometry is gaining traction as a method of fungal identification.[7,11,12,13] Although this method is easy to run, it requires culture growth which is not always possible, and identification is complicated by variable surface proteome expression across the fungal life cycle, making fungi harder to identify against reference databases.[5,7,14]

65 As costs fall and efficiency increases, sequencing is increasingly used for fungal identification.
66 Targeted amplicon sequencing of the internal transcribed spacer region (ITS), ribosomal RNA
67 genes (rDNA), or additional highly conserved genes has been widely adopted by reference
68 laboratories to discriminate between fungal species.[5,9,15,16] However, fungal DNA in primary
69 clinical samples is often very scarce, requiring amplification that can result in false positives
70 from low-level contaminants.[17] Conversely, false negatives can occur due to inhibition of PCR
71 enzymes.[17] Despite these improvements in fungal diagnostics, no single strategy or commercial
72 test can identify a wide variety of fungal organisms rapidly and accurately. Additionally,
73 sequencing assays are often send-out tests with lengthy turnaround times, limiting their clinical
74 utility. There is thus a critical need to explore diagnostic options that balance sensitivity,
75 specificity, scalability, and breadth of coverage.
76
77 Previously, we described a novel rRNA-based strategy for rapid, sensitive bacterial identification
78 directly from primary clinical samples or cultured blood samples utilizing a commercially
79 available assay platform for multiplexed hybridization-based RNA quantification
80 (NanoString).[18,19] This assay, which we named phylogeny-informed rRNA-based strain
81 identification (Phirst-ID), uses 192 probes targeting regions in the 16S and 23S rRNA from 98
82 medically relevant bacterial species.[18] Each target region was selected for specificity at the level
83 of a pathogenic species, or higher taxonomic levels.[18] The fluorescent read-out from these
84 probes results in differential hybridization that creates a unique “fingerprint” pattern of reactivity
85 that accurately identifies each species.[18] We later extended this approach to yeast, designing a
86 Candida Phirst-ID probeset that discretely identified 11 medically relevant *Candida* species with
87 100% accuracy by targeting the 18 and 28S rRNA subunits.[20] By recognizing the highly
88 abundant rRNA itself, rather than the DNA that encodes it, this approach can detect and identify
89 <30 cells without amplification.[20]
90

91 Here we further expanded this approach by creating a panel of 91 probes targeting 86 medically
92 relevant fungi, targeting regions of the 18S and 28S rRNA subunits with differential conservation
93 across taxonomic groupings, ranging from class to species (**Fig. 1**). This species list targeted
94 the Fungal Priority Pathogen List from the WHO and intentionally excluded 56 common
95 environmental fungi as an outgroup in the design process. The resulting "Pan-Fungal Phirst-ID"
96 assay was first tested on a curated Training Set of 93 clinical isolates, then verified on a
97 Validation Set of 54 independent isolates. Using Pearson correlations, we achieved a positive
98 species match rate of 84%. For some isolates, higher order taxonomic probes dominated the
99 signature, compromising accuracy by overpowering more specific signals from genus and
100 species level probes. To address this, we developed an optimized classifier, increasing species-
101 level accuracy to 89%. To assess potential utility on primary specimens, we piloted the assay on
102 archived clinical formalin-fixed paraffin-embedded (FFPE) tissue samples.
103
104 **Materials and Methods:**
105 **Probeset Design**
106 We compiled a comprehensive list of medically relevant fungi, using the WHO Fungal Priority
107 Pathogen list as a guide, for training and validation of our probeset. We collated the 18S and
108 partial 28S sequences for 86 medically relevant fungal species primarily from the Silva
109 database; where sequences were too short or not available, we retrieved regions from NCBI
110 assemblies for the following: *Mucor circinelloides*, *Emergomyces orientalis*, *Epidermophyton*
111 *floccosum*, *Coccidioides posadasii*, *Coccidioides immitis*, *Histoplasma capsulatum*,
112 *Emergomyces pasteurianus*, *Exophiala dermatitidis*, *Trichophyton indotiniae*, *Kodamaea*
113 *ohmeri*, *Wickerhamomyces anomalus*, *Kluyveromyces marxianus*, *Malassezia dermatis*,
114 *Enterocytozoon bieneusi*. Additional 18S sequences were selected from Silva for 56
115 environmental fungal species to serve as an outgroup during probe design to minimize
116 unintentional cross reactivity. Due to inconsistent annotation of the 28S sequences, we were

117 unable to obtain full-length 28S sequences for all ingroup species. For outgroup species, we
118 were only able to obtain reliable 18S sequences. We compiled a full taxonomic classification of
119 each species (**Fig. S1, Table S1**). Employing a similar strategy used in the Pan-Bacterial Phirst-
120 ID probeset design[18]. We profiled each desired target sequence using NanoString's proprietary
121 probe design algorithm to identify all putative pairs of consecutive 50mer probe-binding regions
122 that meet parameter specifications. Specifically, probe sequences are assessed for predicted
123 binding kinetics, thermodynamics, secondary structure, and sequence composition for the
124 desired targets while minimizing predicted cross-reactivity against non-targeted outgroup
125 species, or the human genome. For species-specific probes, outgroups included and all other
126 pathogenic and environmental fungi; we sought probes predicted to bind exclusively to the
127 targeted species in a given rRNA region, whose sequence was unique to that species (**Table**
128 **S2**). For probes targeting higher taxa (genus, family, order, or class), we sought probes
129 predicted to bind all members of that taxon, (including either pathogen or environmental fungi,
130 since including only pathogens would likely impose unrealistic constraints in sequence space),
131 while excluding all outgroups not belonging to that taxon. For *Candida* species, which have
132 undergone considerable phylogenetic reclassification of late, we kept the clinically familiar
133 genus classification *Candida* for all (*Candida, Candidozyma, Pichia, Nakeseomyces,*
134 *Meyerozyma*, etc.), which likely changed design constraints on the genus-level probe. This
135 decision could be revisited in future iterations of the probeset.
136
137 The final probeset includes 60 probes designed to be species-specific, as well as 31 probes
138 designed to recognize rRNA regions conserved across higher taxonomic groupings, including
139 14 genus-level probes, 8 family probes, 5 order probes, and 4 class probes (**Fig. S1**). During
140 testing, the *Mucor circinelloides* 28S probe yielded substantial cross-reactivity against nearly all
141 tested species. We thus removed this probe analytically, resulting in a final set of 91 probes
142 used for all analyses (**Table S1**). Also included in the probeset are 12 control probe pairs (6

143 negative, 6 positive) directed at External RNA Controls Consortium (ERCC) targets as
144 previously described.[18] The 6 positive controls are directed at ERCC spike-ins included in every
145 hybridization reaction at pre-specified concentrations; the 6 negative controls are directed at
146 biorthogonal targets not included in the hybridization.
147
148 **Sample acquisition**
149 Most of our clinical isolates were collected from the Massachusetts General Hospital (MGH)
150 Clinical Microbiology Laboratory, for both Training and Validation Sets. These isolates were
151 provided as cell pellets suspended in Trizol solution after culture isolation on standard fungal
152 media in the clinical laboratory. Isolates were identified through standard clinical microbiology
153 workflows as part of routine patient care and shared with the investigators under Mass General
154 Brigham Institutional Review Board protocol 2015P002215. In keeping with diagnostic
155 limitations in these standard clinical workflows, some isolates were only identified to the genus
156 level, including all *Curvularia*, *Coccidioides*, *Fusarium*, *Cunninghamella*, *Mucor*, *Rhizopus* and
157 *Talaromyces* isolates. Some *Candida* species were supplemented from our own laboratory
158 collection and the Antibiotic Resistance Isolate Bank drug resistant *Candida* (CAN) and *Candida*
159 *auris* (CAU) collections. Additional samples were generously shared with us from the following
160 sources: *Histoplasma capsulatum* cell pellets suspended in Trizol from Dr. Sinem Beyhan
161 (JCVI)[21], *Pneumocystis jirovecii* gDNA from human autopsy samples from Dr. Joseph Kovacs
162 (NIH), and *Blastomyces dermatiditis* purified RNA*, Cryptococcus gattii, Paracoccidioides*
163 *brasiliensis, Sporothrix schenckii* and *Talaromyces marneffei* purified gDNA from the Fungal
164 Genomics Group at the Broad Institute.[21–25] Per our biosafety regulations, all cultured isolates
165 from external collaborators were received in Trizol. If we were unable to obtain a cultured
166 isolate, we utilized either gDNA or RNA as available. Coded/de-identified FFPE tissue samples
167 were obtained from the Brigham & Women's Hospital (BWH) Pathology Department, after visual
168 screening to identify archived clinical samples from 2019-2025 with varying levels of visible

169 fungal forms (ranging from rare to confluent). Fungal identification of these autopsy and surgical
170 pathology cases were based on integration of all available laboratory results from anatomic
171 pathology histochemical stains, culture isolation from concurrently collected tissue samples, and
172 molecular testing of FFPE tissue, frozen tissue, or culture isolates.[26] **Table S3** lists all samples
173 and sources.
174
175 **Training and Validation Set parameters**
176 To test the Pan-Fungal Phirst-ID probeset, we acquired as many samples from our fungal
177 pathogen list as possible. In all, we collected 171 samples (isolates, gDNA, or RNA) from 53
178 species. To assess assay performance as rigorously as possible, we divided these samples into
179 Training and Validation Sets based on the number of samples of each species we were able to
180 collect (**Table S3a, b**). In the ideal case, we included 3 samples of the same genus and species
181 in the Training Set, with any remaining samples (up to 3 more) used for validation. If we were
182 only able to obtain 3 samples total, we included 2 in the Training Set and kept the third in the
183 Validation Set (**Table S3a, b**). In the end, this led to 93 samples across 33 species in the
184 training set, with 54 samples across 32 species in the Validation Set. For 21 additional species,
185 we were only able to obtain 1 (n = 18) or 2 (n = 3) samples. Since this was insufficient to
186 assess classification accuracy, we did not include these samples in either the Training or
187 Validation Sets, but we do report their results separately. For assessment of classification
188 performance, background subtraction and normalization was carried out exactly as previously
189 described,[18,20] and classifiers were developed and iterated using only data from the Training
190 Set, before being run on the Validation Set with no changes to any analytical processes. All
191 Validation Set analyses were done with authors blinded to sample identification until after
192 predictions were finalized. For FFPE samples, pathology co-authors (MLD-M and IHS) selected
193 samples, whereas the remaining co-authors (notably EAY, BJB, CAC, RPB) stayed blinded to

194 the identities of all species until all thresholding decisions, classification algorithms, and identity
195 predictions were finalized.
196
197 **Sample processing**
198 Cultured isolates collected from MGH
199 Clinical isolates were received from the clinical microbiology laboratory in Trizol and processed
200 as follows: 125µL of each sample in Trizol was added to a 2ml screw cap tube (MP
201 Biomedicals, CAT# 76044-692) with an additional 125µL of Trizol reagent (ThermoFisher
202 CAT#15596026) with 100µl of 0.1uM zirconia/silica beads (Biospec, CAT# 11079101z), then
203 bead-beaten for 2 rounds on the FastPrep 5G (MP Biosciences) for 90s, at 10 m/s. 50µL of
204 chloroform was added to each sample, inverted, and the mixture was spun down at 12,000 x g
205 for 15 minutes at 4C. 50µL of the aqueous phase was transferred to a fresh 1.5 ml Eppendorf
206 tube. To prevent degradation of the RNA in this crude lysate, 50µL RLT buffer (Qiagen CAT#
207 79216) with 1% β-mercaptoethanol (β-ME), an irreversible chemical RNase inhibitor, was added
208 to each lysate prior to storage at -80ºC.
209 gDNA samples of selected isolates
210 gDNA samples from the Fungal Genomics group at the Broad Institute were used to assist in
211 species coverage for members of the Fungal Priority Pathogen list. Approximate concentrations
212 were measured using the Nanodrop One system (ThermoFisher CAT#ND-ONE-W). Dilutions of
213 some gDNA samples were made in 1x Phosphate Buffered Saline (PBS) to ensure probe
214 reactivity patterns were within control cutoffs and max probe limits as described below.
215 Cultured isolates prepared from ARBank isolates
216 The ARBank *C. auris* (CAU) and drug-resistant *Candida* (CAN) collections were used to
217 supplement *Candida* species for Training and Validation sets (**Table S3a, b**). Cultured isolates
218 prepared in house and were processed as previously described.[20]
219

220 **NanoString Data generation**

221 Our custom Pan-Fungal Phirst-ID probeset was run using the Elements assay variation on the
222 standard NanoString assay for multiplexed RNA detection. Briefly, lysates from cultured
223 organisms were diluted first at a 1:10 ratio in PBS to avoid assay saturation. For gDNA and
224 RNA samples, dilution was not always necessary depending on the concentration of the
225 sample. For all samples, 1.5μL of lysate or nucleic acid was incubated with unlabeled probe
226 pairs for each target (IDT) and Elements TagSet-96 reagents (NanoString CAT#121000608).
227 Hybridization conditions were standard, except for 2 modifications: lysates were incubated at 95
228 °C for 2 minutes immediately prior to hybridization to denature secondary structural elements
229 and disrupt protein binding in rRNA targets, and hybridizations were incubated for one hour
230 instead of the recommended 16-24 hours, exactly as we have previously done[18,20]. Hybridized
231 samples were then run on the NanoString nCounter platform and subsequently analyzed using
232 the nCounter Digital Analyzer.

233

234 Cell densities from samples cultured by collaborators, and from paraffin scrolls, were unknown
235 prior to running on the nCounter platform. We established three parameters to address sample
236 overloading: (1) positive control signals out of the expected order given known spike-in
237 concentrations, suggesting cross-reactivity; (2) any negative control value > 100 total read
238 counts; or (3) maximum probe count >100,000, where saturation might affect relative ratios. If
239 any of these 3 parameters were met, the affected sample was re-run with increased dilutions
240 until the issue resolved. The maximum number of re-runs required for a given sample was up to
241 three to meet our stated parameters. Across training and Validation Sets, 62 lysates required at
242 least one re-run, with four samples requiring additional runs. All such decisions were made
243 solely based on these three criteria, prior to any assessment of prediction accuracy.

