Genomic and phenotypic insights into the expanding phylogenetic landscape of the *Cryptococcus* genus

Marco A. Coelho[1,§,*], Márcia David-Palma[1,§], Andrey M. Yurkov[2] and Joseph Heitman[1*]

[1]Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina, USA

[2]Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

[§]These authors contributed equally as first authors.

*Corresponding authors

Emails: marco.dias.coelho@duke.edu and heitm001@duke.edu

Running title: Integrative taxonomy of novel *Cryptococcus* species

Keywords: Fungal speciation, human fungal pathogens, comparative genomics, chromosome evolution, rRNA architecture, digital DNA–DNA hybridization

Abstract

The fungal genus *Cryptococcus* includes several life-threatening human pathogens as well as diverse saprobic species whose genome architecture, ecology, and evolutionary history remain less well characterized. Understanding how some lineages evolved into major pathogens remains a central challenge and may be advanced by comparisons with their nonpathogenic counterparts. Integrative approaches have become essential for delimiting species and reconstructing evolutionary relationships, particularly in lineages with cryptic diversity or extensive chromosomal rearrangements. Here, we formally characterize six *Cryptococcus* species representing distinct evolutionary lineages, comprising both newly discovered and previously recognized but unnamed taxa, through a combination of phylogenomic analyses, divergence metrics, chromosomal comparisons, mating assays, and phenotypic profiling. Among pathogenic taxa, we formally name *Cryptococcus hyracis* sp. nov., corresponding to the previously characterized VGV lineage within the *C. gattii* complex. In parallel, we describe five saprobic, nonpathogenic species isolated from fruit, soil, and bark beetle galleries, spanning four phylogenetic clades. We identify a strong ecological association with bark beetles for *Cryptococcus porticicola* sp. nov., the only newly described nonpathogenic species with multiple sequenced strains from diverse sites. In this species, we detect strain-level chromosomal variation and evidence of sexual reproduction, along with population-level signatures of recombination consistent with ongoing genetic exchange. Across the genus, chromosome-level comparisons reveal extensive structural variation, including species- and strain-specific rearrangements that may restrict gene flow. We also identify multiple instances of chromosome number reduction, often associated with centromere inactivation following interchromosomal rearrangements. Comparative metabolic profiling with Biolog phenotype microarrays reveals clade-level differentiation and distinct substrate preferences, which may reflect metabolic divergence and habitat-specific diversification. Notably, we confirm that thermotolerance is restricted to clinically relevant taxa. These findings refine the species-level taxonomy of *Cryptococcus*, broaden its known genomic and ecological diversity, and strengthen the framework for investigating speciation, adaptation, and the emergence of pathogenicity within the genus.

Introduction

The basidiomycete genus *Cryptococcus* was formally established by Vuillemin in 1901 to describe encapsulated, non-fermentative yeasts that lacked ascospore formation—a key distinction from the genus *Saccharomyces* [1, 2]. This taxonomic revision unified two independent discoveries made in 1894: German physicians Otto Busse and Abraham Buschke had isolated a *Saccharomyces*-like yeast from a tibial lesion in a young woman with a chronic granulomatous infection [3], while Italian microbiologist Francesco Sanfelice recovered a similar organism from peach juice, naming it *Saccharomyces neoformans* based on its distinctive colony morphology [4]. Vuillemin's renaming of the organism as *Cryptococcus neoformans* formalized the genus and established its association with human disease.

Over the following decades, *C. neoformans* became recognized as the causative agent of cryptococcal meningoencephalitis, a life-threatening fungal infection with tropism for the central nervous system, particularly in immunocompromised individuals [5]. Its global burden escalated during the HIV/AIDS pandemic of the 1980s and 1990s, when it became a common AIDS-defining illness [2]. Early classifications relied on phenotypic traits such as capsule morphology and serological markers, which were valuable for identification but lacked resolution to resolve cryptic diversity or deeper evolutionary relationships.

A major advance came in 1975, when K.J. Kwon-Chung successfully induced the sexual state of *C. neoformans* (serotype D, now *C. deneoformans*), leading to the description of the teleomorph *Filobasidiella neoformans* [6]. The observed basidia resembled those of the genus *Filobasidium*, prompting the proposal of a new genus to accommodate the sexual state. However, subsequent revisions to the International Code of Nomenclature of algae, fungi and plants (ICNafp)—which moved toward unifying asexual and sexual names—led to the synonymization of *Filobasidiella* under the older name *Cryptococcus*, which has nomenclatural priority. The discovery of the sexual cycle not only clarified the reproductive biology of this organism but also emphasized the ecological and clinical

relevance of spore production, which facilitates environmental dispersal and likely exposure of mammalian hosts [7].

Today, *Cryptococcus* is best known for its emblematic pathogenic species, particularly *C. neoformans* and *C. gattii*, which together cause over 112,000 deaths annually, primarily among individuals with advanced HIV/AIDS, with the highest burden in sub-Saharan Africa [5, 8]. Owing to their substantial global impact, *C. neoformans* and *C. gattii* have been designated as critical- and medium-priority pathogens, respectively, on the WHO fungal priority list [9]. These two species belong to a broader pathogenic clade presently comprising seven recognized human pathogens: *C. neoformans*, *C. deneoformans*, and five members of the *C. gattii* complex (*C. gattii*, *C. bacillisporus*, *C. deuterogattii*, *C. tetragattii*, and *C. decagattii*) [10]. These taxa differ in virulence, antifungal susceptibility, geographic distribution, and host susceptibility, differences increasingly illuminated through comparative genomics and population-level studies. For instance, *C. deuterogattii* and *C. gattii* sensu stricto frequently cause disease in immunocompetent individuals, the former species being responsible for a well-documented outbreak on Vancouver Island that later expanded to mainland Canada and the Pacific Northwest of the United States [11, 12]. In contrast, *C. neoformans*, *C. deneoformans*, *C. bacillisporus*, *C. tetragattii*, and *C. decagattii* are more often associated with infections in immunocompromised hosts [10, 13, 14]. A recently described, deeply divergent lineage within the *C. gattii* complex (VGV) has not yet been linked to human disease but exhibits thermotolerance and other traits suggestive of pathogenic potential [15].

As taxonomic resolution improves, formally naming new *Cryptococcus* species has become both necessary and important. While concerns have been raised about the added complexity for clinical practice [16], failure to recognize distinct phylogenetic lineages can impede diagnosis and obscure clinical and ecological patterns. Even within the *C. gattii* complex, species differ in virulence, antifungal susceptibility, geographic range, and host susceptibility (i.e., immunocompetent vs. immunocompromised) [10, 13, 14], reinforcing the need for accurate species-level distinctions in both public health and evolutionary research.

Beyond these pathogens, the genus harbors several nonpathogenic taxa. Some have been recognized for decades—including *C. wingfieldii*, *C. amylolentus*, *C. depauperatus*, and *C. luteus* [17-20]—while others, such as *C. floricola* [21], have been more recently described. These species are known exclusively from environmental isolates, typically recovered from soil, plants, or insects, and are presumed saprobic, although *C. depauperatus* and *C. luteus* have been proposed as potential mycoparasites [22-24]. These taxa fall into two distinct clades: one comprising *C. amylolentus*, *C. floricola,* and *C. wingfieldii*, and another represented by *C. depauperatus*. The phylogenetic placement of *C. luteus* remains unresolved due to the lack of genome sequence data. Phylogenetically, the closest known relatives to *Cryptococcus* are found in the genus *Kwoniella*, with both genera forming a well-supported clade within the *Tremellales* (class *Tremellomycetes*) [19, 25].

Building on this foundation, recent work and database mining have revealed additional deeply divergent *Cryptococcus* lineages that remain undescribed [25-27]. High-quality genome assemblies from these lineages have enabled broader comparative analyses of gene content, chromosomal organization, and mating-type locus (*MAT*) structure, offering new insights into the genomic basis of ecological adaptation and potential pathogenicity [25, 27]. Here, we build upon and expand these genomic efforts by integrating newly sequenced genomes with existing data to formally describe six new *Cryptococcus* species representing deeply diverging lineages across multiple clades. Through genome-scale phylogenies, divergence metrics, and chromosomal structural comparisons, we analyze species boundaries, identify evidence of sexual reproduction within a previously uncharacterized bark beetle-associated species, and uncover lineage-specific reductions in chromosome number. We further document phenotypic differentiation across clades, including variation in substrate assimilation profiles, thermotolerance, and melanin production. In doing so, we combine the biological and phylogenetic species concepts to establish a stable, well-supported classification within the genus *Cryptococcus* and provide a broader framework for exploring ecological adaptation and the evolutionary emergence of pathogenicity.

Results

Species phylogeny reveals novel *Cryptococcus* diversity and deep evolutionary divergence

We previously analyzed the genomes of 17 *Cryptococcus* species, among which six represent novel lineages that remained undescribed and unnamed [25, 27]. These candidate species are distributed across multiple major clades of the genus (defined as clades A to D). To further resolve their phylogenetic placement and support formal species recognition, we reconstructed a robust species phylogeny for *Cryptococcus* based on 2,687 single-copy orthologs (SC-OGs) identified across 39 representative strains (**Figure 1A**). Where possible, two strains per species were included to assess monophyly and capture intraspecific heterogeneity. For *C. neoformans*, we further incorporated two representatives from each of the major lineages (VNI, VNBI, VNBII, VNII) to reflect known within-species diversity. Sampling within clade D was also expanded by sequencing, assembling, and annotating the genomes of three additional strains. Two of these (DSM108352 and DSM111204) were sequenced with Oxford Nanopore technology, polished with Illumina reads, and assembled to telomere-to-telomere completeness. A third strain (DSM111205) was sequenced with Illumina only, resulting in a high-quality draft assembly. In addition, the publicly available genome assembly of DSM117830 (GCA_024271845.1) was retrieved from GenBank and annotated in this study.

To infer the species phylogeny, we applied a stringent ortholog selection pipeline (**Figure 1C**), starting with 3,510 SC-OGs and filtering based on minimum protein length, coverage relative to the *C. neoformans* H99 reference proteome, and parsimony informativeness. A final set of 2,687 high-confidence SC-OGs was retained for phylogenetic inference. Both Maximum Likelihood (ML) and coalescent-based analyses (ASTRAL) produced largely congruent topologies with strong support across most nodes (**Figure 1A–B**). These analyses robustly recovered the major lineages of *Cryptococcus* pathogens (clade A)—*C. neoformans*, *C. deneoformans*, and the various *C. gattii* species (sensu lato)—and clearly separated nine saprobic species into three distinct, deeply diverging clades. Several of the newly proposed taxa formed long branches or well-supported monophyletic groups

distinct from previously known species, supporting their recognition as separate evolutionary lineages. Clade B includes *C. depauperatus* and strain CMW60451, representing a new species we designate *Cryptococcus cederbergensis* sp. nov., isolated from a bark beetle (*Lanurgus* sp.) in the Cederberg Mountains of South Africa. Clade C comprises *C. floricola*, *C. wingfieldii*, and *C. amylolentus*, along with a divergent lineage represented by a single strain, now named *Cryptococcus maquisensis* sp. nov., previously isolated from soil in Parque Natural da Arrábida, Portugal [28]. Clade D, sister to all other *Cryptococcus* lineages, contains three well-resolved branches, each representing a new species. One of these includes six closely related strains that form a strongly supported monophyletic lineage, designated *Cryptococcus porticicola* sp. nov., named for its association with bark beetle galleries, from which several strains were isolated in Russia and Sweden. An additional strain was isolated from the gut of a bark beetle larva collected in the Czech Republic. The other two branches represent *Cryptococcus fructiofici* sp. nov., isolated from a fig in Umbria, Italy, and *Cryptococcus jaspeensis* sp. nov., isolated from soil at Pedreira do Jaspe in Parque Natural da Arrábida, Portugal, with the species name referencing its topographic origin. Additional strain details are provided in **Table S1** and in their respective descriptions in the Taxonomy section.

Among the four major clades, the placement of clades B and C relative to clade A was consistent between phylogenetic methods, but this relationship (node N43 in **Figure 1A**; node N32 in **Figure 1B**) exhibited comparatively lower gene and site concordance factors (gCF = 29.9%, sCF = 33.05%; **Figure 1D**), and lower quartet support (35.64%; **Figure 1E**), indicating persistent conflict at this deep node despite topological agreement. While not a strict cutoff, a threshold near 33% is often applied in both frameworks to assess phylogenetic signal. This value reflects the theoretical expectation that, among the three possible unrooted topologies for a given internode, support below this level suggests that alternative relationships are more frequently favored than the main topology.

To assess whether similar patterns of conflict extended to shallower nodes within clade A, we examined a pruned view of the ML and ASTRAL trees, restricted to species in the *Cryptococcus* pathogenic clade (**Figure S1A–B**). The overall topology was identical between methods, but several

internal nodes (N47, N52, N61 in the ML tree) showed relatively low concordance. At these nodes, the main topology was most frequently supported by gene trees (gCF = 26.6–32.7%), yet a substantial proportion of trees were discordant or paraphyletic (gDFP ~46%) (**Figure S1C**). Likewise, site concordance factors (sCF) only modestly favored the main topology. ASTRAL quartet support for the corresponding nodes (N28, N23, N16; **Figure S1D**) consistently identified the primary topology (t1) as best supported, with normalized support ranging from 44.8% to 58.3%, and all nodes having local posterior probabilities (LPP) above 0.9. These results suggest that while gene tree conflict is common at these shallow nodes the dominant signal still supports a stable species tree. Such patterns are frequently associated with incomplete lineage sorting (ILS) and recent speciation events [29, 30].

Genome-wide divergence metrics support species boundaries in the *Cryptococcus gattii* complex and the formal recognition of *Cryptococcus hyracis* sp. nov.

In addition to the newly identified nonpathogenic species, we formally recognize a previously described but unnamed lineage within the *C. gattii* complex. This lineage, referred to as VGV by Farrer et al. [15], was discovered in the Central Zambezian Miombo Woodlands of Zambia and isolated from tree holes and hyrax middens (stratified fecal and urinary deposits formed at communal latrines). In that study, VGV was shown to be genetically distinct from other *C. gattii* lineages (VGI–VGIV), based on both population genetic analyses and its placement as a monophyletic group in genome-wide phylogenies. Species names for all major molecular types (VGI–VGIV and VGVI) had been formally proposed by Hagen et al. in 2015 [10], establishing a precedent for taxonomic recognition within the complex. Because a subsequent publication expressed caution regarding the stability and interpretation of species boundaries in this group [16], the VGV lineage remained unnamed despite clear evidence of genetic divergence. To maintain consistency with the framework proposed by Hagen et al. [10, 13], to reflect our view that recognizing distinct species within the *C. gattii* complex is advantageous for

genomic and ecological clarity, we propose to designate the VGV lineage as *Cryptococcus hyracis* sp. nov., in reference to its frequent association with hyrax-inhabited environments.

In our phylogenetic analyses, *C. hyracis* formed a well-supported clade distinct from its closest relative, *C. tetragattii* (VGIV) (**Figure 1A–B** and **Figure S1A–B**). The two species were resolved as sister groups in the ASTRAL tree with moderate quartet support (58.3%, node N23; **Figure S1B** and **S1D**), and gene concordance fell just below the 33% threshold for inferring a dominant phylogenetic signal (gCF = 32.7%, node N52; **Figure S1A** and **S1C**), again pointing to potential ILS or rapid divergence.

To complement the phylogenetic analyses and further evaluate species boundaries, we calculated genome-wide divergence metrics including average nucleotide identity (ANI), average amino acid identity (AAI), and digital DNA–DNA hybridization (dDDH) across species. These metrics, widely adopted in prokaryotic taxonomy, are increasingly applied to yeast systematics under the taxogenomic framework to quantify genetic relatedness beyond traditional barcode sequences [31]. ANI measures nucleotide-level similarity between homologous genomic regions, with values above ~95–96% typically indicating conspecificity in prokaryotes. Similarly, dDDH, originally developed to approximate wet-lab DNA–DNA hybridization, estimates intergenomic relatedness using BLAST-based comparisons and retains the conventional 70% threshold for species delineation [32]. However, the thresholds developed for prokaryotes are not universally applicable to fungi, and their application in yeast remains underdeveloped. Broader application will likely require integrating additional criteria, such as sexual compatibility and reproductive isolation under the biological species concept. While limited in scope and largely focused on ascomycetous yeasts, initial studies suggest these metrics can be informative when aligned with phylogenetic and phenotypic data [31, 33].

To contextualize these metrics for species delimitation and ensure their appropriate application to basidiomycete fungi, *C. neoformans* was selected as a benchmark. This well-circumscribed species exhibits a documented sexual cycle and reproductive isolation. Genome comparisons included eight *C.*

neoformans strains, with two representatives from each of the major lineages (VNI, VNII, VNBI, and VNBII), providing a reference for intraspecific variation. Although these lineages exhibit substantial population structure and limited recombination between them, they are recognized as a single species [34]. Thus, the lowest similarity observed among these strains provides a conservative upper bound for intraspecific genomic divergence within *Cryptococcus*. The minimum ANI and AAI values observed among *C. neoformans* lineages were 97.7% and 97.4%, respectively, corresponding to divergences of 2.3% and 2.6% (**Figure 2A–C** and **Table S2**). Similarly, to provide a reference point for the lower bound of interspecific divergence, we used the species pair *C. bacillisporus* and *C. decagattii,* which are considered taxonomically distinct despite relatively high genomic similarity. The highest observed ANI and AAI values between strains of these two species were 96.5% and 95.5%, respectively, corresponding to divergences of 3.5% and 4.5%. These values fall below the lowest ANI and AAI observed in any intraspecific comparison in our dataset (**Figure 2C** and **Table S2**), reinforcing the interpretation of *C. bacillisporus* and *C. decagattii* as a species-level boundary case.