244

245 **Data processing and visualization**

246 Raw binding data (counts per probe) were compiled using NanoString nSolver software (v3.0).
247 Raw counts were normalized using positive and negative control spike-ins provided by
248 NanoString using custom scripts in R, and background signal for each probe was subtracted
249 based on the average of three blank lanes, exactly as previously described.[18] Pearson
250 correlations for the resulting normalized, background-subtracted probeset reactivity profiles
251 were subsequently calculated across samples using the "cor" function in R (version 4.3.1).
252 Heatmaps were generated by the Pheatmap package v1.0.12 in R. Probes and samples were
253 arranged phylogenetically.
254
255 **Co-PILOT Classification**
256 We developed a hierarchical classification model based on Pearson correlations (**Fig. S4**)
257 termed **Co**mplementarity-based **P**hylogeny-**I**nformed **L**evel-**O**riented **T**raversal (Co-PILOT),
258 coded entirely in R (version 4.3.1 github) Initial Pearson correlations were calculated for each
259 isolate to identify most similar samples based on probe-binding signatures. The taxonomic
260 group (starting with class) was then found for the closest match. All other isolates not in this
261 taxonomic group were then removed, as were probes not specific to that group (**Fig. S4a-b**).
262 This process was repeated until the species level, or if there were no more samples to compare
263 to. The Co-PILOT algorithm was developed and iterated using only our Training Set, then
264 deployed without modification on our Validation Set, considering each Validation sample one at
265 a time and using matches to the Training Set as its basis for identification decisions. When the
266 best match was to a sample from the MGH clinical laboratory for which identification was only
267 determined to the genus level, Co-PILOT was unable to make a species prediction but could
268 predict identity to the genus level.
269
270 **Pilot study of FFPE samples**

271 FFPE scrolls (60μm per block in each tube) were obtained from the Brigham and Women's
272 Hospital Pathology Department. Samples FP001-FP012, FP016-019 were processed using the
273 RNeasy FFPE kit (Qiagen CAT# 73504), per NanoString recommendations. Given inconsistent
274 results at an interim assessment, we processed the remaining samples (FP020-030) using the
275 quick DNA/RNA FFPE kit (Zymo CAT# R1009) (**Table S3c**). Samples were first deparaffinized
276 using proprietary deparaffinization solutions provided by each kit followed by heating tissue to
277 either 56ºC followed by 80ºC (Qiagen) or 54ºC (Zymo), separating wax from tissue. Prior to
278 moving to column purification, samples underwent two rounds of bead beating as described
279 previously and spun down at 13,000xg for 1 min before loading onto column for final purification.
280 All lysates were stored at -80ºC after elution. FFPE samples were analyzed as a separate
281 dataset relative to Training and Validation Sets. Compiled FFPE data was analyzed for species
282 classification against a “superset” containing Training, Validation, and rare incidence isolate
283 datasets, mirroring how we would expect to deploy this method on primary tissue samples going
284 forward. After visual inspection of resulting PSRPs for all samples, we had low confidence in
285 samples that had <1000 raw counts from any probe (see **Fig. S9**). Using this metric as a
286 baseline, we divided samples based on their max probe counts into low and high confidence
287 groups, though we report Co-PILOT’s best match for each sample in each group. As described
288 above, if a species identification was not defined for a given sample, the final best match was
289 made at the genus level.
290
291 **ITS sequencing workflow to confirm species ID**
292 Clinical isolates were received in Trizol, and additional cultured material was unavailable. For
293 those unable to be identified to the species level by standard workflows, we sought to obtain
294 genomic DNA (gDNA) to enable ITS sequencing. As these samples were intended for extraction
295 of RNA rather than DNA, we omitted the chloroform extraction, which improves partitioning of
296 RNA from gDNA, hoping this would result in enough gDNA for PCR. 100μL of remaining

297 inactivated sample not used in original lysate prep was combined with an additional 300µL of
298 Trizol, then underwent 2 rounds of bead beating as previously described, then processed using
299 the Zymo DirectZol RNA kit (CAT#: R2050). Samples were eluted in 25ul of Elution buffer as
300 specified. ITS1 and ITS2 regions were amplified using the NEB Q5 hot start master mix
301 (CAT#:M0491S) using primers and cycling conditions as previously described.[20,27] PCR
302 products were visualized on an E-Gel Power Snap Electrophoresis Device (Invitrogen
303 CAT#:G8100) using a 1% Agarose E-Gel (Invitrogen CAT#: G401001) for the presence of
304 bands for ITS1 or ITS2. Samples with bands corresponding to either ITS1, ITS2 or both were
305 purified using the Zymo DNA Clean and Concentrate (CAT#: D4033) and sent to Plasmidsaurus
306 for amplicon sequencing. Results were first analyzed in Geneious and trimmed for quality using
307 standard parameters; trimmed sequences then were submitted to UNITE using the regular
308 output style and 2 sequence alignments per sample for analysis.[28]
309
310 **Results**
311 **Pathogen selection and Pan-Fungal Phirst-ID probeset design**
312 We selected 86 medically relevant fungi using the WHO Priority Pathogen List as key species to
313 target with our panel, and an outgroup of 56 environmental and non-medically relevant fungi
314 (**Table S1**). Targeting species-specific variable regions of the 18 and 28S rRNA subunits, we
315 designed 61 species-specific probes, one of which was removed from analyses due to
316 excessive cross-reactivity (see Methods). For the remaining targeted species, we could not
317 design a species-specific probe within the constraints of the NanoString assay due to
318 phylogenetic constraints. For the closely related *Candida albicans* and *Candida dubliniensis,*
319 and *Aspergillus fumigatus* and *Aspergillus clavatus,* we designed single probes predicted to
320 selectively recognize each pair of species.
321

322 As previously described,[18] we also designed higher-order taxonomic probes to improve
323 accuracy in identifying target species through multiple recognition events, creating a unique
324 probeset reactivity profile (PSRP) that serves as a "fingerprint" for each isolate tested. These
325 higher taxonomic order probes can also enable partial characterization of pathogens not
326 included in probeset design.[18] To do so, we analyzed the 18 and 28S sequences for variable
327 regions shared across species of the same genus, family, order, or class conserved enough to
328 recognize all members within a given taxonomic classification, while excluding outgroup
329 species.[18,20] This resulted in 14 genus probes, 8 family probes, 5 order probes and 4 class
330 probes, culminating in a Pan-Fungal Phirst-ID probeset of 91 probes (**Fig. 1**; **Fig. S1, Table S2**).
331
332 **Collation of Training and Validation Sets**
333 To rigorously test our panel, we collected isolates from 18 of 19 members of the WHO Fungal
334 Priority Pathogen List, as well as 35 other clinically important species available from the MGH
335 clinical microbiology laboratory. Isolates were split into Training and Validation Sets based on
336 availability, with three isolates per species used for Training when possible and all remaining
337 isolates used for Validation; when only three total isolates were available, two were used for
338 Training and one for Validation (**Table S3a, b**; see Methods). Our Training Set contains 93
339 isolates spanning 18 unique genera and 26 defined species, and our Validation Set contains 54
340 isolates with the same genus and species numbers. For 7 genera, we were unable to get
341 species characterization from the MGH clinical lab, due to standard clinical identification
342 methodology. In addition, we collected several clinical isolates that were too rare (n = 1 or 2 in
343 our collection) to include in our Training and Validation rubric. Though unable to rigorously
344 validate classification performance, we tested 24 of these isolates, spanning an additional 14
345 genera and 21 species not included in our Training and Validation Sets (**Fig. S2**).
346
347

348 **Unique reactivity profiles accurately distinguish species**

349 To quantify differences in PSRPs between species, we first calculated a pairwise Pearson
350 correlation coefficient for each sample against every other sample in the Training Set (**Fig. S3**).
351 This simple analytical approach incorporates data from all probes tested in the probeset across
352 the entire reactivity profile. Because of sequence conservation patterns in rRNA genes, PSRPs
353 of this probeset should reflect species identity. The strongest correlations should exist within
354 isolates of the same species, whereas non-species matched pairs should have lower correlation
355 coefficient (**Fig. 2**).[18,20] In a "leave one out" analysis for all isolates in the training set, 83%
356 matched best to members of the same species, with most mismatches occurring to closely
357 related species (**Fig. 3**). For 15 samples not identified to the species level, the best match in all
358 15 was to the same genus, the closest available taxonomic match in the Training Set. In all,
359 93% of isolates matched at the genus level, and match rates steadily increased for broader
360 taxa, reaching 100% at the class level (**Fig. 3a**).

361

362 **Leveraging taxonomy level information to improve classification accuracy**

363 For many isolates in our training set, the higher taxonomic order probes resulted in greater
364 signal than the species-specific probes (**Fig. 4a**). One notable example is the *Aspergillus* genus,
365 where class probes dominated the signal readout compared with species probes (**Fig. 4b**). This
366 systematic difference in probe binding by taxonomic level of the probe target, despite perfect
367 complementarity between probe and target, was not observed in bacteria.[18] This posed a
368 unique challenge to accurate classification by Pearson correlation, which intrinsically weights
369 higher signals and thus undervalues species-level probes with lower maximal reactivities,
370 reducing species-level accuracy.

371

372 To address this uneven signal distribution across probes targeting different taxonomic levels, we
373 developed a hierarchical classifier, still based on Pearson correlations but progressing stepwise

through taxonomic classifications. We called this classifier **Co**mplementarity-based **P**hylogeny-**I**nformed **L**evel-**O**riented **T**raversal (Co-PILOT). This approach begins with the full PSRP, but traverses down the taxonomic classification, at each step eliminating any species and probes not included in the best-matched taxon. **Fig S4** illustrates a specific example of how Co-PILOT analyzes an *Aspergillus* sample, narrowing from all probes and samples (**Fig. S4a**), through class (best match = *Eurotiomycetes*), order (*Eurotiales*), and family (*Aspergillaceae*)-specific ones (**Fig. S4b-d**), to only those from the *Aspergillus* genus (**Fig. S4e**). At each level, only probes and samples relevant to the selected taxon are used to recalculate Pearson correlation coefficients. The best-matched sample from the Training Set at the final step, usually at the species level, is taken as Co-PILOT's final identification. If the best match from the Training Set is a clinical sample only identified to the genus level (e.g., *Mucor* spp.), then Co-PILOT's final step would be at the genus level. In this manner, the entire probeset is still used to classify each sample, but at each taxonomic level, only the probes designed to be informative in distinguishing among members of that taxon are considered. This method allows these more precise reactivity patterns to inform identity, rather than idiosyncratic differences between higher-order probes that might otherwise dominate a Pearson correlation. Critically, this novel classifier was devised and iteratively refined based only on the Training Set, before data from the Validation Set was examined. Thus, although Co-PILOT improved classification accuracy for the Training Set (**Fig. S5**), this should be viewed as "over-trained". However, its performance on the Validation Set represents an independent assessment of classification accuracy.

The Pan-Fungal Phirst-ID assay accurately classifies an independent Validation Set of diverse fungal pathogens

Fifty-four independent clinical isolates, representing the same genera and species as the Training Set, were tested as a Validation Set for the Pan-Fungal Phirst-ID assay (**Fig. S6**). The resulting PSRPs for these isolates were compiled into a separate dataset that was only released

400 for analysis after the entire Training Set had been analyzed and the Co-PILOT algorithm had
401 been finalized. Pearson correlations identified a similar fraction of species (85%) and higher-
402 order taxa in the Validation Set as the Training Set (**Fig. 5a, b**). Co-PILOT performed better,
403 identifying 89.3% of species in this independent Validation Set, with 96.2% concordance at the
404 family level and >98% for all higher-order taxa (**Fig. 5c**).
405
406 **Investigation of discrepant classifications in the Validation Set**
407 In the Validation Set, Co-PILOT misidentified 4 samples out of 54, all among species also
408 misidentified in leave-one-out analysis in the Training Set. Incorrect ID matches were correctly
409 binned at high order taxonomic levels but mismatched to different but related genera or species.
410 Specifically, misclassifications were made between *Cryptococcus neoformans* and
411 *Cryptococcus gattii*, and between *Scedosporium apiospermum* and *Lomentospora prolificans*.
412 Both pairs are closely related phylogenetically; *C. gattii* was formerly referred to as a serovar of
413 *C. neoformans,*[29,30] and *L. prolificans* was formerly classified as a *Scedosporium*.[31] In examining
414 reactivity profiles (**Fig. 2, Fig. S6**), despite probes designed to distinguish between each
415 species pair, no real distinction exists between their PSRPs. For the pair of *Cryptococcus*
416 species, a probe designed to selectively recognize the *C. gattii* 28S rRNA subunit failed to
417 generate a signal for either species. This *C. gattii* species-specific probe matches a *C.*
418 *deuterogattii* (VGII) sequence. Our training set included one *C. deuterogattii* sample, but this did
419 not react with the species-specific probe, suggesting this probe has a general binding issue and
420 should be redesigned and confirmed for alignment and potential reactivity to the other *C. gattii*
421 species complex members. For *S. apiospermum* and *L. prolificans*, two problems arose in these
422 closely related species: a probe designed to be specific for the *L. prolificans* 28S subunit cross-
423 reacted with *S. apiospermum* as strongly as with *L. prolificans*; whereas a probe designed to
424 selectively recognize the *S. apiospermum* 18S subunit generated no signal for either species.
425 Relevant family and class probes recognized both species comparably, as expected.

426 We investigated one additional discrepant classification in the Validation Set, which both
427 Pearson correlations and Co-Pilot called a *Rhizopus*, but the clinical identification labeled it a
428 *Mucor*. On visual inspection, its reactivity pattern clearly resembled the *Rhizopus* from our
429 Training Set more than the *Mucor* (**Fig. S7**). Given the difficulty in clinically distinguishing
430 among the *Mucorales* by morphology,[32,33] we pursued ITS sequencing to clarify, which showed
431 100% match to *Rhizopus microsporus* (**Table S3b**), not a *Mucor* species, consistent with our
432 assay's call. Reclassifying this isolate as a *Rhizopus* instead of a *Mucor* improves the accuracy
433 of our Co-Pilot classifications on the Validation Set (to 91%, 94%, and 98% at the species,
434 genus, and family levels respectively).
435
436 **Pilot use of Pan-Fungal Phirst-ID on clinical Formalin- Fixed Paraffin- Embedded (FFPE)**
437 **tissue samples**
438 Fixation of tissues using formalin and embedding in paraffin blocks is the gold standard method
439 for preparing human tissues for histopathologic diagnosis, including microorganism
440 identification.[34,35,36] However, this process prevents subsequent growth of microbes; thus, if
441 infection is only suspected after tissue fixation (e.g., in a lung nodule removed for suspicion of
442 neoplasia), culture-based pathogen identification is precluded. Often such samples are obtained
443 via surgical procedures that are impractical or even dangerous to repeat, meaning pathogen
444 identification must be done either imprecisely by morphology, or via nucleic acid extraction and
445 PCR from the fixed tissue.[37,38] Since Pan-Fungal Phirst-ID requires only hybridization and no
446 enzymology, we hypothesized that it might perform comparably to PCR on crude nucleic acid
447 preparations from FFPE tissue. We thus investigated the performance of our Pan-Fungal Phirst-
448 ID probeset on nucleic acids extracted from archived clinical FFPE samples on which fungal
449 forms were observed via microscopic examination, plus two negative controls with no fungal
450 forms (**Fig. S8**). Pan-Fungal Phirst-ID correctly identified FFPE samples, largely depending on
451 the quality of sample and NanoString read counts we were able to extract from each sample.