Comparatively, pairwise analysis between *C. hyracis* and its sister species *C. tetragattii* yielded ANI values ranging from 95.32% to 95.51% and AAI values from 95.21% to 95.61%, depending on the strain pair. These are lower than the values observed within *C. neoformans* lineages (97.7% ANI and 97.4% AAI) and also lower than those between *C. bacillisporus* and *C. decagattii* (96.3–96.5% ANI; 95.4–95.5% AAI), the most closely related interspecies pair in our dataset (**Figure 2A–B** and **Table S2**). The ANI vs. AAI scatterplot (**Figure 2C**) reinforced this pattern: *C. hyracis*–*C. tetragattii* comparisons fell well below the intraspecies cluster and were among the most similar interspecies comparisons, yet still more divergent than *C. bacillisporus*–*C. decagattii*. The strong correlation between ANI and AAI across all comparisons (Spearman $\rho = 0.8966$, $P < 3.9e{-}237$; Pearson $r = 0.9912$, $P < 1e{-}300$) supports the consistency of these metrics in capturing species-level genomic divergence.

Estimates based on dDDH further supported the distinctiveness of *C. hyracis*. Pairwise dDDH values between *C. hyracis* and *C. tetragattii* ranged from 62.1% to 62.8%, with corresponding probabilities of exceeding the 70% species threshold between 59.1% and 61.3% (**Figure 2D–F** and

Table S2). For comparison, *C. bacillisporus* and *C. decagattii* yielded a dDDH value of approximately 69.3%, just below the conventional cutoff, but with a greater than 75% probability that the true value exceeds the 70% threshold. Nonetheless, these interspecific values remain significantly lower than those observed between *C. neoformans* strains, where dDDH ranges from 79% (probability ≥70% = 89.75%) to 99.1% (probability ≥70% = 98.07) (**Figure 2E** and **Table S2**). This suggests that *C. bacillisporus* and *C. decagattii* lie near the boundary of species-level divergence. In contrast, *C. hyracis* and *C. tetragattii* show a more pronounced level of genomic separation. Together, these results support the recognition of *C. hyracis* as a distinct species, building on previous work by Farrer et al. [15] that first identified it as a divergent lineage within the *C. gattii* complex.

Although all currently available strains of *C. hyracis* are of the *MAT*α mating type and no sexual cycle has been observed thus far, we tested whether *C. hyracis* could engage in mating with strains from other *C. gattii* (sensu lato) species. Our assays showed that *C. hyracis* is capable of forming sexual structures in pairings with *MAT***a** strains of *C. bacillisporus*, *C. decagattii*, *C. tetragattii,* and *C. deuterogattii* (**Figure S2**). Notably, strain MF13 consistently exhibited stronger mating responses than MF34, indicating possible strain-level differences in mating competence, consistent with prior observations of intraspecific variation and their placement in distinct subclades [15]. While the fertility and ploidy status of resulting progeny were not assessed here, these preliminary results suggest that *C. hyracis* retains the ability to mate with other species in the complex. A more systematic investigation of post-zygotic outcomes will be needed to assess reproductive compatibility and isolation across the *C. gattii* complex and to test species boundaries under the biological species concept.

Chromosomal rearrangements and karyotypic diversification in clade A species

Structural variation at the chromosomal level provides a complementary axis of genome divergence and may offer important insights into reproductive isolation and the emergence of species boundaries. Although all pathogenic *Cryptococcus* species (clade A) maintain a karyotype of 14

chromosomes (**Figure 1A**) [27], the extent and pattern of large-scale chromosomal rearrangements across these species have not been systematically compared based on chromosome-level assemblies. To address this, we first performed comparative synteny analyses across clade A species using *C. neoformans* strain 125.91 as a reference. This strain was selected over the more commonly studied H99, which carries a private translocation and therefore does not represent the typical karyotype of the species (**Figure S3A–B**). Interspecies differences are summarized in **Figure 3** with one representative strain per species. **Figure S3** presents additional karyotype comparisons among *C. neoformans* and *C. deneoformans* strains, while **Figure S4** includes all analyzed strains from the *C. gattii* complex to capture both species- and strain-specific variation.

Our comparisons revealed multiple species- and strain-specific structural differences, including large translocations and inversions. Three translocations shared by *C. bacillisporus* and *C. decagattii* appear to involve centromeric regions of chr. pairs 1–2, 4–5, and 12–13 in *C. neoformans* 125.91 (**Figure 3A–B**), consistent with intercentromeric recombination (ICR) as a known mechanism of chromosome reshuffling in *Cryptococcus* [35, 36]. Two of these events (1–2 and 4–5) are older, as they are present in all analyzed *C. gattii* (sensu lato) species, suggesting that they originated in their last common ancestor. In contrast, the third event (chr. pair 12–13) is absent in *C. gattii* (sensu stricto), *C. tetragattii*, *C. hyracis*, and *C. deuterogattii*, indicating that it occurred later, in the lineage leading to *C. bacillisporus* and *C. decagattii*.

Additional translocations were identified among *C. gattii* lineages (**Figure S4**). For example, chrs. 1 and 11 in *C. deuterogattii* appear to derive from a translocation between corresponding chrs. 3 and 9 of *C. gattii* WM276, which are conserved in related species and likely represent the ancestral state (anc3 and anc9) (**Figure 3A** and **Figure S4B**, event a). Two further rearrangements involving anc12–anc13 and anc8–anc10 define the common ancestor of *C. bacillisporus* and *C. decagattii* (**Figure S4B**, events b and c). In *C. decagattii*, one of the resulting chromosomes (corresponding to extant *C. bacillisporus* chr. 5) underwent a subsequent translocation with anc7, giving rise to extant chrs. 4 and 9 in *C. decagattii* (**Figure S4B**, event d). We also detected two strain-specific translocations:

one in *C. decagattii* CBS11687 (between anc1 and anc11; **Figure S4B**, event e) and another in *C. tetragattii* CBS11718 (between chrs. anc6 and anc13; **Figure S4B**, event f). As a result, *C. bacillisporus* differs from *C. decagattii* by one fixed translocation (relative to strain 7685027) or two (relative to strain CBS11687), which further impose postzygotic reproductive barriers.

In contrast, *C. hyracis* strain MF34 and *C. tetragattii* IND107 share a largely collinear karyotype, differing only by a single inversion on chr. 2. Similarly, *C. neoformans* 125.91 and *C. deneoformans* NIH433 differ only by two large inversions (on chrs. 1 and 9; **Figure 3**), with otherwise conserved chromosomal structures. Although these species pairs show limited karyotypic divergence, their clear sequence-level divergence, particularly in the case of *C. neoformans* and *C. deneoformans*, supports their recognition as distinct species, suggesting that large-scale structural changes may not always accompany speciation.

Together, these results reveal that karyotypic diversification across clade A involves a mixture of lineage-specific rearrangements and ongoing structural changes at the strain level. While fixed chromosomal rearrangements likely reinforce reproductive isolation in some lineages—such as between *C. bacillisporus* and *C. decagattii*—their absence in others, such as *C. hyracis* and *C. tetragattii*, underscores a complex and potentially asynchronous relationship between genome structural variation and species divergence in *Cryptococcus*.

Intraspecific variation and evidence of recombination in *Cryptococcus porticicola*

Among members of newly described clade D, *Cryptococcus porticicola* is the only species for which multiple strains are currently available and have been sequenced, enabling analyses of intraspecific genome variation, chromosomal structural dynamics, and reproductive potential. To explore these aspects, we first conducted pairwise genome comparisons among the six sequenced strains revealing high overall similarity, with ANI values ranging from 99.29% to 99.94% and dDDH

values from 94.6% to 97.2% (**Figure 2E** and **Table S2**), confirming that all isolates belong to the same species despite some geographic separation.

Next, we conducted mating assays between strain pairs predicted to be compatible based on sequence differences at both the pheromone/receptor (*STE3*) and homeodomain (*HD*) mating-type loci (**Figure 4A**). These two unlinked *MAT* loci were previously characterized and shown to follow a tetrapolar configuration in *C. porticicola* and other clade D species [25]. Among the tested combinations (**Figure S5**), only the cross between strains DSM108352 (*a1b3*) and NCYC1536 (*a2b2*) sporadically produced visible hyphae and putative sexual structures embedded in the agar, resembling basidia (**Figure 4B–C**). These structures emerged only after extended incubation (~1 month) and were not consistently observed across replicates. While their identity could not be definitively confirmed, their morphology is consistent with sexual development observed in related *Kwoniella* species [37-39]. These structures were observed on V8 medium pH 5.0, whereas growth on MS medium led to hyphal formation but without basidial-like clusters (**Figure S5D**). Notably, some hyphal segments appeared to contain unfused clamp connections (**Figure 4C**), potentially indicating a monokaryotic mycelium.

To further investigate genetic diversity and potential recombination in *C. porticicola*, we analyzed genome-wide single nucleotide polymorphisms (SNPs) across the six available strains. High-confidence SNPs were identified with Snippy pipeline using DSM108351 as the reference, generating a core genome SNP matrix. A neighbor-net network constructed from this alignment revealed reticulate relationships, suggesting conflicting phylogenetic signals and potential recombination events (**Figure 4D**). This interpretation is further supported by the Pairwise Homoplasy Index (PHI) test implemented in SplitsTree, which returned a highly significant result ($P < 1e\text{-}6$), suggesting that the observed pattern of site incompatibilities is unlikely to arise under a strictly clonal (non-recombining) model. Principal component analysis (PCA) also indicated genetic differentiation among strains, with DSM108352 and NCYC1536—the only cross producing putative sexual structures—showing the greatest separation (**Figure 4E**).

To assess patterns of historical recombination more directly, we next examined linkage disequilibrium (LD) and allelic compatibility among SNPs. LD decay analysis, conducted after applying a minor allele frequency (MAF) threshold of 0.167 to reduce noise from rare variants, revealed a clear decline in r^2 values with increasing physical distance between SNPs, consistent with recombination progressively breaking down associations between alleles (**Figure 4F**). Based on the smoothed decay curve, r^2 dropped to 50% of its initial maximum (LD50) at approximately 13.9 kb. This signal was reinforced by the four-gamete test (FGT), which showed an increasing proportion of SNP pairs violating the test at greater genomic distances. Such violations are expected when recombination occurs, as the presence of all four possible haplotypes between two biallelic SNPs is unlikely to arise without it, although rare recurrent or back mutations (homoplasy) can also contribute. Taken together, and with the caveat of limited sample size, these findings support the hypothesis that recombination occurs in *C. porticicola*, consistent with the morphological features observed in mating assays that resemble, but do not definitively confirm, sexual development.

Finally, to assess whether chromosomal structural variation parallels the sequence-level diversity observed in *C. porticicola*, we compared chromosome-scale assemblies from four strains spanning the phylogenetic breadth of the species (**Figure 4A** and **4H**). All strains possess 9 chromosomes, indicating a conserved karyotype across the species. Synteny comparisons, using DSM108351 as the reference, revealed a largely conserved chromosomal architecture between DSM108351 and NCYC1536, which are more closely related. Nevertheless, several intraspecific rearrangements were identified (**Figure 4H**). Two reciprocal translocations are shared between DSM108352 and DSM111204, suggesting these events arose in their common ancestor. Additionally, DSM111204 harbors a third, strain-specific rearrangement with centromere-proximal breakpoints, consistent with an intercentromeric recombination (ICR) event (**Figure 4I**). These results demonstrate that *C. porticicola* harbors detectable chromosomal structural variation among strains, including translocations that may have functional consequences during meiosis. Notably, the rearrangements observed between DSM108352 and NCYC1536 (the only pair producing putative sexual structures)

383 could contribute to postzygotic isolation through segregation defects, potentially explaining the rarity
384 and incompleteness of sexual development. These findings suggest that *C. porticicola* may comprise
385 distinct populations (e.g. geographic or host-related) with established genomic and structural
386 divergence.

387

388 **Chromosome number reduction and centromere dynamics in clade D species**

389 The chromosomal variation identified within *C. porticicola* prompted us to investigate broader
390 trends in karyotype evolution across clade D. All three newly described species in this clade exhibit
391 marked reductions in chromosome number relative to *Cryptococcus* species of clades A and C, which
392 maintain the ancestral 14-chromosome karyotype inferred for the genus [27]. Specifically, *C. porticicola*
393 has 9 chromosomes, while *C. fructiofici* and *C. jaspeensis* have only 8 and 7 chromosomes, respectively
394 (**Figure 5A**).

395 To elucidate the genomic changes underlying these reductions, we compared chromosome-
396 scale assemblies using *C. porticicola* DSM108351 as a reference and mapped centromere positions
397 (defined as long ORF-free regions enriched in LTR retrotransposons and 5-methylcytosine (5mCG)
398 DNA methylation [35, 40-43]) across all available strains of clade D. In *C. jaspeensis*, the region syntenic
399 to *CEN7* of *C. porticicola* and *C. fructiofici* lacks hallmark centromeric features and likely no longer
400 functions as a centromere (**Figure 5B** and **Figure S6F**). Similarly, both *C. fructiofici* and *C. jaspeensis*
401 contain additional inactivated centromeres (*iCENs*) corresponding to *CEN9* and *CEN8* of *C. porticicola*.
402 These *iCENs* retain conserved flanking genes but are depleted of LTR retrotransposons and show only
403 residual 5mCG methylation, consistent with erosion of centromeric identity while preserving local gene
404 order (**Figure 5D**, panel III, and **Figure S6G**).

405 Notably, the number of inactivated centromeres (*iCENs*) in *C. fructiofici* (two) and *C. jaspeensis*
406 (three) exceeds the reduction in chromosome number relative to *C. porticicola*, potentially suggesting
407 that a new centromere formed in their common ancestor. Indeed, the chromosomes that formerly carried

408 *CEN9* (chr. 4 in *C. fructiofici*, chr. 5 in *C. jaspeensis*) now harbor an apparently active centromere at a
409 novel location (**Figure 5C** and **5D**, panel II). This putative centromere (*nCEN*) is conserved in both *C.*
410 *fructiofici* and *C. jaspeensis* and exhibits hallmark centromeric features, including transposon
411 enrichment and 5mCG methylation. Interestingly, the syntenic region in *C. porticicola* displays spatially
412 restricted 5mCG enrichment and a small number of transposable elements but does not coincide with
413 any predicted centromere. This raises two possible scenarios: either the region was newly co-opted as
414 a neocentromere in the ancestor of *C. fructiofici* and *C. jaspeensis*, or it reflects an ancestral centromere
415 that was retained in these species but lost or inactivated in *C. porticicola*. Importantly, gene order is
416 preserved across these chromosomes, suggesting that centromere repositioning, if it occurred, did so
417 without major local rearrangement. Together, these findings suggest that chromosome number
418 reduction in clade D species, potentially contributing to species diversification, is linked to structural
419 rearrangements followed by centromere inactivation, consistent with a broader model supported by
420 observations in *C. depauperatus* [27] and several other fungi, in which centromere identity erodes
421 through the progressive loss or excision of repeat-rich DNA [44-46].

422

423 **Variation in rRNA gene organization across *Cryptococcus* species**

424 In addition to chromosome number changes, variation in ribosomal RNA (rRNA) gene
425 organization represents another axis of genome structural divergence in *Cryptococcus* [27]. To explore
426 this in more detail, we applied a unified pipeline that integrates rRNA gene prediction, clustering
427 detection, and statistical enrichment tests (see Material and Methods) to survey the chromosomal
428 arrangement of 5S, 5.8S, 18S, and 28S rRNA genes across 20 chromosome-level genome assemblies
429 representing all major clades and species (**Figure 6** and **Figure S7**).

430 Fully clustered rDNA arrays (defined as tandem repeats of the canonical 28S–5.8S–18S–5S
431 gene units) were observed exclusively in pathogenic species from clade A. In contrast, other species
432 exhibited unclustered architectures, with the 5S gene physically separated from the 28S–5.8S–18S unit

433 and exhibiting lineage-specific variation in chromosomal distribution. Among clade C species, the 5S
434 genes remained significantly enriched on a single chromosome (chi-squared test, $P < 0.0001$ for all
435 species; **Figure 6B**, **Figure S7**, and **Table 3**), suggesting that despite dispersion, the ancestral
436 chromosomal context of the rDNA array may be partly retained. For instance, in *C. floricola*, over 50%
437 of all 5S genes are located on a single chromosome (standardized residual > +11; **Table S3**), which
438 also harbors the canonical rDNA array. By contrast, in other clade C species, the 28S–18S–5.8S array
439 appears relocated to alternate chromosomal ends, consistent with interchromosomal mobility potentially
440 mediated by subtelomeric recombination (**Figure S7**).

441 Clade D species also display unclustered rRNA gene organization but with differing 5S patterns.
442 In *C. porticicola*, 5S genes are distributed relatively evenly across chromosomes (chi-squared test, $P =$
443 0.87; **Table S3**). Conversely, *C. fructiofici* and *C. jaspeensis* exhibit significant 5S gene enrichment on
444 specific chromosomes (chi-squared test, $P = 0.0006$ and 0.011, respectively) coupled with a strong
445 subtelomeric bias (Fisher's exact, $P < 1.12e^{-37}$ and $P < 3.03e^{-15}$, respectively; **Table S3**). In both
446 species, the majority of 5S genes reside near chromosome ends, suggesting relocation to subtelomeric
447 regions, possibly through recombination or transposition. Together, these results indicate that 5S rRNA
448 gene organization varies substantially across *Cryptococcus*, with clustered arrays retained in
449 pathogenic species and more dispersed patterns in nonpathogenic lineages, suggesting lineage-
450 specific shifts in rDNA architecture alongside broader chromosomal evolution.

451

452 **Phenotypic differentiation among *Cryptococcus* clades**

453 To examine whether genomic divergence across the *Cryptococcus* genus is accompanied by
454 phenotypic differentiation, we performed comparative growth and metabolic profiling across 16 of the
455 17 recognized species. Eighteen strains were tested, including duplicate representatives for *C.*
456 *porticicola* and *C. amylolentus* (**Figure 7** and **Table S4**). *C. depauperatus* was excluded due to its
457 markedly slower, filamentous growth, which was incompatible with standardized assay conditions.