452 When plotting heatmaps of read counts, 55% displayed visually poor probe binding behavior;
453 thus, prior to running Co-PILOT, we opted to exclude these as non-informative and focus our
454 analysis on samples with sufficient NanoString reads (**Fig. S9**). Of the 12 "high confidence"
455 samples, 8 were correctly identified, with 2 others notably close: an *A. fumigatus* was called *A.*
456 *clavatus*; and a *Madurella tropicana*, not included in our expanded Training Set, matched with
457 the closest possible phylogenetic match, an environmental *Chaetomium* isolate). We also found
458 a negative control with a large *Candida albicans* signature (**Fig. 6**), a possible contaminant in
459 sample collection or processing; the other negative control showed low signal as expected (**Fig.**
460 **S9**).
461
462 **Discussion**
463 Fungal infections are a growing global health problem, and current diagnostic methods are
464 unable to meet the demands of healthcare systems. Currently available testing measures have
465 significant limitations from financial burdens for instrument costs or expert personnel, frequent
466 false positives for biomarkers, and limited organism range for which specific tests are available.
467 We previously developed a molecular diagnostic approach based on rRNA hybridization, Phirst-
468 ID, to rapidly and accurately identify bacterial pathogens[18] and *Candida* species[20] using a
469 commercial detection system present in clinical pathology laboratories primarily for oncological
470 use.[39–41] Here, we extended this approach to target 86 medically relevant fungi in our Pan-
471 Fungal Phirst-ID panel.
472
473 rRNA is an appealing target for hybridization-based identification due to its extreme
474 conservation within species,[42] and its high abundance, allowing amplification-free detection.
475 Experimental conditions for Phirst-ID only require base pairing, enabling detection from crude
476 lysate with <30 minutes hands-on time and no nucleic acid extraction or purification, with a total
477 assay time of <8 hours.[18] By comparison, we previously reported a median of 28 hours from

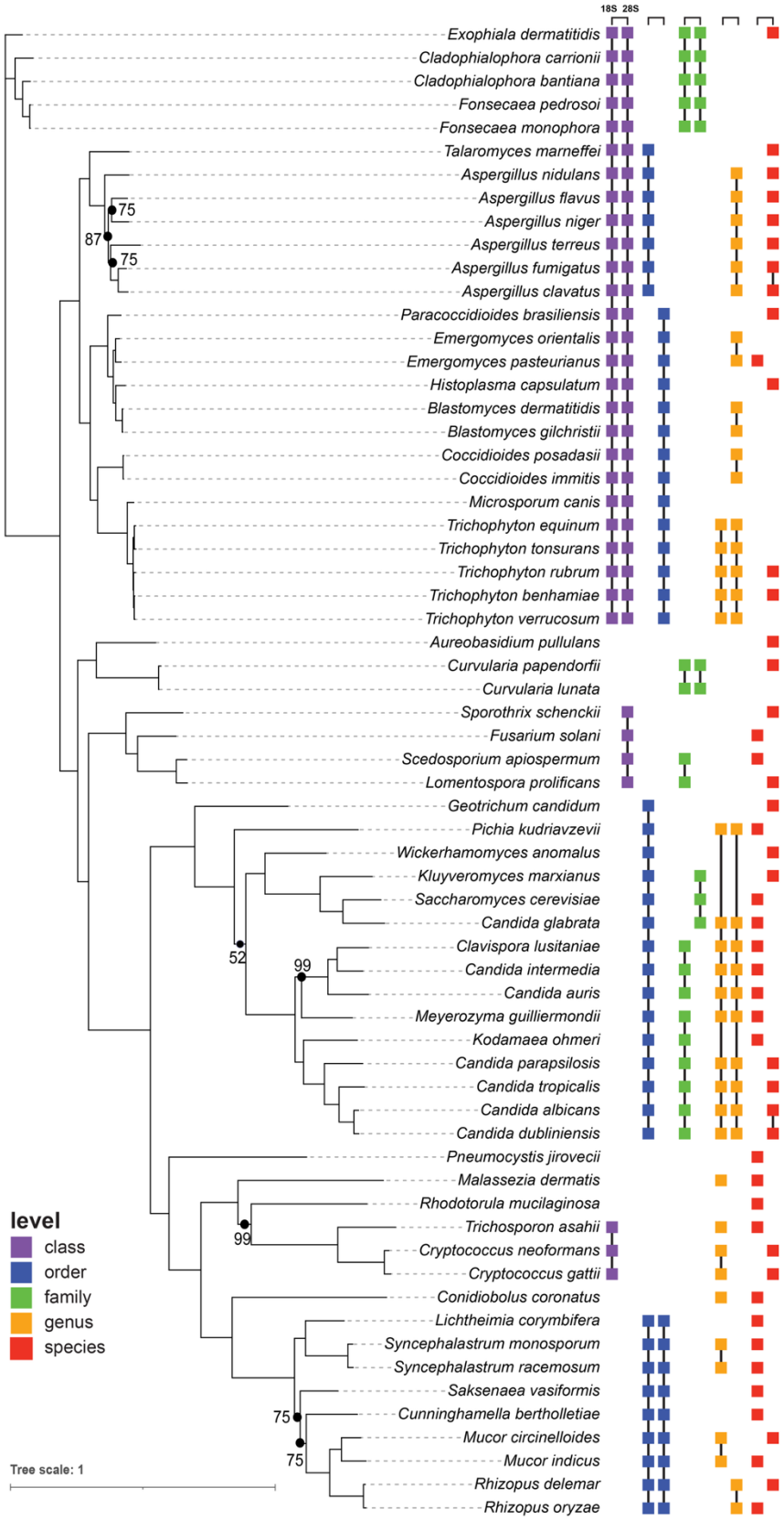
478 culture to identification for *Candida* bloodstream infections;[20] other fungi take longer to identify.
479 Unlike amplification-based methods, the lack of enzymology makes Phirst-ID compatible with
480 chemical RNase inhibitors for robust RNA stabilization. And whereas mass spectrometry-based
481 methods must contend with surface proteomes that can vary across the fungal life cycle [5,7,9,14]
482 rRNA is invariant.
483
484 Prior iterations of Phirst-ID identified organisms based on the closest Pearson correlation of
485 PSRPs of a query organism to that of organisms in a reference panel.[18,20] By contrast, upon
486 extending this approach to diverse fungal pathogens, we found that systematic variations in
487 probe intensity across the Pan-Fungal Phirst-ID panel led to avoidable errors using this
488 relatively simplistic classification scheme (**Fig. 4**). We therefore designed Co-PILOT, a classifier
489 that integrates Pearson correlations of full and partial PSRPs with known taxonomic
490 relationships between species to improve classification accuracy (**Fig. S4**). After iteratively
491 designing Co-PILOT based on a Training Set of 93 isolates across 32 species of common
492 fungal pathogens, Co-PILOT improved identification on an independent Validation Set of 54
493 isolates representing the same species compared with the simpler scheme. Two pairwise
494 misidentifications were made at the species level, each between close phylogenetic relatives:
495 Co-PILOT mistook *C. neoformans* for *C. gattii* and vice-versa, and *S. apiospermum* for *L.*
496 *prolificans* and vice-versa. Both errors appeared to be due to probes designed to distinguish
497 each pair, not performing as expected. In the former case, the distinction between *C.*
498 *neoformans* and *C. gattii* has minimal clinical consequences; though their geographic
499 distribution is somewhat distinct, and their propensity to infect immunocompetent patients may
500 somewhat differ, they both cause similar types of illness, with similar severity and management
501 options.[43] By contrast, the latter case is more clinically critical: though they cause a similar
502 spectrum of illness in similarly immunocompromised patients, the intrinsic antifungal resistance
503 profiles of *S. apiospermum* and *L. prolificans* differ meaningfully, leading to different empiric

504 therapies.[44,45] In order to distinguish these, iterations on the probeset would be required, and if
505 no rRNA-directed probe is able to make this clinically important distinction, mRNA targets may
506 be needed, at a cost of assay sensitivity; or a separate diagnostic test may be required.
507
508 In a pilot study to test additional clinical sample types on our Pan-Fungal Phirst-ID probeset, we
509 tested archived FFPE tissue samples containing fungal forms. This sample type is critical for
510 patient diagnosis when cultures are either not attempted or negative, or to identify species prior
511 to culture, which for fungi can take days to weeks.[1,5,6,37,38] FFPE tissue is optimized for
512 morphological examination, but molecular methods are increasingly used to identify
513 pathogens.[38] Typical workflows involve DNA extraction followed by rDNA-directed PCR with
514 amplicon sequencing, which generally has limited success, ranging from 13-34% in published
515 series,[46–50] typically worse for fungi.[47] Of the 27 FFPE samples we tested, 12 passed our
516 threshold of confidence (>1000 raw max probe counts); of these, Co-PILOT correctly identified
517 5 to the species level, 9 to the genus level, and one more to the closest available phylogenetic
518 match (**Fig. 6**). For 14 samples positive for fungal forms by microscopy, we were unable to
519 identify an organism due to lack of signal. Adding a bead beating step after deparaffinization
520 helped increase signal for some samples, but more work is required to optimize RNA extraction
521 for best results on this clinically important sample type, before testing head-to-head against the
522 current standard of practice.
523
524 This study has several limitations. First, although Pan-Fungal Phirst-ID was designed to detect
525 86 fungal pathogens, we have thus far only been able to rigorously test and independently
526 validate it against multiple isolates of 32 species. Its accuracy in identifying these 32 key
527 species should be well-predicted by its performance on our Validation Set and should improve
528 as the Training Set gets larger with further use. However, its performance on species not
529 included in this Training Set is uncertain. First, we did observe predictable recognition patterns

530 on an additional 18 species (**Fig. S2**), for which we were only able to obtain 1-2 examples for
531 testing. When trained on sufficient examples of new species, Pan-Fungal Phirst-ID should
532 perform well, as long as the 91 probes provide a unique PSRP. However, recognizing when an
533 off-panel species is encountered remains a key challenge. Second, although Pan-Fungal Phirst-
534 ID with Co-PILOT accurately identified >90% of clinical isolates samples to the species level
535 and 98% at the family level in an independent Validation Set, it clearly struggled with the
536 clinically relevant distinction between *S. apiospermum* and *L. prolificans*. Third, several isolates
537 in our Training and Validation Set were only identified to the genus level by the clinical
538 microbiology laboratory, and therefore Phirst-ID accuracy could only be measured to the genus
539 level. With further training, it is possible that species-level distinctions could be made, but
540 genus-level identification for these isolates meets or even exceeds current clinical laboratory
541 reporting standards for these genera (e.g., *Mucor*, *Rhizopus*, and *Cunninghamella* are often
542 reported as order *Mucorales* from clinical laboratories), whereas Pan-Fungal Phirst-ID
543 accurately identified them to the genus level - better than clinical morphology-based
544 assessments in at least one case (**Fig. S7**). Fourth, aside from the pilot on FFPE tissue, we thus
545 far only trained and tested this assay on monocultures. Although *Candida* Phirst-ID was able to
546 recognize mixed cultures as linear combinations of PSRPs from each species,[20] the increased
547 complexity of PSRPs in the Pan-Fungal probeset will likely make this more challenging. Fifth, in
548 this initial work, we rigorously tested this assay primarily on cultured isolates, although the
549 hybridization-based assay in other contexts has worked well despite an excess of off-target
550 nucleic acids (such as human tissue). Still, despite some promising results on a small number of
551 pilot samples, further optimization is needed for deployment on primary clinical samples,
552 including the high-value clinical use case of FFPE tissue. Finally, this assay is currently
553 research-use-only, like its bacterial and *Candida* counterparts. However, this broad-range,
554 sensitive assay could be readily applied to agricultural, environmental or biodefense monitoring.
555 It would be feasible to combine our bacterial probeset[18] with this Pan-Fungal probeset to

556 characterize complex samples or polymicrobial infection.[20] Additional work beyond the scope of
557 our academic laboratory would be required to optimize process and reagents, and ideally
558 automate portions of the workflow, prior to broader scope implementation. Once optimized and
559 scaled, Pan-Fungal Phirst-ID offers a novel approach to fungal identification that combines the
560 lack of enzymology of mass spectrometry with the invariant and high-abundance targets of
561 rRNA-based methods, implementable on either cultured isolates or primary specimens.
562

563 **Acknowledgments:**
564 We thank Dr. Sinem Beyhan and Jacob Durazo from the J. Craig Venter Institute, and Dr.
565 Joseph Kovacs, Dr. Shelly Curran, and Dr. Liang Ma from the NIH, for generously sharing
566 samples with us, and Andrew Grootsky, Aleksandr Kutchma, Richard Boykin (current or former
567 employees of NanoString, now Bruker Spatial Biology, the company that manufactures the RNA
568 detection platform used in this manuscript) for their assistance with our non-standard probe
569 design. This work was supported in part by the National Institutes of Health [grant numbers
570 1R01AI153405 to R.P.B. and C.A.C., and T32 AI007061 to E.A.Y]. The content is solely the
571 responsibility of the authors and does not necessarily represent the official views of the National
572 Institutes of Health. The funders had no role in study design, data collection, and interpretation,
573 or decision to submit this work for publication. R.P.B, E.A.Y and M.L.D-M are co-inventors on
574 subject matter in BI-11335, and R.P.B is a co-inventor on subject matter in BI-11339, filed by
575 the Broad Institute directed to methodology of identifying fungal pathogens using rRNA and
576 disclosure of fungal probes, respectively, as described in this manuscript.
577
578
579
580
581


18S
28S
Exophiala dermatitidis
Cladophialophora carrionii
Cladophialophora bantiana
Fonsecaea pedrosoi
Fonsecaea monophora
Talaromyces marneffei
Aspergillus nidulans
Aspergillus flavus
Aspergillus niger
Aspergillus terreus
Aspergillus fumigatus
Aspergillus clavatus
Paracoccidioides brasiliensis
Emergomyces orientalis
Emergomyces pasteurianus
Histoplasma capsulatum
Blastomyces dermatitidis
Blastomyces gilchristii
Coccidioides posadasii
Coccidioides immitis
Microsporum canis
Trichophyton equinum
Trichophyton tonsurans
Trichophyton rubrum
Trichophyton benhamiae
Trichophyton verrucosum
Aureobasidium pullulans
Curvularia papendorfii
Curvularia lunata
Sporothrix schenckii
Fusarium solani
Scedosporium apiospermum
Lomentospora prolificans
Geotrichum candidum
Pichia kudriavzevii
Wickerhamomyces anomalus
Kluyveromyces marxianus
Saccharomyces cerevisiae
Candida glabrata
Clavispora lusitaniae
Candida intermedia
Candida auris
Meyerozyma guilliermondii
Kodamaea ohmeri
Candida parapsilosis
Candida tropicalis
Candida albicans
Candida dubliniensis
Pneumocystis jirovecii
Malassezia dermatis
Rhodotorula mucilaginosa
Trichosporon asahii
Cryptococcus neoformans
Cryptococcus gattii
Conidiobolus coronatus
Lichtheimia corymbifera
Syncephalastrum monosporum
Syncephalastrum racemosum
Saksenaea vasiformis
Cunninghamella bertholletiae
Mucor circinelloides
Mucor indicus
Rhizopus delemar
Rhizopus oryzae
75
87
75
52
99
99
75
75
level
class
order
family
genus
species
Tree scale: 1