We first assessed the metabolic capacity of all strains with Biolog Phenotype MicroArrays, testing growth across 153 substrates. Of these, 113 yielded variable responses across the 18 strains and were retained for comparative analyses (**Table S4**). Hierarchical clustering of the binarized profiles revealed clade-level groupings, indicating that species within the same clade tend to share more similar metabolic profiles (**Figure 7A**). However, the branching order in the dendrogram does not mirror the species phylogeny (**Figure 1**), suggesting that while combination of metabolic traits can discriminate clades, they may reflect ecological adaptation or functional convergence rather than shared evolutionary history. Principal component analysis (PCA) of the same dataset (**Figure 7B**) supported broad clade separation along PC1 (23.7%) and PC2 (10.9%). The only exception was *C. cederbergensis* (clade B), which clustered within the 95% confidence ellipse of clade C, suggesting metabolic similarity despite phylogenetic distance. This pattern may, however, change with the discovery and characterization of additional clade B species in the future.

To further explore clade-specific metabolic differentiation, we applied a Random Forest classifier to the binarized Biolog dataset. This analysis identified a ranked set of informative features, with glycerol, glucuronamide, L-serine, xylitol, and D-galacturonic acid among the most discriminating substrates (**Figure S8A**). Visualization of their average presence across clades revealed clear clade-specific patterns (**Figure S8B**). For example, glycerol and xylitol were utilized by all clades except clade A; glucuronamide utilization was largely restricted to clade A; and clades B and C showed impaired growth in 4% and 8% NaCl, consistent with heightened osmotic stress sensitivity. Conversely, D-melibiose utilization was restricted to clade C species. These trends suggest that substrate utilization profiles may reflect underlying phylogenetic structure and offer phenotypic markers for clade discrimination. However, because Random Forest identifies correlates of classification rather than causal determinants, further experiments will be needed to validate these candidate traits and assess their diagnostic or taxonomic utility.

In parallel, we evaluated strains from all recognized *Cryptococcus* species (excluding *C. depauperatus*) for two traits broadly associated with pathogenicity: thermotolerance and melanin

484 production. Growth at 37 °C was tested on rich medium, and melanin synthesis was assessed on Niger
485 seed agar and L-DOPA (**Figure 7C** and **Figure S8C**). Thermotolerance was sharply restricted to clade
486 A pathogens: all clade A strains grew robustly at 35 °C and 37 °C, while none of the clade B, C, or D
487 strains did. Melanin production showed a similar distribution: all clade A strains produced some level of
488 melanin on both media, whereas all nonpathogenic species lacked detectable production. This pattern
489 was not explained by variation in laccase gene copy number. Laccases are oxidoreductase enzymes
490 that catalyze melanin biosynthesis by oxidizing diphenolic compounds such as L-DOPA [47, 48]. For
491 instance, clade C species such as *C. amylolentus* encode up to four predicted laccase genes but failed
492 to produce melanin under the tested conditions. Together, these results align with prior studies linking
493 thermotolerance and melanin biosynthesis to virulence in *Cryptococcus* and suggest that these traits
494 underlie both ecological adaptation and the evolution of pathogenic potential.

495

496 **Taxonomic description of novel *Cryptococcus* species**

497 **Description of *Cryptococcus hyracis* M.C. Fisher, M. Vanhove, R.A. Farrer, M. David-Palma, M.A.**
498 **Coelho, J. Heitman, and K.J. Kwon-Chung, sp. nov.**

499 Etymology: from hyrax (*Dendrohyrax arboreus*), an herbivorous mammal native to Africa and
500 the Middle East; the species name refers to hyrax-associated environments, as several strains,
501 including the type, were isolated from habitats such as tree holes and middens linked to hyrax activity.

502 After 1 week at 25 °C on PD and GPY streak is white to cream, mucoid with a smooth glistening
503 surface. Margins are smooth and entire. Cells are globose to ovoid 3-8 µm is size, occurring singly or
504 in pairs, and proliferating by multilateral budding. Cells are surrounded by polysaccharide capsules.
505 Pseudohyphae and true hyphae were not observed. Ballistoconidia were not observed. Teleomorph
506 was not observed, as currently available strains are of the same mating type (*MAT*α).

507 Glucose is not fermented. Positive growth on D-Glucose, D-Galactose, L-Sorbose, D-Xylose, L-
508 Arabinose, L-Rhamnose, Sucrose, Maltose, a,a-Trehalose, Me a-D-Glucoside, Cellobiose, Salicin,

Arbutin spliting, Raffinose, Melezitose, Inulin, Starch, Ribitol, Arabinitol, D-Glucitol, D-Mannitol, Galactitol, myo-Inositol, 2-Keto-D-Gluconate, D-Gluconate, D-Glucuronate, D-Galacturonate, Succinate, Citrate, Ethanol, D-glucarate, Galactonic Acid Lactone, Palatinose, L-Malic acid, L-Tartaric acid, Galactaric acid, Gentobiose, and Tween 80. No growth on D-Glucosamine, D-Ribose, D-Arabinose, Melibiose, Lactose, Glycerol, Erythritol, Xylitol, DL-Lactate, Methanol, Propane 1,2 diol, Quinic acid, D-Tartaric acid, and Tween 40.

No growth in the presence of 10% sodium chloride, but on 50% and 60% D-Glucose. Urea hydrolysis and Diazonium Blue B reaction are positive. Starch-like compounds are produced. Maximum growth temperature: 37 °C.

Type: holotype, strain MF34, isolated from a tree hole associated with the Southern tree hyrax (*Dendrohyrax arboreus*), collected on 20 September 2013 in Mutinondo (latitude −12.45, longitude 31.29), Central Zambezian Miombo Woodlands; preserved in a metabolically inactive state in the German Collection of Microorganisms and Cell Cultures (DSMZ), Braunschweig, Germany as DSM 117690.

Description of *Cryptococcus cederbergensis* S. Marincowitz, N.Q. Pham, M. David-Palma, M.A. Coelho, J. Heitman, and M.J. Wingfield, sp. nov

Etymology: refers to the Cederberg Mountains in South Africa, where the sample that yielded the type strain was collected. Formed as a Latinized geographic adjective.

After 1 week at 25 °C on PD and GPY agars, streak is white to cream, mucoid with a smooth glistering surface. On malt extract agar, culture is more mucoid. Margins are smooth and entire. Cells are globose to ovoid 3-8 µm is size, occurring singly or in pairs, and proliferating by multilateral budding. Cells are surrounded by polysaccharide capsules. Pseudohyphae and true hyphae were not observed. Ballistoconidia were not observed. Teleomorph was not observed.

533 Glucose is not fermented. Positive growth on D-Glucose, D-Galactose, D-Glucosamine, D-
534 Ribose, D-Xylose, L-Arabinose, D-Arabinose, L-Rhamnose, Sucrose, Maltose, a,a-Trehalose, Me a-D-
535 Glucoside, Cellobiose, Salicin, Arbutin spliting, Melibiose, Lactose, Raffinose, Melezitose, Inulin, Starch,
536 Glycerol, Erythritol, Ribitol, Xylitol, Arabinitol, D-Glucitol, D-Mannitol, Galactitol, myo-Inositol, 2-Keto-D-
537 Gluconate, D-Gluconate, D-Glucuronate, Succinate, Citrate, Methanol, Ethanol, D-glucarate,
538 Galactonic Acid Lactone, Palatinose, L-Malic acid, Galactaric acid, Gentobiose, and Tween 80. No
539 growth on L-Sorbose, D-Galacturonate, DL-Lactate, Propane 1,2 diol, Quinic acid, L-Tartaric acid, D-
540 Tartaric acid, and Tween 40.

541 No growth in the presence of 10% sodium chloride, but not on 50% D-Glucose. Urea hydrolysis
542 and Diazonium Blue B reaction are positive. Starch-like compounds are produced. Maximum growth
543 temperature: 30 °C (weak, slow).

544 Type: holotype, strain CMW60451 isolated from a bark beetle (*Lanurgus* sp.) infesting twigs of
545 the endangered conifer *Widdringtonia cedarbergensis*, collected on 17 October 2022 in the Cederberg
546 Mountains of South Africa, preserved in the H.G.W.L. Schweickerdt herbarium (PRU). Ex-holotype
547 CMW 60451, preserved in a metabolically inactive state in the culture collection of the Forestry and
548 Agricultural Biotechnology Institute (FABI) at the University of Pretoria, South Africa.

549

550 **Description of *Cryptococcus maquisensis* M.A. Coelho, M. David-Palma, O. Röhl, J.P. Sampaio,**
551 **J. Heitman, and A.M. Yurkov, sp. nov.**

552 Etymology: from *maquis*, the French term for dense Mediterranean shrubland, referring to the
553 habitat from which the species was isolated.

554 After 1 week at 25 °C on PD and GPY agars, streak is white to cream, butyrous to mucoid with
555 a smooth surface. Margins are smooth and entire. Cells are globose to ovoid 3-8 µm is size, occurring
556 singly or in pairs, and proliferating by multilateral budding. Cells are surrounded by polysaccharide

557 capsules. Short pseudohyphae observed in old cultures. No true hyphae were observed. Ballistoconidia
558 were not observed. Teleomorph was not observed.

559 Glucose is not fermented. Positive growth on D-Glucose, D-Galactose, D-Glucosamine, D-
560 Ribose, D-Xylose, D-Arabinose, L-Rhamnose, Sucrose, Maltose, a,a-Trehalose, Cellobiose, Salicin,
561 Arbutin spliting, Melibiose, Raffinose, Melezitose, Starch, Glycerol, Erythritol, Ribitol, Xylitol, Arabinitol,
562 D-Glucitol, D-Mannitol, Galactitol, myo-Inositol, 2-Keto-D-Gluconate, D-Gluconate, D-Glucuronate,
563 Succinate, Citrate, Ethanol, D-glucarate, Galactonic Acid Lactone, Palatinose, L-Malic acid, L-Tartaric
564 acid, Galactaric acid, and Gentobiose. No growth on L-Sorbose, L-Arabinose, Me a-D-Glucoside,
565 Lactose, Inulin, D-Galacturonate, DL-Lactate, Methanol, Propane 1,2 diol, Quinic acid, D-Tartaric acid,
566 Tween 40, and Tween 80.

567 Growth in the presence of 10% sodium chloride (weak), and on 50% and 60%D-Glucose. Urea
568 hydrolysis and Diazonium Blue B reaction are positive. Starch-like compounds are produced. Maximum
569 growth temperature: 30 °C.

570 Type: isotypes of strain OR849 isolated from soil collected in September 2013 in Serra da
571 Arrábida, Setubal, Portugal, preserved in a metabolically inactive state in the German Collection of
572 Microorganisms and Cell Cultures (DSMZ), Braunschweig, Germany (DSM 104548) and the
573 Portuguese Yeast Culture Collection (PYCC), Caparica, Portugal.

574

575 **Description of *Cryptococcus fructiofici* B. Turchetti, M. David-Palma, M.A. Coelho, A.M. Yurkov,**
576 **and J. Heitman, sp. nov.**

577 Etymology: from Latin *fructus* (fruit) and *ficus* (fig), Latinized and used as a noun in apposition,
578 referring to the fig fruit, the source of isolation.

579 After 1 week at 25 °C on PD and GPY agars, streak is white to cream, butyrous to mucoid with
580 a smooth surface. Margins are smooth and entire. Cells are globose to ovoid 3-8 µm is size, occurring
581 singly or in pairs, and proliferating by multilateral budding. Cells are surrounded by polysaccharide

capsules. Pseudohyphae and true hyphae were not observed. Ballistoconidia were not observed. Teleomorph was not observed.

Glucose is not fermented. Positive growth on D-Glucose, D-Galactose, D-Ribose, D-Xylose, D-Arabinose, L-Rhamnose, Sucrose, Maltose, a,a-Trehalose, Cellobiose, Salicin, Arbutin spliting, Raffinose, Melezitose, Starch, Glycerol, Erythritol, Ribitol, Xylitol, Arabinitol, D-Glucitol, D-Mannitol, Galactitol, myo-Inositol, 2-Keto-D-Gluconate, D-Gluconate, D-Glucuronate, D-Galacturonate, Succinate, Ethanol, Palatinose, L-Malic acid, Gentobiose, and Tween 80. No growth on L-Sorbose, D-Glucosamine, L-Arabinose, Me a-D-Glucoside, Melibiose, Lactose, Inulin, DL-Lactate, Citrate, Methanol, Propane 1,2 diol, Quinic acid, D-glucarate, Galactonic Acid Lactone, L-Tartaric acid, D-Tartaric acid, Galactaric acid, and Tween 40.

Utilization of nitrogen sources: positive growth on potassium nitrate, sodium nitrite, ethylamine, lysine, and cadaverine.

No growth in the presence of 10% sodium chloride (weak), and on 50% and 60%D-Glucose. Urea hydrolysis and Diazonium Blue B reaction are positive. Starch-like compounds are produced. Maximum growth temperature: 30 °C.

Type: holotype DBVPG 10121 isolated from figs collected in September 2011 in Brufa, Perugia, Italy, preserved in a metabolically inactive state in the Industrial Yeasts Collection DBVPG, Department of Agricultural, Food and Environmental Sciences, University of Perugia, Perugia, Italy. Ex-holotype is preserved in a metabolically inactive state in the German Collection of Microorganisms and Cell Cultures (DSMZ), Braunschweig, Germany as DSM 109063.

Description of *Cryptococcus jaspeensis* M.A. Coelho, M. David-Palma, O. Röhl, J.P. Sampaio, J. Heitman, and A.M. Yurkov, sp. nov.

Etymology: from Alto do Jaspe, a locality in the Arrábida Natural Park (Portugal) where the type strain was isolated. The epithet *jaspeensis* is a Latinized toponym referring to the place of origin.

After 1 week at 25 °C on PD and GPY agars, streak is white to cream, butyrous to mucoid with a smooth surface. Margins are smooth and entire. Cells are globose to ovoid 3-8 µm is size, occurring singly or in pairs, and proliferating by multilateral budding. Cells are surrounded by polysaccharide capsules. Pseudohyphae and true hyphae were not observed. Ballistoconidia were not observed. Teleomorph was not observed.

Glucose is not fermented. Positive growth on D-Glucose, D-Galactose, D-Glucosamine, D-Ribose, D-Xylose, D-Arabinose, L-Rhamnose, Sucrose, Maltose, a,a-Trehalose, Me a-D-Glucoside, Cellobiose, Salicin, Arbutin spliting, Raffinose, Melezitose, Inulin, Starch, Glycerol, Erythritol, Ribitol, Xylitol, Arabinitol, D-Glucitol, D-Mannitol, Galactitol, myo-Inositol, 2-Keto-D-Gluconate, D-Gluconate, D-Glucuronate, D-Galacturonate, Succinate, Ethanol, D-glucarate, Palatinose, L-Malic acid, Galactaric acid, Gentobiose, Tween 40, and Tween 80. No growth on L-Sorbose, L-Arabinose, Melibiose, Lactose, DL-Lactate, Citrate, Methanol, Propane 1,2 diol, Quinic acid, Galactonic Acid Lactone, L-Tartaric acid, and D-Tartaric acid.

Utilization of nitrogen sources: positive growth on potassium nitrate, sodium nitrite, ethylamine, and lysine.

No growth in the presence of 10% sodium chloride (weak), and on 50% and 60%D-Glucose. Urea hydrolysis and Diazonium Blue B reaction are positive. Starch-like compounds are produced. Maximum growth temperature: 25 °C.

Type: isotypes of strain OR918 isolated from soil collected in September 2013 in Serra da Arrábida, Setubal, Portugal, preserved in a metabolically inactive state in the German Collection of Microorganisms and Cell Cultures (DSMZ), Braunschweig, Germany (DSM 104549) and the Portuguese Yeast Culture Collection (PYCC), Caparica, Portugal

Description of *Cryptococcus porticicola* M.A. Coelho, M. David-Palma, A. Kachalkin, M. Kolarik, A.M. Yurkov, and J. Heitman, sp. nov.

Etymology: from Latin *porticus* (meaning gallery, passage, or corridor) referring to the burrows or tunnels made by bark beetles, which is the substrate where the species has been repeatedly found.

After 1 week at 25 °C on PD and GPY agars, streak is white to cream, butyrous to mucoid with a smooth surface. Margins are smooth and entire. Cells are globose to ovoid 3-8 µm is size, occurring singly or in pairs, and proliferating by multilateral budding. Cells are surrounded by polysaccharide capsules. Pseudohyphae and true hyphae were not observed under these conditions. Ballistoconidia were not observed. Under mating conditions on V8 agar (pH 5.0), hyphae and basidia-like structures were sporadically observed after ~1 month of incubation in co-culture with strain NCYC1536. These putative sexual structures appeared embedded in the agar, forming small clusters of globose cells approximately 8–10 µm in diameter. Their morphology was consistent with basidia but their identity could not be definitively confirmed. No sexual structures were observed on MS medium.

Glucose is not fermented. Positive growth on D-Glucose, D-Galactose, L-Sorbose, D-Glucosamine, D-Ribose, D-Xylose, L-Arabinose, D-Arabinose, L-Rhamnose, Sucrose, Maltose, a,a-Trehalose, Me a-D-Glucoside, Cellobiose, Salicin, Arbutin spliting, Raffinose, Melezitose, Starch, Glycerol, Erythritol, Ribitol, Xylitol, Arabinitol, D-Glucitol, D-Mannitol, Galactitol, myo-Inositol, 2-Keto-D-Gluconate, D-Gluconate, D-Glucuronate, D-Galacturonate, Succinate, Ethanol, D-glucarate, Galactonic Acid Lactone, Palatinose, L-Malic acid, Galactaric acid, Gentobiose, Tween 40, and Tween 80. No growth on Melibiose, Lactose, Inulin, DL-Lactate, Citrate, Methanol, Propane 1,2 diol, Quinic acid, L-Tartaric acid, and D-Tartaric acid.