582

583 **Figure 1. Phylogenetic coverage of Pan-Fungal Phirst-ID probeset.** Fungal species included
584 in probeset design, arranged by maximum-likelihood phylogeny inferred using IQ-TREE. To the
585 right of each species, each solo square or set of connected squares denotes a unique probe;
586 black lines connecting squares indicate a single probe designed to target multiple species within
587 a taxonomic group. Colors indicate taxonomic level of the probes (see key at bottom left). For
588 each level, squares at left indicate an 18S rRNA target and at right a 28S rRNA target. Tree
589 branch lengths are in expected substitutions per site.
590
591
592
593
594
595
596

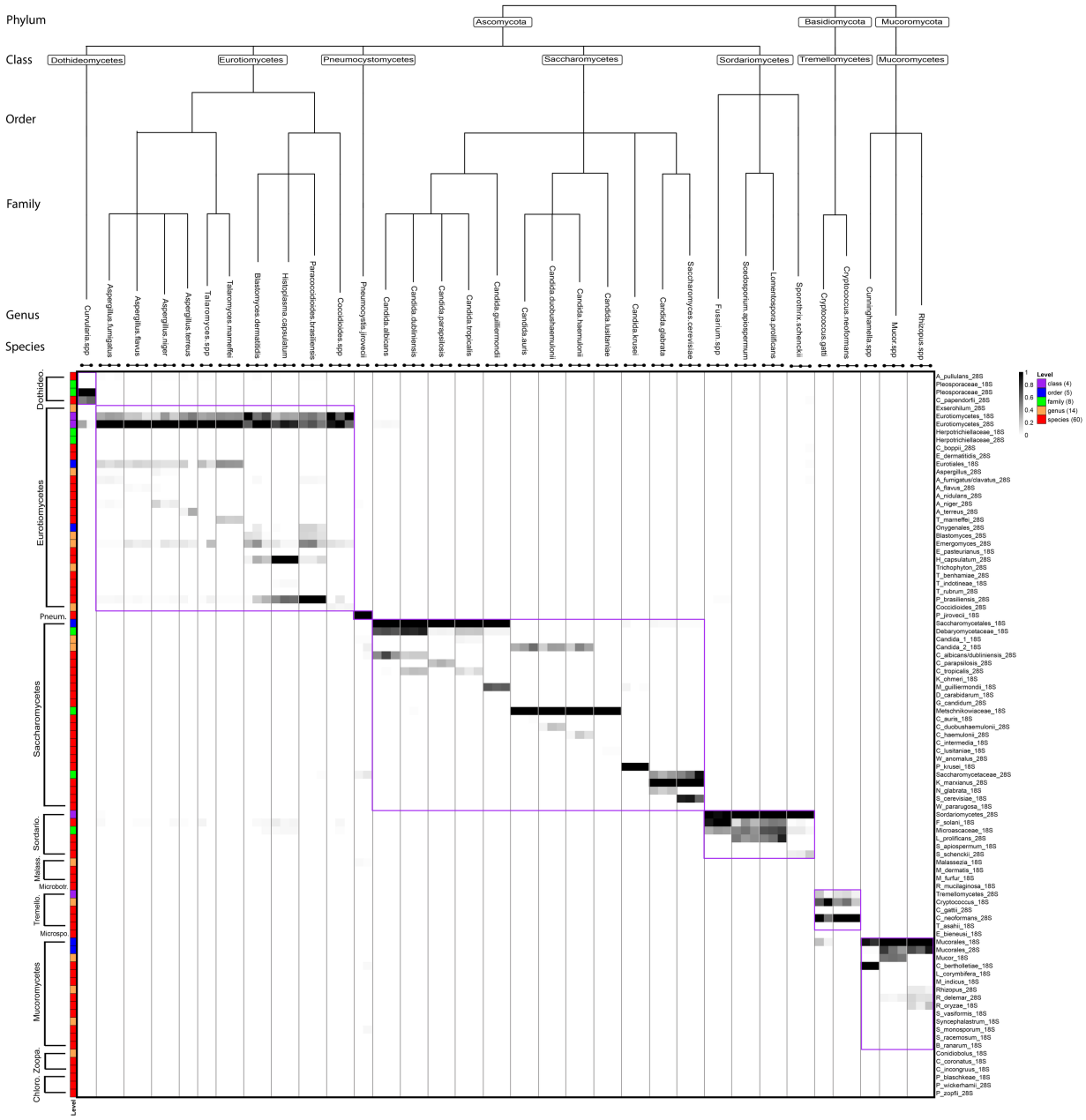


Figure 2. Pan-Fungal Phirst-ID panel generates unique PSRPs for 33 clinically relevant species. Heatmap shows scaled binding intensities of 91 probes (vertical axis) tested against the Training Set of 94 samples from 33 species (horizontal axis). Taxonomic level of each probe is indicated by colored box at left (numbers in parentheses = total probes designed at each level); probe name is listed at right. Each axis is ordered by taxonomic hierarchy, with species displayed in a cladogram. Horizontal black bars with dots indicate unique isolates of each species tested; vertical lines indicate transitions between species. Purple boxes indicate regions where the class of samples and intended probe targets match.

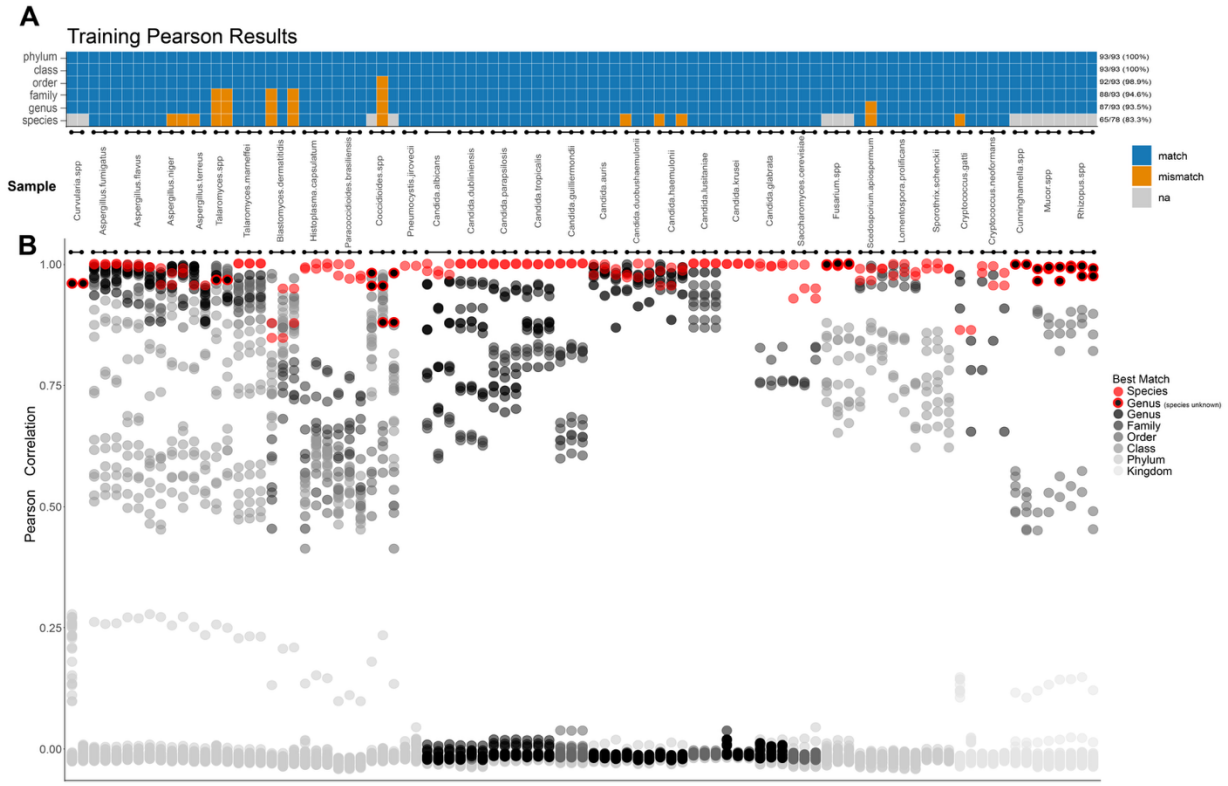


Figure 3. Pan-Fungal Phirst-ID PSRPs identify most species. Leave-one-out analysis of Training Set data showing closest taxonomic matches based on highest Pearson correlation. **(A)** Tile plot indicating accuracy of best non-self matches within the Training Set for identifying each individual isolate at each taxonomic level. **(B)** Dot plot of Pearson correlations of each indicated sample against all other samples in the Training Set, colored by closest-matched taxonomic level of the comparator. When the species of a test strain was unknown, the best possible match is to the genus level, indicated by a gray box at the species level in (A), and by a black circle with red border in (B).

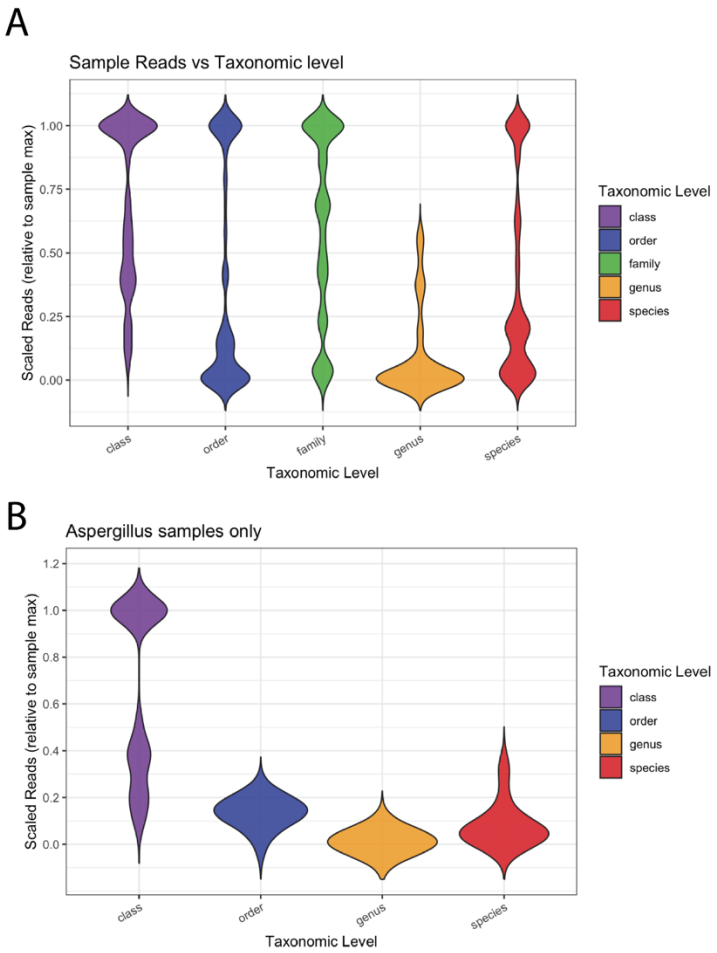


Figure 4. Probes designed for regions conserved at higher taxonomic levels exhibit higher binding signal. Violin plots of normalized signal intensity of probes targeting each taxonomic level of **(A)** all samples in Training Set and **(B)** only *Aspergillus* samples in Training Set. Y-axis for both graphs indicates non-zero probe reactivities normalized to the maximally reactive probe for each sample.

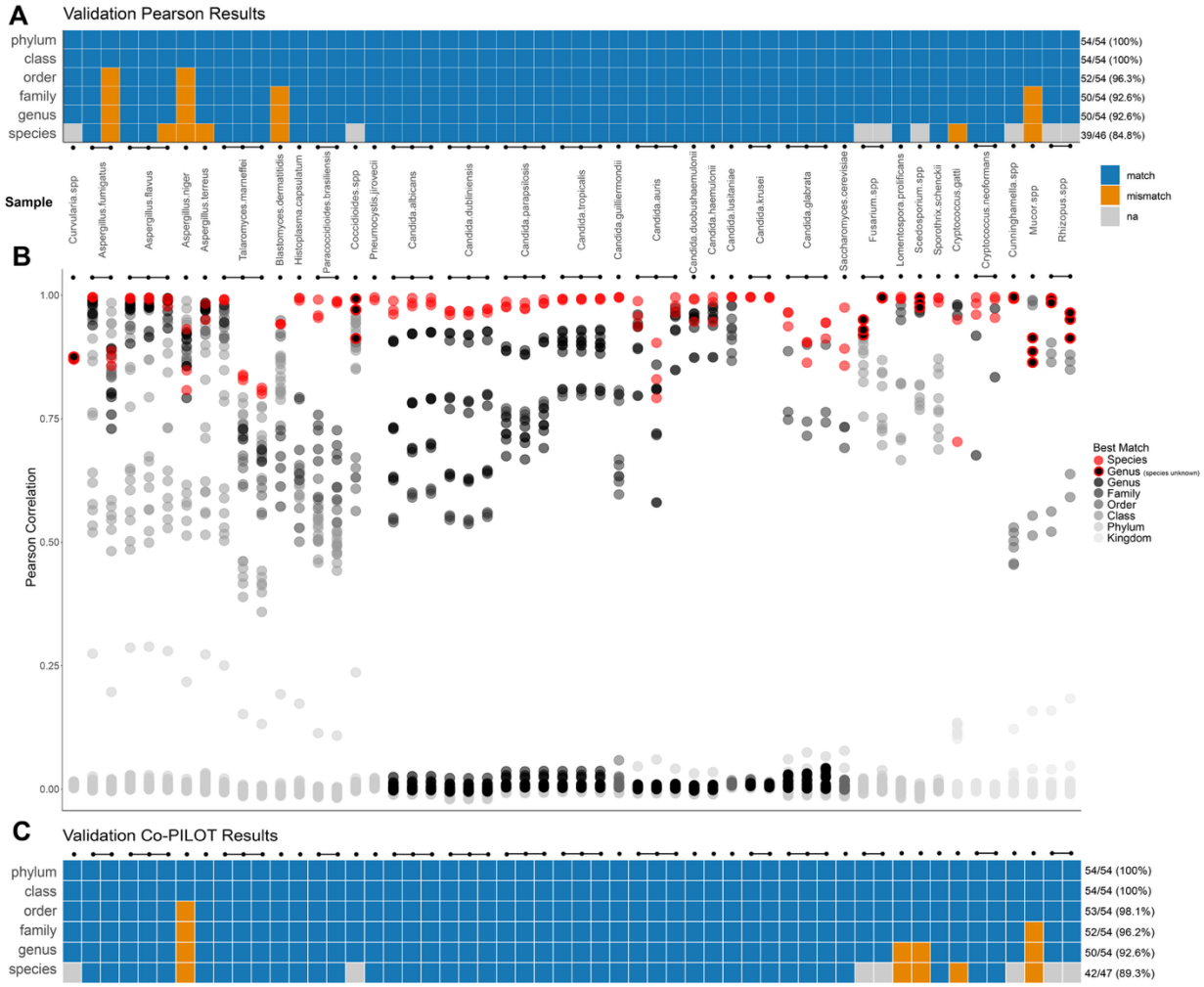


Figure 5: Pan-Fungal Phirst-ID PSRPs identify most species in independent Validation Set. Each sample in the Validation Set was compared by Pearson correlation with that of the entire Training Set. **(A)** Tile plot indicating accuracy of best matches at each taxonomic level based on highest Pearson correlation. **(B)** Dot plot of Pearson correlations of each Validation sample against all samples in the Training Set, colored by closest taxonomic level of the comparator. **(C)** Tile plot shows that Co-PILOT improves accuracy relative to simple Pearson correlations (A) for identifying the best match. When the species of a test strain was unknown, the best possible match is to the genus level, indicated by a gray tile at the species level in (A and C), and by a black circle with red border in (B).

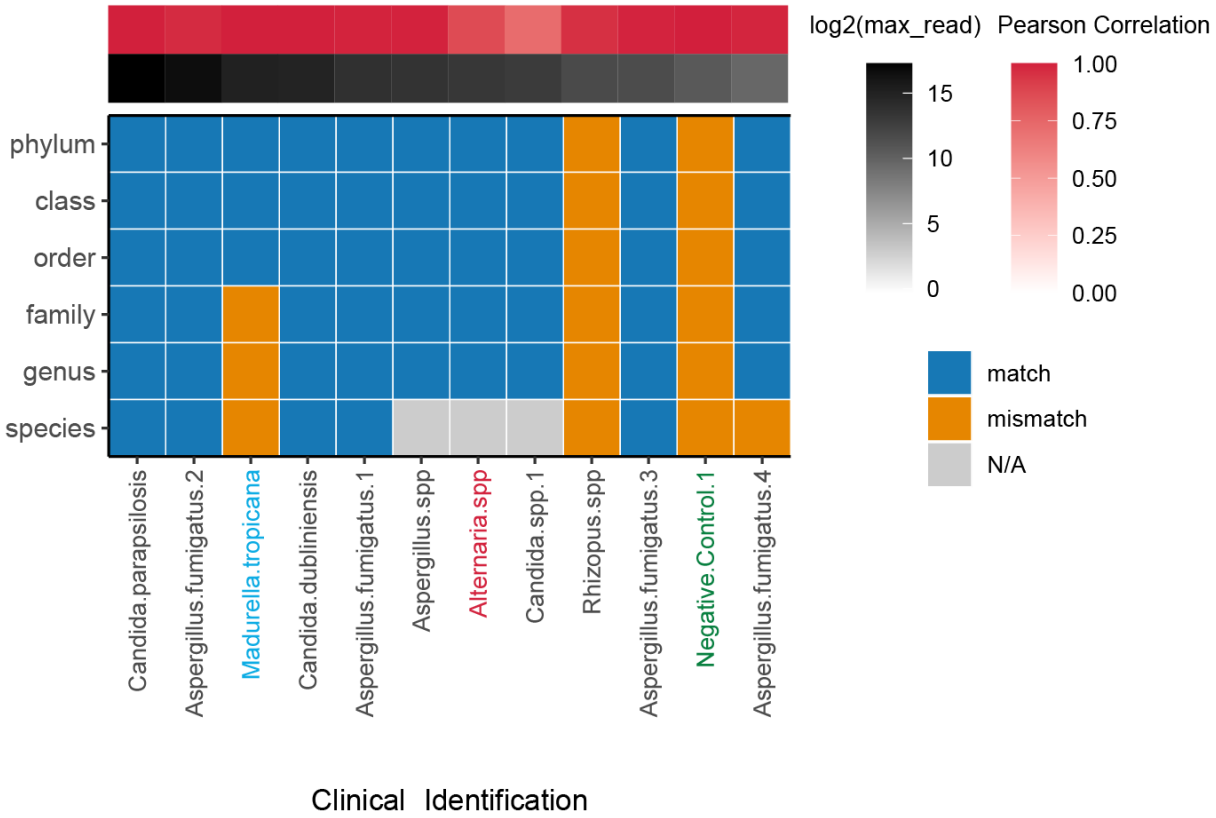


Figure 6. Pilot application of Pan-Fungal Phirst-ID to FFPE samples. Tile plot indicating prediction accuracy at each taxonomic level for the subset of FFPE samples with high enough probe reactivity to pass our threshold for Co-PILOT analysis. Sample order is based the highest probe read count (green heatmap), which was used to set this threshold. Red heatmap shows Pearson correlation of each sample to its best match. Clinical laboratory identifications are listed below (see Table S3c); red text indicates an organism not included in probeset design, green text indicates a negative control FFPE sample (no fungal forms seen), blue text indicates an organism that was neither included in probeset design nor previously tested by Pan-Fungal Phirst-ID (and thus not available to Co-PILOT algorithm for comparison).

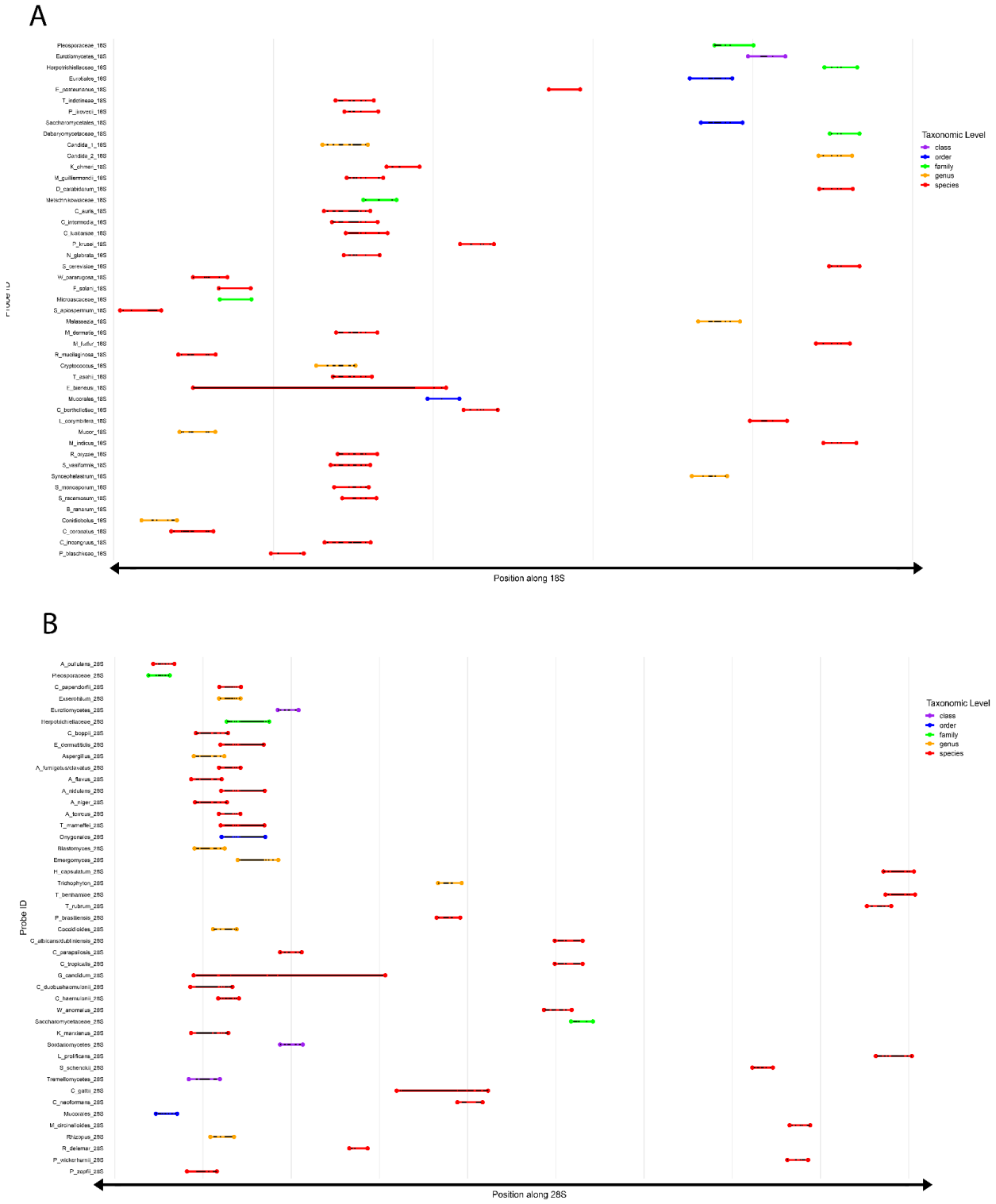


Figure S1. Coverage maps of Pan-Fungal Phirst-ID probe sequences along 18S and 28S rRNA subunits. Probe sequences mapped along target rRNA sequence of (A) 18S and (B) 28S subunits. Color of line indicates taxonomic level specificity of probe. Longer mappings indicate gaps in alignment due to genetic variation in the consensus sequence. Light gray vertical lines

indicate 500 bp distance along the consensus 18/28S axes (including gaps). Probe sequences are available in Supplementary Table S2.

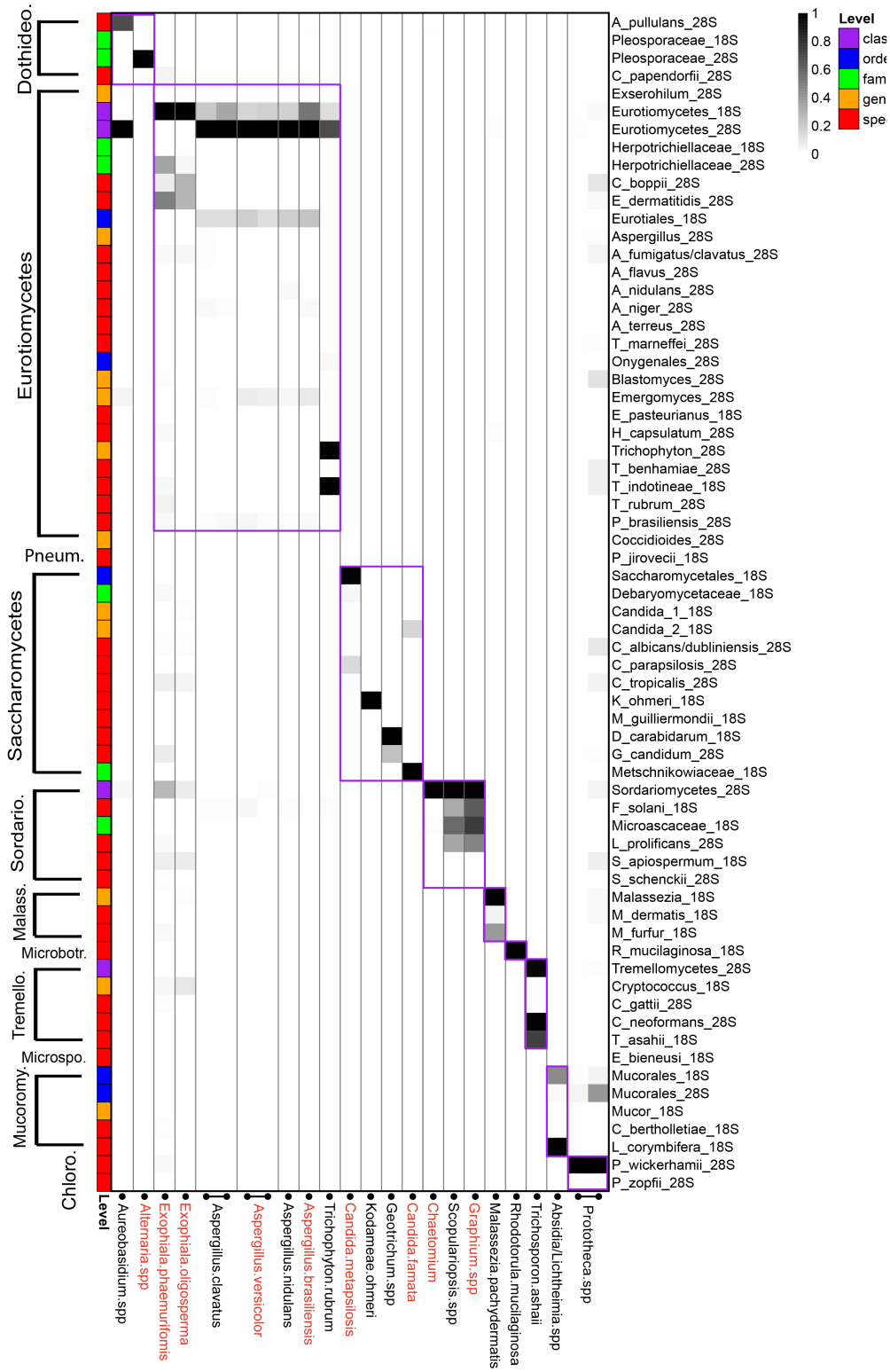

Level
clas
orde
fam
gen
spe
1
0.8
0.6
0.4
0.2
0
A_pullulans_28S
Pleosporaceae_18S
Pleosporaceae_28S
C_papendorfii_28S
Exserohilum_28S
Eurotiomycetes_18S
Eurotiomycetes_28S
Herpotrichiellaceae_18S
Herpotrichiellaceae_28S
C_boppii_28S
E_dermatitidis_28S
Eurotiales_18S
Aspergillus_28S
A_fumigatus/clavatus_28S
A_flavus_28S
A_nidulans_28S
A_niger_28S
A_terreus_28S
T_marneffei_28S
Onygenales_28S
Blastomyces_28S
Emergomyces_28S
E_pasteurianus_18S
H_capsulatum_28S
Trichophyton_28S
T_benhamiae_28S
T_indotineae_18S
T_rubrum_28S
P_brasiliensis_28S
Coccidioides_28S
P_jirovecii_18S
Saccharomycetales_18S
Debaryomycetaceae_18S
Candida_1_18S
Candida_2_18S
C_albicans/dubliniensis_28S
C_parapsilosis_28S
C_tropicalis_28S
K_ohmeri_18S
M_guilliermondii_18S
D_carabidarum_18S
G_candidum_28S
Metschnikowiaceae_18S
Sordariomycetes_28S
F_solani_18S
Microascaceae_18S
L_prolificans_28S
S_apiospermum_18S
S_schenckii_28S
Malassezia_18S
M_dermatis_18S
M_furfur_18S
R_mucilaginosa_18S
Tremellomycetes_28S
Cryptococcus_18S
C_gattii_28S
C_neoformans_28S
T_asahii_18S
E_bieneusi_18S
Mucorales_18S
Mucorales_28S
Mucor_18S
C_bertholletiae_18S
L_corymbifera_18S
P_wickerhamii_28S
P_zopfii_28S
Dothideo.
Eurotiomycetes
Pneum.
Saccharomycetes
Sordario.
Malass.
Microbotr.
Tremello.
Microspo.
Mucoromy.
Chloro.
Level
Aureobasidium.spp
Alternaria.spp
Exophiala.phaeomuriformis
Exophiala.oligosperma
Aspergillus.clavatus
Aspergillus.versicolor
Aspergillus.nidulans
Aspergillus.brasiliensis
Trichophyton.rubrum
Candida.metapsilosis
Kodameae.ohmeri
Geotrichum.spp
Candida.famata
Chaetomium
Scopulariopsis.spp
Graphium.spp
Malassezia.pachydermatis
Rhodotorula.mucilaginosa
Trichosporon.ashaii
Absidia/Lichtheimia.spp
Prototheca.spp

Figure S2 Pan-Fungal Phirst-ID PSRPs against additional samples not included in the Training Set. Heatmap displaying normalized probe binding intensities of 24 additional samples, for which we were only able to obtain 1 or 2 examples and thus did not meet criteria for inclusion in the Training or Validation Sets (because we could not rigorously apply our sample identification algorithms). Purple boxes indicate regions where the class of samples and intended probe targets match. Red text indicates species not included in probeset design but obtained for testing.