Utilization of nitrogen sources: positive growth on potassium nitrate, sodium nitrite, ethylamine, and lysine.

No growth in the presence of 10% sodium chloride (weak), and on 50% and 60%D-Glucose. Urea hydrolysis and Diazonium Blue B reaction are positive. Starch-like compounds are produced. Maximum growth temperature: 30 °C.

Type: from gallery of great spruce bark beetle, *Dendroctonus micans* collected in May 2017 in Moscow region, Russia; preserved as two metabolically inactive isotypes in the yeast collection of the Lomonosov Moscow State University (KBP Y-6209) and the All-Russian Collection of Microorganisms (VKM), Pushchino, Russia (VKM Y-3025). Ex-type culture is deposited in the German Collection of Microorganisms and Cell Cultures (DSMZ), Braunschweig, Germany (DSM 108351).

Paratypes: KBP Y-6300 (= DSM 108352) from the gallery of great spruce bark beetle, *Dendroctonus micans* (Moscow region, Russia), KBP Y-6370 (= DSM 111204) and KBP Y-6556 (= DSM 111205) from spruce bark beetle gut, *Ips typographus* (Moscow region, Russia), CCUG 11125 (= NCYC 1536 = DSM 117691) from spruce bark beetle gut, *Ips typographus* (Göteborg, Sweden), and CCF 6641 (= DSM 117830) from spruce bark beetle gut, *Ips typographus* (Nižbor, Czechia).

Discussion

This study expands the known diversity of the genus *Cryptococcus* by formally describing six new species distributed across four distinct clades of the phylogeny. Although initially identified as potential new *Cryptococcus* species through rRNA and later genome sequencing, these taxa had remained undescribed, limiting their integration into broader ecological and evolutionary contexts. Defining species boundaries in fungi, especially among closely related lineages, typically requires a holistic approach that integrates phylogenetic relationships, evidence of reproductive isolation, and other biological traits. However, such data are often difficult to obtain, particularly for non-model organisms with limited experimental tractability or for lineages represented by few isolates. In these cases, genomic data may offer a useful alternative, though interpretation is not without challenges.

As fungal genome resources continue to grow, fungal taxonomy is entering a phylogenomic era in which genome-scale data and quantitative divergence metrics provide a reproducible and evolutionarily grounded basis for species delineation [30, 31]. In this context, taxogenomic approaches anchored in genome-wide comparisons and robust phylogenetic inference are particularly valuable for uncovering hidden diversity and resolving taxonomic uncertainty in yeasts [49, 50]. Still, metrics such as ANI and AAI must be benchmarked against external data that reflect fungal biology, including ecology and reproduction—traits traditionally employed for species circumscription under the phenotypic and biological species concepts [51, 52]. By integrating phylogenomic inference, chromosomal comparisons, divergence metrics, and phenotypic profiling, this study establishes a robust framework for species recognition. While applied here to the genus *Cryptococcus*, this integrative approach can likely be extended to other basidiomycete genera.

In our phylogenomic analyses of *Cryptococcus*, strong overall concordance was observed across inference methods, but residual conflict remained at both deep and shallow nodes. For instance, while the placement of clades B and C relative to clade A was topologically consistent, reduced gene and site concordance and lower quartet support at this deep node point to lingering uncertainty, likely reflecting ancient incomplete lineage sorting (ILS). Within clade A, several internal nodes also exhibited gene tree discordance, likely due to a combination of ILS and recent, rapid speciation. Although historical introgression cannot be entirely ruled out within the *Cryptococcus gattii* complex, available evidence suggests it is rare and limited in scope [15, 53]. Notably, coalescent-based inference with ASTRAL yielded higher resolution across all clade A nodes, with quartet support values exceeding the 33% threshold. These patterns suggest that gene tree discordance, while common in recently diverged fungal lineages, does not necessarily compromise species tree inference, particularly when relationships are reconstructed with coalescent-aware approaches like ASTRAL, which explicitly model the effects of ILS.

Genome-wide similarity metrics such as ANI and ddDDH provide a complementary and quantitative framework for evaluating species boundaries, as demonstrated in prokaryotes. Although

704 fungi lack a universally accepted genomic threshold for species delimitation, a 95% ANI cutoff has been
705 proposed based on limited comparisons across a few selected ascomycetous and basidiomycetous
706 yeasts. However, our comparative analyses support emerging evidence that this threshold may not be
707 suitable and underestimate species-level divergence in some basidiomycete lineages [31, 54]. For
708 example, the lowest ANI and dDDH values observed among *Cryptococcus neoformans* strains (97.7%
709 and 79%, respectively) provide a conservative lower bound for intraspecific genomic similarity, while
710 the divergence between *C. bacillisporus* and *C. decagattii* (96.5% ANI; 69.4% dDDH), two recognized
711 species, establishes a practical lower bound for interspecific separation.

712 The newly described species *C. hyracis* sp. nov. illustrates the utility of this framework.
713 Phylogenomic analyses resolved *C. hyracis* as sister to *C. tetragattii*, with ANI values ranging from
714 95.32% to 95.51%, and dDDH values from 62.1% to 62.8%. These values fall below both the
715 intraspecific thresholds observed in *C. neoformans* and even the interspecific values seen between *C.*
716 *bacillisporus* and *C. decagattii*, providing quantitative support for species-level distinction. Although all
717 currently available *C. hyracis* strains are *MAT*α, precluding intraspecific mating assays, interspecific
718 crosses with *MAT***a** strains from other species produced sexual structures, suggesting the lineage
719 retains sexual competence. Further studies evaluating progeny viability and reproductive compatibility
720 with *C. tetragattii* are needed but are constrained by the current lack of *C. hyracis MAT***a** isolates for
721 control crosses. More broadly, systematic reproductive assessments across the *C. gattii* complex would
722 help refine the interpretive power of genome-based divergence thresholds. When standardized and
723 applied in a clade-specific context, these genomic similarity metrics should provide a scalable and
724 reproducible framework for species delimitation, particularly in lineages where experimental crosses
725 are infeasible. The presence of hybrid strains remains a caveat, underscoring the need to integrate
726 genomic evidence with other lines of data. Together, these findings extend earlier support for the
727 distinctiveness of *C. hyracis* [15] and reinforce the broader utility of empirical, lineage-specific
728 benchmarks for taxonomic resolution in *Cryptococcus* and other fungal taxa.

729 Chromosomal architecture and genome-wide sequence divergence do not always evolve in
730 concert, and both can independently shape species boundaries. For instance, *C. hyracis* and *C.*
731 *tetragattii* are resolved as sister species and differ by ~4.5% at the genome level, yet their karyotypes
732 are nearly collinear apart from a single large inversion on chromosome 2, which would only impact
733 fertility in the rare case of crossover within the inverted segment. A similar decoupling of sequence and
734 structural evolution is observed between *C. neoformans* 125.91 and *C. deneoformans* NIH433, which
735 diverge by ~12% but differ structurally by only two large-scale inversions (on chromosomes 1 and 9).
736 This pattern echoes observations in other fungal systems. In *Saccharomyces*, for instance, S.
737 *cerevisiae* and *S. paradoxus* retain collinear genomes but produce sterile hybrids due to anti-
738 recombination effects mediated by mismatch repair [55, 56]. In *Cryptococcus*, deletion of *MSH2* (a key
739 component of the mismatch repair system) enhances hybrid viability between *C. neoformans* and *C.*
740 *deneoformans*, yet fails to restore normal levels of meiotic recombination, implying the presence of
741 additional post-zygotic barriers [57]. Adding further nuance, some *C. neoformans* strains (e.g., H99,
742 Bt65, Bt81) harbor private rearrangements [36, 58], demonstrating that chromosomal changes can arise
743 independently within species and may not reflect broader lineage evolution. In contrast, *C. bacillisporus*
744 and *C. decagattii*, the most closely related species pair in our dataset (~3.5% divergence), exhibit up to
745 two fixed reciprocal translocations between their sequenced strains, despite high genome-wide
746 similarity. Whether these rearrangements initiated speciation or arose secondarily remains unresolved.
747 However, experimental work in *C. neoformans* shows that targeted centromere breakage can generate
748 novel karyotypes that reduce meiotic viability in the absence of sequence divergence, supporting a
749 causal role for chromosomal changes in reproductive isolation [36]. Although often underappreciated,
750 these findings draw further attention to the role of chromosomal rearrangements in fungal speciation
751 and illustrate the importance of broad within-species sampling to fully capture karyotypic dynamics and
752 evolutionary trajectories.