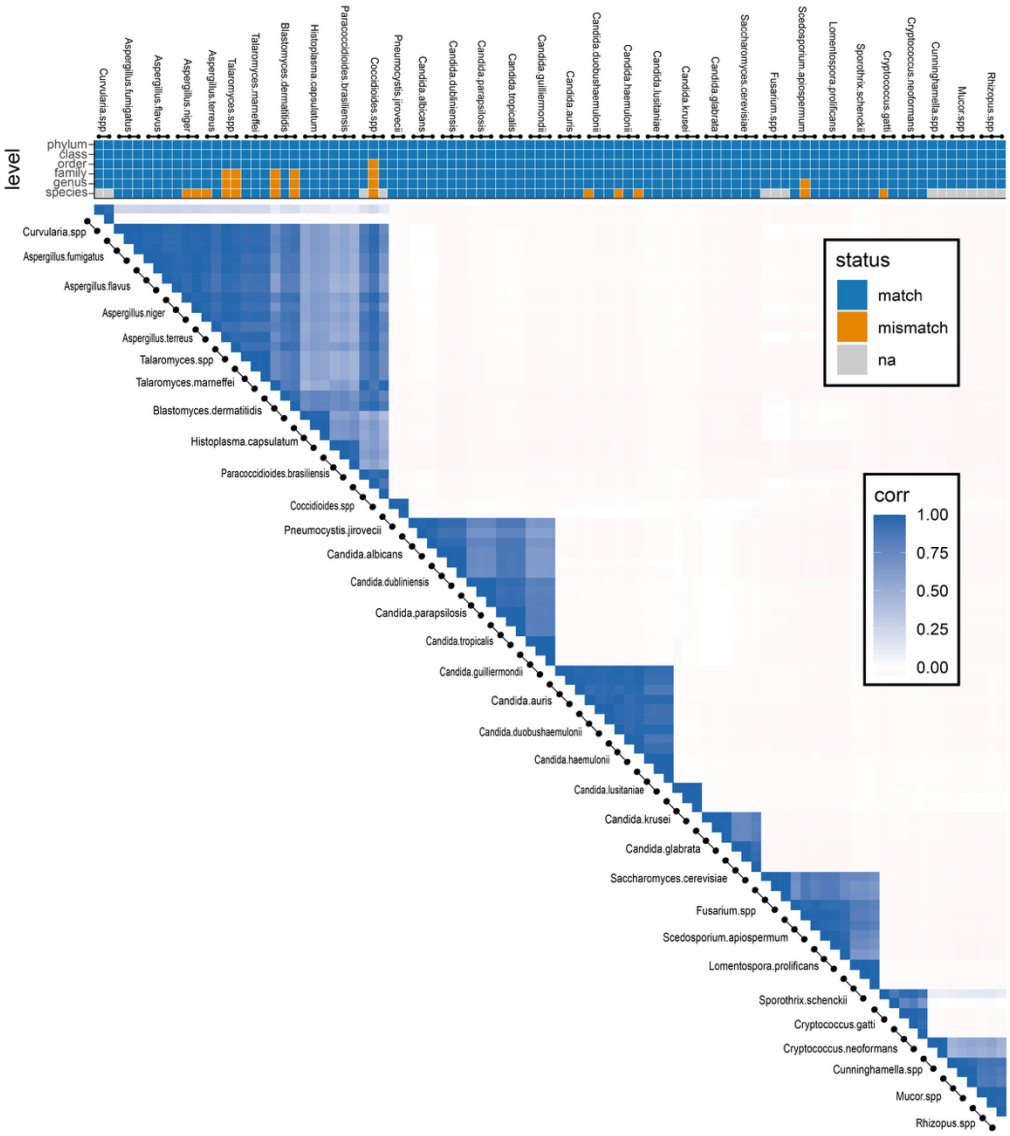


Figure S3. Pearson correlations of Training Set PSRPs are highest between closely related species. Samples with similar taxonomic classification had higher pairwise Pearson correlations than those that were more dissimilar. **(A)** Tile plot indicates prediction accuracy for the single best-matched (highest) non-self-Pearson correlation at each taxonomic level, as in Fig 3a. **(B)** Heatmap of pairwise Pearson correlations, showing that the highest correlations map to members of closely related species. Training Set samples are ordered by taxonomic classification as in Fig 2.

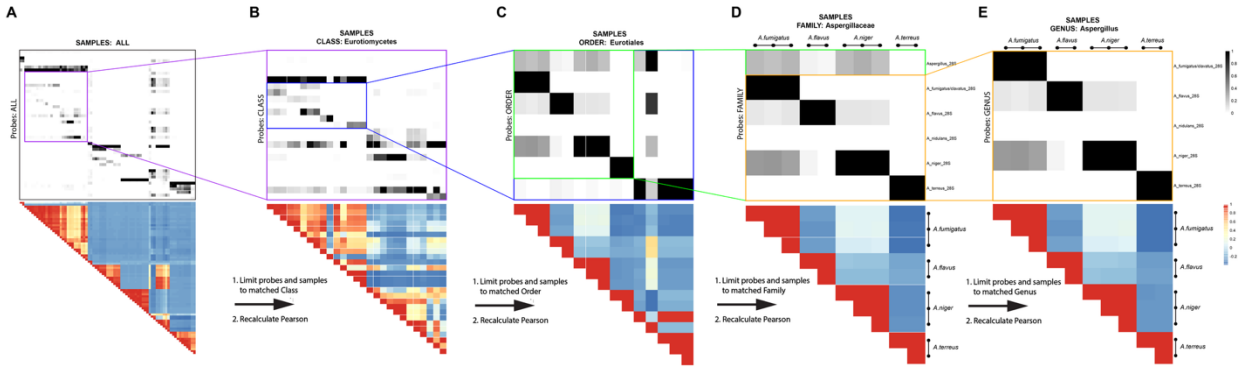


Figure S4. Schematic of our hierarchical Pearson correlation classifier Co-PILOT. Visual representation of how Co-PILOT moves through taxonomic levels, demonstrated on an *Aspergillus fumigatus* sample. Co-PILOT begins by considering all probes and samples **(A)** and chooses the best-matched class based on the class of the single (non-self) sample with the highest Pearson correlation to the query sample. It then restricts further analysis to only samples from that class, and only probes nested within that class (i.e., those designed to target regions matching an order, family, genus, or species within the selected class), indicated by the box (colored by taxonomic level, as in other figures; here, purple box = class); data from all other samples and probes are removed at this stage **(B)**, including probe(s) matching selected class. In subsequent steps, Co-PILOT selects only the subset of samples & probe(s) containing the best-matched order (**B** → **C**, blue box), family (**C** → **D**, green box), and genus (**D** → **E**, orange box). Top panels = read count heatmaps, rescaled by sample at each step to maximize remaining probe; bottom panels = Pearson correlation heatmaps only using selected data at each step.

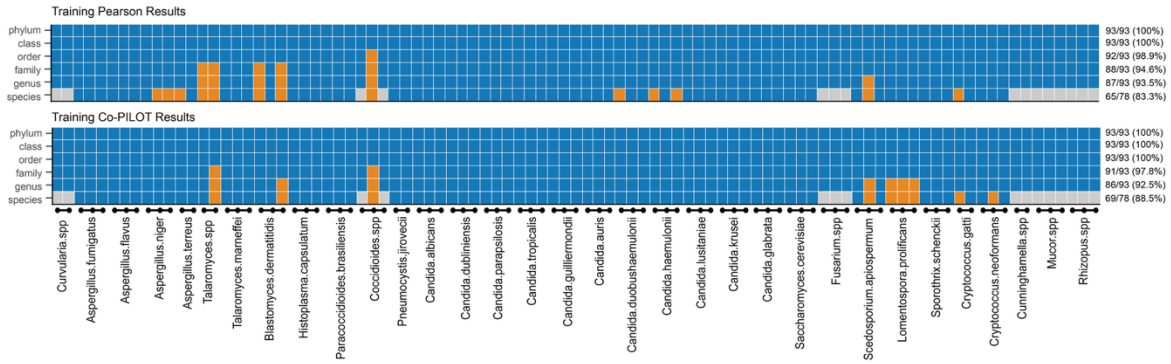


Figure S5. The Co-PILOT classifier improved identification accuracy of the Pan-Fungal Phirst-ID probeset on Training Set. Tile plot indicating prediction accuracy at each taxonomic level, as in Fig 3a. Co-PILOT (bottom panel) improved accuracy relative to simple Pearson correlations (top panel), though it was iteratively designed based on the Training Set, so these results are best viewed as “overtrained” - see main text and Methods for details. By contrast, its performance on the Validation Set (Fig 6) is an independent assessment of Co-PILOT’s accuracy on data it had not seen until its algorithm (Supp Fig S4) was fully designed.

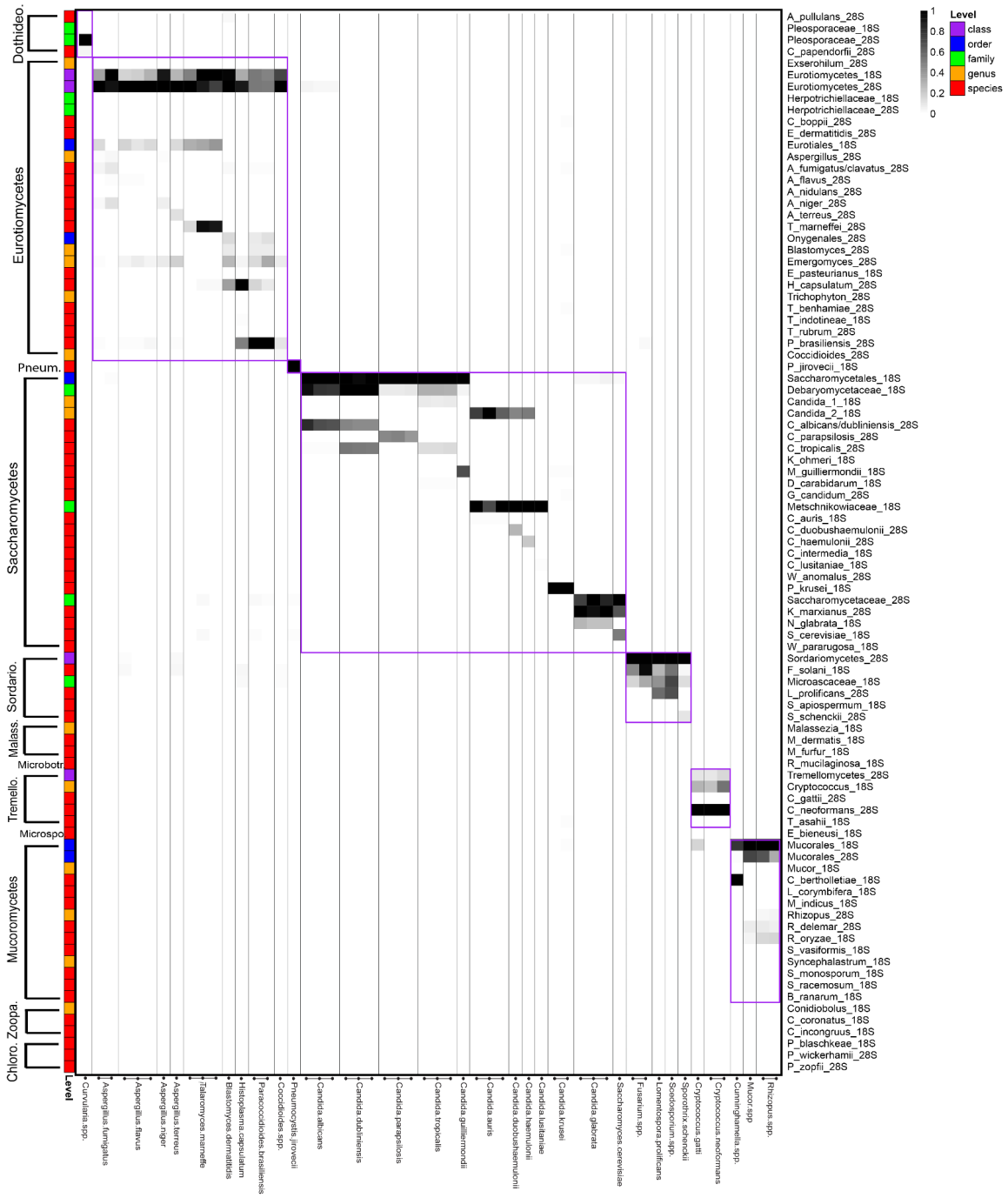


Figure S6. Pan-Fungal Phirst-ID panel generates unique PSRPs for 32 clinically relevant species in our Validation Set. Heatmap of scaled binding intensities of the 54 samples comprising our Validation Set. As in Figure 2, samples and probes are ordered taxonomically, though the cladogram is omitted here. Taxonomic level of each probe is indicated by colored box at left; probe name is listed at right. Purple boxes indicate regions where the class of samples and intended probe targets match.

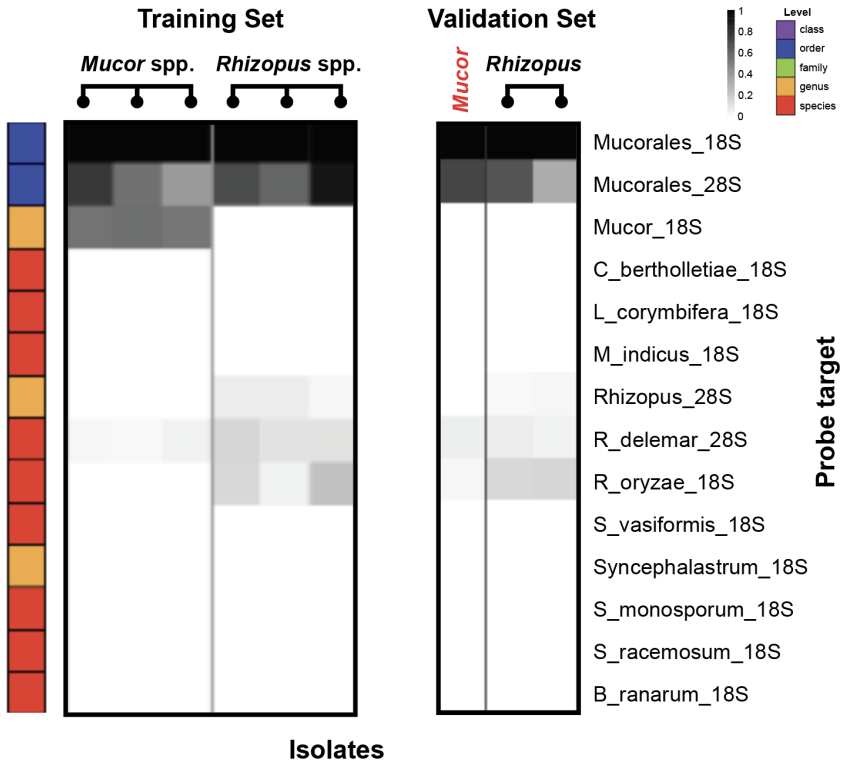


Figure S7 Errant clinical identification of *Mucor* from Validation Set. Pan-Fungal Phirst-ID data from all probes targeting subsets of the *Mucoromycetes* class, shown for all samples from the Training and Validation Sets identified by the clinical microbiology laboratory as either *Mucor* or *Rhizopus*. The red label indicates a Validation sample identified by the clinical laboratory as a *Mucor*, but with a PSRP better resembling a *Rhizopus* (specifically in the lack of Mucor_18S probe reactivity, and perhaps greater R_oryzae_18S probe reactivity). ITS sequencing identified this sample as *Rhizopus microsporus*, more consistent with the Pan-Fungal Phirst-ID prediction than the clinical classification.

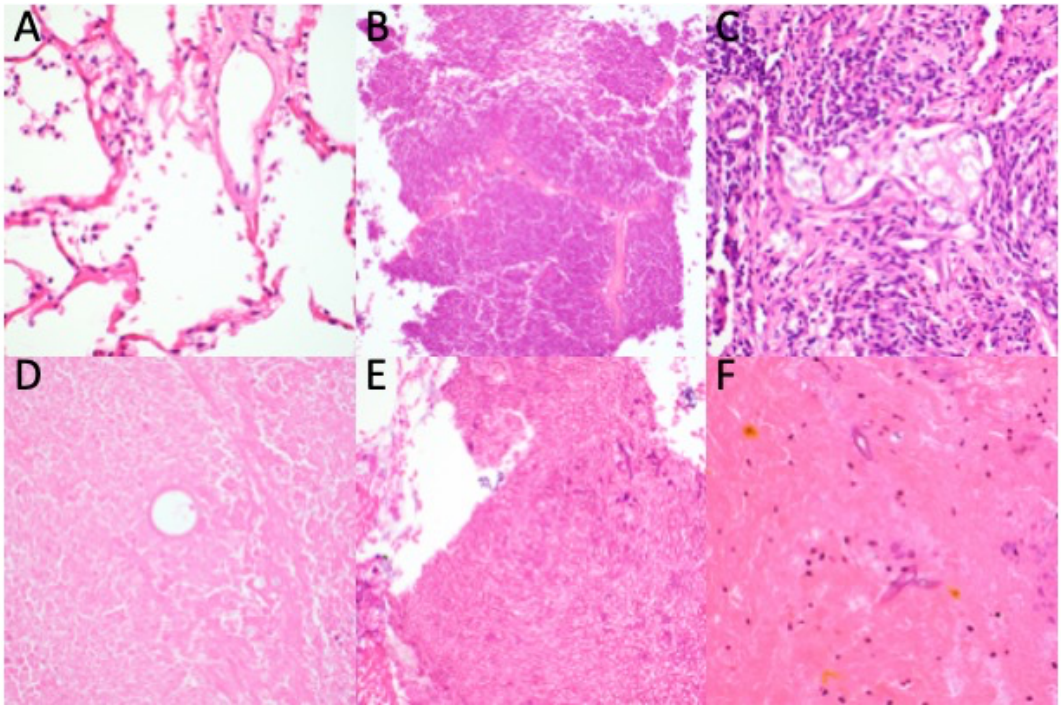


Figure S8. Representative hematoxylin and eosin (H&E)-stained sections from formalin-fixed, paraffin-embedded tissues included in the study. (A) Lung tissue, negative control (Fig S9, right panel, column 11; Supp Table S3d, sample FP002). **(B)** Aortic valve tissue with *Candida parapsilosis* (Fig S9, left panel, column 7; Supp Table S3d, sample FP024. **(C)** Lung tissue with *Cryptococcus neoformans* (Fig S9, right panel, column 8; Supp Table S3d, sample FP011). **(D)** Lung tissue with *Coccidioides* sp (Fig S9, right panel, column 4; Supp Table S3d, sample FP019. **(E)** Sinus tissue with *Aspergillus fumigatus* (Fig S9, left panel, column 5; Supp Table S3d, sample FP028). **(F)** Sinus tissue with *Rhizopus* sp (Fig S9, right panel, column 15; Supp Table S3d, sample FP023). All images were acquired using a 40× objective.

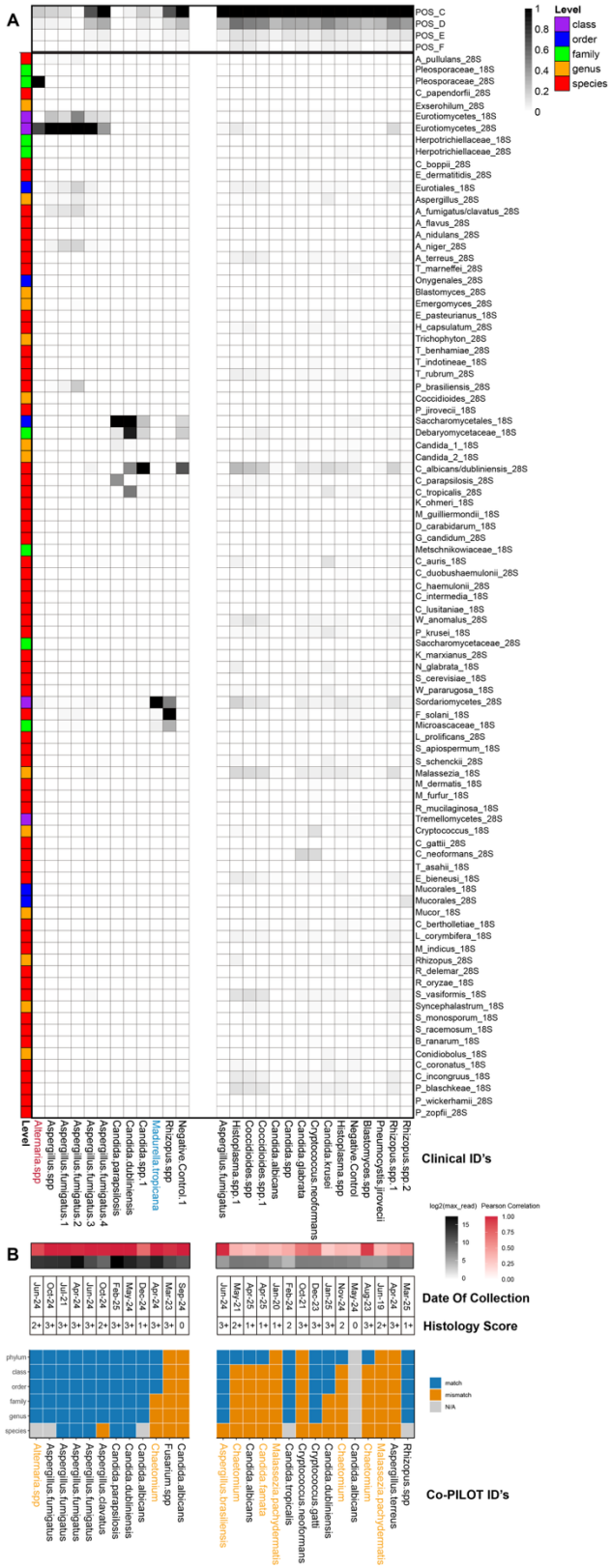

A
POS_C
POS_D
POS_E
POS_F
Level
class
order
family
genus
species
Clinical ID's
B
log2(max_read)
Pearson Correlation
Date Of Collection
Histology Score
phylum
class
order
family
genus
species
match
mismatch
N/A
Co-PILOT ID's

Figure S9: Pilot use of Pan-Fungal Phirst-ID for FFPE samples. (A) Heatmap of scaled probe binding intensities of 27 FFPE samples. For column-scaling to better visually represent total read counts, the four lowest-abundance positive control spike-ins are included for each sample (top 4 rows, above horizontal line). Bottom axis indicates clinical identification provided by the Pathology department (see Supplementary Table S3d). Red text indicates an organism not included in probeset design, blue text indicates an organism neither included in probeset design for nor previously tested by Pan-Fungal Phirst-ID (and thus not available to the Co-PILOT algorithm for comparison). **(B)** Tile plot indicating prediction accuracy of Pan-Fungal Phirst-ID with Co-PILOT at each taxonomic level, as in Fig 3a. Date of sample collected is indicated above and final Co-PILOT identification (best match from Training Set) is listed below the tile plot. 12 samples at left, which exceeded our probe detection threshold, are separated from 17 samples at right, which did not. For visualization purposes, all samples are scaled to the third-most-abundant positive control spike-in (shown above black horizontal line at top). Samples are sorted by taxonomic order of the identification from the clinical microbiology laboratory. Orange text indicates a Co-PILOT prediction of a species not included in probeset design.

826 **References**

827 1. Almeida, F., Rodrigues, M. L., & Coelho, C. (2019). The Still Underestimated Problem of
828 Fungal Diseases Worldwide. *Frontiers in Microbiology*, *10*.
829 https://doi.org/10.3389/fmicb.2019.00214
830 2. Arvanitis, M., Anagnostou, T., Fuchs, B. B., Caliendo, A. M., & Mylonakis, E. (2014).
831 Molecular and Nonmolecular Diagnostic Methods for Invasive Fungal Infections. *Clinical*
832 *Microbiology Reviews*, *27*(3), 490–526. https://doi.org/10.1128/CMR.00091-13
833 3. Zhang, S. X., Babady, N. E., Hanson, K. E., Harrington, A. T., Larkin, P. M. K., Leal, S.
834 M., Luethy, P. M., Martin, I. W., Pancholi, P., Procop, G. W., Riedel, S., Seyedmousavi, S.,
835 Sullivan, K. V., Walsh, T. J., & Lockhart, S. R. (2021). Recognition of Diagnostic Gaps for
836 Laboratory Diagnosis of Fungal Diseases: Expert Opinion from the Fungal Diagnostics
837 Laboratories Consortium (FDLC). *Journal of Clinical Microbiology*, *59*(7), e01784-20.
838 https://doi.org/10.1128/JCM.01784-20
839 4. Badiee, P., & Hashemizadeh, Z. (2014). Opportunistic invasive fungal infections:
840 Diagnosis & clinical management. *The Indian Journal of Medical Research*, *139*(2), 195–
841 204.
842 5. Kozel, T. R., & Wickes, B. (2014). Fungal diagnostics. *Cold Spring Harbor Perspectives*
843 *in Medicine*, *4*(4), a019299. https://doi.org/10.1101/cshperspect.a019299
844 6. Barnes, R. A. (2008). Early diagnosis of fungal infection in immunocompromised
845 patients. *Journal of Antimicrobial Chemotherapy*, *61*(Supplement 1), i3–i6.
846 https://doi.org/10.1093/jac/dkm424
847 7. Freeman Weiss, Z., Leon, A., & Koo, S. (2021). The Evolving Landscape of Fungal
848 Diagnostics, Current and Emerging Microbiological Approaches. *Journal of Fungi*, *7*(2), 127.
849 https://doi.org/10.3390/jof7020127
850 8. Hage, C. A., Carmona, E. M., Epelbaum, O., Evans, S. E., Gabe, L. M., Haydour, Q.,
851 Knox, K. S., Kolls, J. K., Murad, M. H., Wengenack, N. L., & Limper, A. H. (2019).

852 Microbiological Laboratory Testing in the Diagnosis of Fungal Infections in Pulmonary and
853 Critical Care Practice. An Official American Thoracic Society Clinical Practice Guideline.
854 *American Journal of Respiratory and Critical Care Medicine*, *200*(5), 535–550.
855 https://doi.org/10.1164/rccm.201906-1185ST
856 9. Wickes, B. L., & Wiederhold, N. P. (2018). Molecular diagnostics in medical mycology.
857 *Nature Communications*, *9*(1), 5135. https://doi.org/10.1038/s41467-018-07556-5
858 10. Theel, E. S., & Doern, C. D. (2013). Point-Counterpoint: β- D -Glucan Testing Is
859 Important for Diagnosis of Invasive Fungal Infections. *Journal of Clinical Microbiology*,
860 *51*(11), 3478–3483. https://doi.org/10.1128/JCM.01737-13
861 11. Yaman, G., Akyar, I., & Can, S. (2012). Evaluation of the MALDI TOF-MS method for
862 identification of Candida strains isolated from blood cultures. *Diagnostic Microbiology and*
863 *Infectious Disease*, *73*(1), 65–67. https://doi.org/10.1016/j.diagmicrobio.2012.01.013
864 12. Lau, A. F. (2021). Matrix-Assisted Laser Desorption Ionization Time-of-Flight for Fungal
865 Identification. *Clinics in Laboratory Medicine*, *41*(2), 267–283.
866 https://doi.org/10.1016/j.cll.2021.03.006
867 13. Huang, Y., Zhang, M., Zhu, M., Wang, M., Sun, Y., Gu, H., Cao, J., Li, X., Zhang, S.,
868 Wang, J., & Lu, X. (2017). Comparison of two matrix-assisted laser desorption ionization-
869 time of flight mass spectrometry systems for the identification of clinical filamentous fungi.
870 *World Journal of Microbiology and Biotechnology*, *33*(7), 142.
871 https://doi.org/10.1007/s11274-017-2297-3
872 14. Klugherz, I., Kölli, B., Reinmüller, M., Willinger, B., Leitner, E., Steinmetz, I., & Dichtl, K.
873 (2025). Identification of Filamentous Fungi: An Evaluation of Three MALDI - TOF Mass
874 Spectrometry Systems. *Mycoses*, *68*(6), e70078. https://doi.org/10.1111/myc.70078
875 15. Kurtzman, C. P., & Robnett, C. J. (1998). Identification and phylogeny of ascomycetous
876 yeasts from analysis of nuclear large subunit (26S) ribosomal DNA partial sequences.
877 *Antonie van Leeuwenhoek*, *73*(4), 331–371. https://doi.org/10.1023/A:1001761008817

16. Cathy A. Petti, MD, Mary E. Brandt, PhD, Deirdre L. Church, MD, PhD, D(ABMM), Stefan Emler, MD, Keith Simmon, MS, PhD, & Adrian M. Zelazny, PhD, D(ABMM). (2018). *Interpretive Criteria for Identification of Bacteria and Fungi by Targeted DNA Sequencing* (2nd ed.). Clinical and Laboratory Standards Institute (CLSI). https://clsi.org/shop/standards/mm18/

17. Alanio, A., & Bretagne, S. (2014). Difficulties with molecular diagnostic tests for mould and yeast infections: Where do we stand? *Clinical Microbiology and Infection*, *20*, 36–41. https://doi.org/10.1111/1469-0691.12617

18. Bhattacharyya, R. P., Walker, M., Boykin, R., Son, S. S., Liu, J., Hachey, A. C., Ma, P., Wu, L., Choi, K., Cummins, K. C., Benson, M., Skerry, J., Ryu, H., Wong, S. Y., Goldberg, M. B., Han, J., Pierce, V. M., Cosimi, L. A., Shoresh, N., … Hung, D. T. (2019). Rapid identification and phylogenetic classification of diverse bacterial pathogens in a multiplexed hybridization assay targeting ribosomal RNA. *Scientific Reports*, *9*(1), 4516. https://doi.org/10.1038/s41598-019-40792-3

19. Geiss, G. K., Bumgarner, R. E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D. L., Fell, H. P., Ferree, S., George, R. D., Grogan, T., James, J. J., Maysuria, M., Mitton, J. D., Oliveri, P., Osborn, J. L., Peng, T., Ratcliffe, A. L., Webster, P. J., Davidson, E. H., … Dimitrov, K. (2008). Direct multiplexed measurement of gene expression with color-coded probe pairs. *Nature Biotechnology*, *26*(3), 317–325. https://doi.org/10.1038/nbt1385

20. Matzko, M. E., Sephton-Clark, P. C. S., Young, E. L., Jhaveri, T. A., Martinsen, M. A., Mojica, E., Boykin, R., Pierce, V. M., Cuomo, C. A., & Bhattacharyya, R. P. (2022). A novel rRNA hybridization-based approach to rapid, accurate *Candida* identification directly from blood culture. *Medical Mycology*, *60*(9), myac065. https://doi.org/10.1093/mmy/myac065

21. Voorhies, M., Cohen, S., Shea, T. P., Petrus, S., Muñoz, J. F., Poplawski, S., Goldman, W. E., Michael, T. P., Cuomo, C. A., Sil, A., & Beyhan, S. (2022). Chromosome-Level

Genome Assembly of a Human Fungal Pathogen Reveals Synteny among Geographically Distinct Species. *mBio*, *13*(1), e02574-21. https://doi.org/10.1128/mbio.02574-21

22. Cuomo, C. A., Rodriguez-Del Valle, N., Perez-Sanchez, L., Abouelleil, A., Goldberg, J., Young, S., Zeng, Q., & Birren, B. W. (2014). Genome Sequence of the Pathogenic Fungus Sporothrix schenckii (ATCC 58251). *Genome Announcements*, *2*(3), e00446-14. https://doi.org/10.1128/genomeA.00446-14

23. Sephton-Clark, P., Nguyen, T., Hoa, N. T., Ashton, P., van Doorn, H. R., Ly, V. T., Le, T., & Cuomo, C. A. (2023). Impact of pathogen genetics on clinical phenotypes in a population of Talaromyces marneffei from Vietnam. *Genetics*, *224*(4), iyad100. https://doi.org/10.1093/genetics/iyad100

24. Muñoz, J. F., Gauthier, G. M., Desjardins, C. A., Gallo, J. E., Holder, J., Sullivan, T. D., Marty, A. J., Carmen, J. C., Chen, Z., Ding, L., Gujja, S., Magrini, V., Misas, E., Mitreva, M., Priest, M., Saif, S., Whiston, E. A., Young, S., Zeng, Q., … Cuomo, C. A. (2015). The Dynamic Genome and Transcriptome of the Human Fungal Pathogen Blastomyces and Close Relative Emmonsia. *PLoS Genetics*, *11*(10), e1005493. https://doi.org/10.1371/journal.pgen.1005493

25. Muñoz, J. F., Gallo, J. E., Misas, E., Priest, M., Imamovic, A., Young, S., Zeng, Q., Clay, O. K., McEwen, J. G., & Cuomo, C. A. (2014). Genome update of the dimorphic human pathogenic fungi causing paracoccidioidomycosis. *PLoS Neglected Tropical Diseases*, *8*(12), e3348. https://doi.org/10.1371/journal.pntd.0003348

26. Hudson, J., Budvytiene, I., & Banaei, N. (2025). Diagnostic stewardship applied to targeted fungal sequencing. *Journal of Clinical Microbiology*, *63*(10), e00896-25. https://doi.org/10.1128/jcm.00896-25

27. Leaw, S. N., Chang, H. C., Sun, H. F., Barton, R., Bouchara, J.-P., & Chang, T. C. (2006). Identification of Medically Important Yeast Species by Sequence Analysis of the

Internal Transcribed Spacer Regions. *Journal of Clinical Microbiology*, *44*(3), 693–699. https://doi.org/10.1128/JCM.44.3.693-699.2006

28. Abarenkov, K., Nilsson, R. H., Larsson, K.-H., Taylor, A. F. S., May, T. W., Frøslev, T. G., Pawlowska, J., Lindahl, B., Põldmaa, K., Truong, C., Vu, D., Hosoya, T., Niskanen, T., Piirmann, T., Ivanov, F., Zirk, A., Peterson, M., Cheeke, T. E., Ishigami, Y., … Kõljalg, U. (2024). The UNITE database for molecular identification and taxonomic communication of fungi and other eukaryotes: Sequences, taxa and classifications reconsidered. *Nucleic Acids Research*, *52*(D1), D791–D797. https://doi.org/10.1093/nar/gkad1039

29. Poplin, V., Smith, C., Caceres, D. H., Herkert, P. F., Jegede, O., Thompson, G. R., Baddley, J. W., Schwartz, I. S., Kubat, R., Deka, M. A., Toda, M., Lockhart, S. R., Chiller, T., Hagen, F., & Bahr, N. C. (2024). Geographical distribution of the Cryptococcus gattii species complex: A systematic review. *The Lancet. Microbe*, *5*(12), 100921. https://doi.org/10.1016/S2666-5247(24)00161-7

30. Kwon-Chung, K. J., Bennett, J. E., Wickes, B. L., Meyer, W., Cuomo, C. A., Wollenburg, K. R., Bicanic, T. A., Castañeda, E., Chang, Y. C., Chen, J., Cogliati, M., Dromer, F., Ellis, D., Filler, S. G., Fisher, M. C., Harrison, T. S., Holland, S. M., Kohno, S., Kronstad, J. W., … Casadevall, A. (2017). The Case for Adopting the “Species Complex” Nomenclature for the Etiologic Agents of Cryptococcosis. *mSphere*, *2*(1), e00357-16. https://doi.org/10.1128/mSphere.00357-16

31. Lackner, M., De Hoog, G. S., Yang, L., Ferreira Moreno, L., Ahmed, S. A., Andreas, F., Kaltseis, J., Nagl, M., Lass-Flörl, C., Risslegger, B., Rambach, G., Speth, C., Robert, V., Buzina, W., Chen, S., Bouchara, J.-P., Cano-Lira, J. F., Guarro, J., Gené, J., … Liporagi Lopes, L. C. (2022). Proposed nomenclature for Pseudallescheria, Scedosporium and related genera. *Fungal Diversity*, *113*(1), 193–194. https://doi.org/10.1007/s13225-022-00505-0

32. Sangoi, A. R., Rogers, W. M., Longacre, T. A., Montoya, J. G., Baron, E. J., & Banaei, N. (2009). Challenges and Pitfalls of Morphologic Identification of Fungal Infections in Histologic and Cytologic Specimens. *American Journal of Clinical Pathology*, *131*(3), 364–375. https://doi.org/10.1309/AJCP99OOOZSNISCZ

33. Safiia, J., Díaz, M. A., Alshaker, H., Atallah, C. J., Sakr, P., Moshovitis, D. G., Nawlo, A., Franceschi, A. E., Liakos, A., & Koo, S. (2024). Recent Advances in Diagnostic Approaches for Mucormycosis. *Journal of Fungi*, *10*(10), 727. https://doi.org/10.3390/jof10100727

34. Mathieson, W., & Thomas, G. A. (2020). Why Formalin-fixed, Paraffin-embedded Biospecimens Must Be Used in Genomic Medicine: An Evidence-based Review and Conclusion. *The Journal of Histochemistry and Cytochemistry: Official Journal of the Histochemistry Society*, *68*(8), 543–552. https://doi.org/10.1369/0022155420945050

35. Fox, C. H., Johnson, F. B., Whiting, J., & Roller, P. P. (1985). Formaldehyde fixation. *The Journal of Histochemistry and Cytochemistry: Official Journal of the Histochemistry Society*, *33*(8), 845–853. https://doi.org/10.1177/33.8.3894502

36. Guarner, J., & Brandt, M. E. (2011). Histopathologic Diagnosis of Fungal Infections in the 21st Century. *Clinical Microbiology Reviews*, *24*(2), 247–280. https://doi.org/10.1128/CMR.00053-10

37. Cabaret, O., Toussain, G., Abermil, N., Alsamad, I. A., Botterel, F., Costa, J.-M., Papon, J.-F., & Bretagne, S. (2011). Degradation of fungal DNA in formalin-fixed paraffin-embedded sinus fungal balls hampers reliable sequence-based identification of fungi. *Medical Mycology*, *49*(3), 329–332. https://doi.org/10.3109/13693786.2010.525537

38. Gomez, C. A., Budvytiene, I., Zemek, A. J., & Banaei, N. (2017). Performance of Targeted Fungal Sequencing for Culture-Independent Diagnosis of Invasive Fungal Disease. *Clinical Infectious Diseases*, *65*(12), 2035–2041. https://doi.org/10.1093/cid/cix728

39. Lænkholm, A.-V., Jensen, M.-B., Eriksen, J. O., Rasmussen, B. B., Knoop, A. S., Buckingham, W., Ferree, S., Schaper, C., Nielsen, T. O., Haffner, T., Kibøl, T., Møller

979 Talman, M.-L., Bak Jylling, A. M., Tabor, T. P., & Ejlertsen, B. (2018). PAM50 Risk of
980 Recurrence Score Predicts 10-Year Distant Recurrence in a Comprehensive Danish Cohort
981 of Postmenopausal Women Allocated to 5 Years of Endocrine Therapy for Hormone
982 Receptor–Positive Early Breast Cancer. *Journal of Clinical Oncology*, *36*(8), 735–740.
983 https://doi.org/10.1200/JCO.2017.74.6586
984 40. Costa, R. L. B., & Gradishar, W. J. (2018). Do Genomic Assays Provide the Necessary
985 Confidence to De-escalate Adjuvant Therapy? *Journal of Clinical Oncology*, *36*(8), 725–727.
986 https://doi.org/10.1200/JCO.2017.76.9802
987 41. Gnant, M., Sestak, I., Filipits, M., Dowsett, M., Balic, M., Lopez-Knowles, E., Greil, R.,
988 Dubsky, P., Stoeger, H., Rudas, M., Jakesz, R., Ferree, S., Cowens, J. W., Nielsen, T.,
989 Schaper, C., Fesl, C., & Cuzick, J. (2015). Identifying clinically relevant prognostic
990 subgroups of postmenopausal women with node-positive hormone receptor-positive early-
991 stage breast cancer treated with endocrine therapy: A combined analysis of ABCSG-8 and
992 ATAC using the PAM50 risk of recurrence score and intrinsic subtype. *Annals of Oncology*,
993 *26*(8), 1685–1691. https://doi.org/10.1093/annonc/mdv215
994 42. Khot, P. D., Ko, D. L., & Fredricks, D. N. (2009). Sequencing and analysis of fungal
995 rRNA operons for development of broad-range fungal PCR assays. *Applied and*
996 *Environmental Microbiology*, *75*(6), 1559–1565. https://doi.org/10.1128/AEM.02383-08
997 43. Kwon-Chung, K. J., Fraser, J. A., Doering, T. L., Wang, Z., Janbon, G., Idnurm, A., &
998 Bahn, Y.-S. (2014). Cryptococcus neoformans and Cryptococcus gattii, the etiologic agents
999 of cryptococcosis. *Cold Spring Harbor Perspectives in Medicine*, *4*(7), a019760.
1000 https://doi.org/10.1101/cshperspect.a019760
1001 44. Wu, Y., Grossman, N., Totten, M., Memon, W., Fitzgerald, A., Ying, C., & Zhang, S. X.
1002 (2020). Antifungal Susceptibility Profiles and Drug Resistance Mechanisms of Clinical
1003 Lomentospora prolificans Isolates. *Antimicrobial Agents and Chemotherapy*, *64*(11),
1004 e00318-20. https://doi.org/10.1128/AAC.00318-20

1005 45. Chen, S. C.-A., Halliday, C. L., Hoenigl, M., Cornely, O. A., & Meyer, W. (2021).
1006 Scedosporium and Lomentospora Infections: Contemporary Microbiological Tools for the
1007 Diagnosis of Invasive Disease. *Journal of Fungi*, *7*(1), 23.
1008 https://doi.org/10.3390/jof7010023
1009 46. Aggarwal, D., Kanitkar, T., Narouz, M., Azadian, B. S., Moore, L. S. P., & Mughal, N.
1010 (2020). Clinical utility and cost-effectiveness of bacterial 16S rRNA and targeted PCR based
1011 diagnostic testing in a UK microbiology laboratory network. *Scientific Reports*, *10*(1), 7965.
1012 https://doi.org/10.1038/s41598-020-64739-1
1013 47. Kubiak, J., Morgan, A., Kirmaier, A., Arnaout, R., & Riedel, S. (2023). Universal PCR for
1014 bacteria, mycobacteria, and fungi: A 10-year retrospective review of clinical indications and
1015 patient outcomes. *Journal of Clinical Microbiology*, *61*(12), e00952-23.
1016 https://doi.org/10.1128/jcm.00952-23
1017 48. Kerkhoff, A. D., Rutishauser, R. L., Miller, S., & Babik, J. M. (2020). Clinical Utility of
1018 Universal Broad-Range Polymerase Chain Reaction Amplicon Sequencing for Pathogen
1019 Identification: A Retrospective Cohort Study. *Clinical Infectious Diseases*, *71*(6), 1554–1557.
1020 https://doi.org/10.1093/cid/ciz1245
1021 49. Naureckas Li, C., & Nakamura, M. M. (2022). Utility of Broad-Range PCR Sequencing
1022 for Infectious Diseases Clinical Decision Making: A Pediatric Center Experience. *Journal of*
1023 *Clinical Microbiology*, *60*(5), e02437-21. https://doi.org/10.1128/jcm.02437-21
1024 50. Weinert-Stein, K., Cater, M., Cherian, S. S., Azem, R., Khazan, A. E., Desai, A. P.,
1025 Tripp, L., Lim, P. P., Stempak, L. M., & Hojat, L. S. (2024). Clinical Utility of Broad-Range
1026 PCR Testing and Impact on Outcomes in Adults with Suspected Infection. *Antibiotics*,
1027 *13*(12), 1166. https://doi.org/10.3390/antibiotics13121166