753 In addition to the structural rearrangements observed among closely related pathogenic species,
754 our analyses revealed more extensive karyotypic shifts in clade D, including multiple reductions in

755 chromosome number. The three newly described clade D species (*C. porticicola*, *C. fructiofici*, and *C.*
756 *jaspeensis*) harbor fewer chromosomes (9, 8, and 7, respectively) than the ancestral 14-chromosome
757 karyotype inferred for *Cryptococcus* and its sister genus *Kwoniella* [27]. Comparative synteny analyses
758 of centromeric regions, which are typically enriched in transposable elements and 5mCG DNA
759 methylation [35, 40, 41], suggest that chromosome number reductions in clade D were primarily driven
760 by structural rearrangements followed by centromere inactivation. Several centromere-syntenic regions
761 retain conserved flanking gene order but lack TE enrichment and display only residual 5mCG,
762 consistent with inactivated centromeres (*iCENs*). This model aligns with patterns reported in *C.*
763 *depauperatus* [27] and other fungi such as *Verticillium*, *Malassezia* and *Candida* [44, 46, 59, 60],
764 supporting a broader role for centromere inactivation in karyotype evolution. We also identified an
765 intercentromeric recombination (ICR) event in *C. porticicola* strain DSM111204, associated with a strain-
766 specific chromosomal rearrangement. ICR has been reported in both pathogenic and nonpathogenic
767 *Cryptococcus* species [27, 35, 36], and its detection here extends the known phylogenetic distribution
768 of this mechanism.

769 In both *C. fructiofici* and *C. jaspeensis*, the total number of inferred *iCENs* surpasses the number
770 of chromosomes lost when compared to *C. porticicola*, suggesting that centromere inactivation alone
771 does not fully account for their current karyotypes. This raises the possibility that neocentromere
772 formation may have accompanied or followed inactivation events to maintain chromosome stability. A
773 candidate centromere shared by both species, located on chromosomes 4 and 5, respectively, exhibits
774 hallmark centromeric features, including enrichment for transposable elements and 5mCG DNA
775 methylation. The syntenic region in *C. porticicola* shows only partial enrichment for these features,
776 suggesting that it may represent either a newly co-opted centromere in the ancestor of *C. fructiofici* and
777 *C. jaspeensis*, or an ancestral centromere that was subsequently lost or inactivated in *C. porticicola*. If
778 this region indeed functions as a centromere, its architecture contrasts with neocentromeres described
779 in *C. deuterogattii*, which were experimentally induced by targeted deletion of native centromeres and
780 typically form at gene-rich loci lacking repeats, transposon accumulation, and DNA methylation, and are

often epigenetically unstable [42, 43]. This distinction may reflect either species-specific differences or contrasts between newly formed neocentromeres and those that have stabilized over extended evolutionary periods. Taken together, our findings suggest that both centromere inactivation and centromere-associated recombination are recurrent processes contributing to genome restructuring and karyotype evolution in *Cryptococcus*, with potential implications for speciation.

Ribosomal RNA (rRNA) genes have long served as core phylogenetic and taxonomic markers in fungal systematics. Mirroring the broader chromosomal and centromeric changes observed across *Cryptococcus*, we found lineage-specific variation in the genomic arrangement of the 5S rRNA gene across *Cryptococcus* species. In clade A pathogens, 5S genes remain physically clustered with the 18S, 5.8S, and 28S components on a single chromosome, a configuration common in fungi [61] and likely maintained by constraints related to nucleolar organization and coordinated rRNA dosage control [62]. In contrast, nonpathogenic species exhibit unlinked and frequently dispersed 5S loci, often enriched in subtelomeric regions as seen in clade D species such as *C. fructiofici* and *C. jaspeensis*. While some clade C species retain partial enrichment of 5S genes on the same chromosome as the canonical rDNA array, others show evidence of array relocation, consistent with rDNA mobility. These patterns may reflect relaxed functional constraints or increased genome plasticity in saprobic lineages. Mechanistically, the genomic decoupling of 5S from the other rRNA genes may be facilitated by its independent transcription by RNA polymerase III, in contrast to the polymerase I-transcribed 18S, 5.8S, and 28S genes. This transcriptional distinction has been proposed as a key reason for the physical separation of 5S genes in many eukaryotes [63]. Additionally, retroelement-mediated mobility has been hypothesized as a mechanism for 5S gene dispersal, based on shared internal promoters between 5S rRNA genes and short interspersed nuclear elements (SINEs), and documented interactions with long interspersed nuclear elements (LINE) machinery [63, 64]. The subtelomeric bias observed in some *Cryptococcus* species is also consistent with known hotspots of recombination and TE activity [58]. Together, our results indicate that 5S rRNA gene organization is evolutionarily labile in *Cryptococcus*, likely influenced by its independent transcription and potential for mobilization. These lineage-specific

807 differences raise important questions about the functional consequences of rDNA organization,
808 including possible effects on nucleolar structure, rRNA transcriptional regulation, or genome stability.
809 While these impacts remain unclear, the extent of observed divergence suggests that rRNA gene
810 architecture may represent an underappreciated axis of genome evolution in *Cryptococcus* and
811 highpoints this genus as a promising model for investigating rRNA gene dynamics and genome
812 compartmentalization in fungi.

813 Beyond structural genome evolution, we also explored the recombination potential of *C.*
814 *porticicola*, the only newly described nonpathogenic species for which multiple isolates of opposite
815 mating types were available. All six strains were isolated from bark beetle galleries or larval guts,
816 indicating a strong ecological association with conifer-infesting beetles. In mating assays, only one
817 cross (DSM108352 × NCYC1536) produced hyphae and embedded basidia-like structures resembling
818 those of *Kwoniella* rather than the typical aerial basidia of *Cryptococcus*, suggesting a potentially
819 divergent mode of sexual development. However, these structures emerged only sporadically and
820 require further validation. Notably, the two mating strains differ by two chromosomal translocations,
821 which are likely to reduce progeny viability. Despite limited morphological evidence, genomic analyses
822 based on a small number of isolates support ongoing sexual reproduction in *C. porticicola*. Genome-
823 wide SNP patterns revealed reticulate relationships, while linkage disequilibrium (LD) decay and four-
824 gamete test violations were consistent with historical recombination. LD decay analysis showed that r^2
825 values dropped to 50% of their initial maximum at approximately 13.9 kb, which is broadly consistent
826 with values observed in recombining *C. neoformans* subpopulations (e.g., VNI, VNBI, and VNBII),
827 where LD50 ranges from 5 to 7.5 kb when analyzed separately [34]. The four-gamete test further
828 supported this inference, revealing a modest but increasing proportion of SNP pairs that violated the
829 four-gamete condition with distance. Although the sample size remains limited, the convergence of
830 these genomic signals suggests that sexual reproduction is maintained in natural populations of *C.*
831 *porticicola*, with meiotic processes likely contributing to genetic cohesion, evolutionary flexibility, and
832 ecological persistence. Future work will focus on optimizing mating conditions to enable successful

833 crosses between strains with more collinear karyotypes, as all other combinations tested to date have
834 not yielded evidence of mating. In parallel, recovering additional strains from bark beetle-associated
835 substrates may help identify more sexually compatible pairs. If these approaches prove unsuccessful,
836 or if sporulation remains rare or progeny are technically difficult to isolate, introducing selectable
837 markers into compatible strains could provide an alternative means to detect genetic exchange following
838 co-culture, as demonstrated in other non-model *Cryptococcus* lineages [20].

839 Complementing our genomic analyses, we evaluated phenotypic variation across the
840 *Cryptococcus* genus to determine whether clade-level divergence is mirrored by metabolic traits.
841 Comparative growth profiling revealed clear clade-level differentiation: hierarchical clustering and
842 principal component analysis (PCA) based on Biolog substrate utilization patterns showed that species
843 within the same clade tend to group together phenotypically, broadly reflecting phylogenetic
844 relationships. Notably, *C. cederbergensis* (clade B) clustered within the clade C phenotypic space,
845 suggesting functional convergence potentially linked to similar environmental conditions. Using a
846 Random Forest classifier, we identified a subset of informative substrates—such as glycerol, xylitol, and
847 D-galacturonic acid—with clade-specific utilization. These findings indicate that metabolic profiles may
848 reflect ecological specialization or historical adaptation. While Biolog assays are sensitive to
849 experimental conditions and may not always replicate physiological responses observed in conventional
850 auxanographic tube-based tests, the consistent clade-level growth responses and clustering observed
851 here supports its utility for comparative phenotyping and points to candidate traits for future genomic
852 dissection. Recent studies in Saccharomycotina yeasts have shown that ecological transitions often
853 coincide with extensive remodeling of metabolic gene repertoires [65, 66]. Our phenotype-first approach
854 in *Cryptococcus* provides a framework to identify traits that may reflect lineage-specific ecological
855 strategies and historical adaptation, although direct links to specific gene family expansions or losses
856 remain to be established.

857 Strikingly, we confirmed that two traits historically associated with virulence in *Cryptococcus*—
858 thermotolerance and melanin production—are restricted to clade A pathogens in our dataset. None of

859 the nonpathogenic species examined displayed these traits under the tested conditions. Yet, in the case
860 of melanin production, they retain a full complement of biosynthetic genes [27], including multiple
861 laccase homologs; several even harbor greater copy numbers than their pathogenic counterparts. In *C.*
862 *neoformans*, melanin synthesis is largely driven by the cell wall–associated laccase *LAC1*, which plays
863 a central role in oxidative stress resistance and survival within macrophages [67, 68], while its paralog
864 *LAC2* can partially compensate under certain conditions [69]. The retention and potential redundancy
865 of these laccase genes in nonpathogenic lineages suggest an ancestral, potentially preadaptive function
866 shaped by environmental pressures, in line with models where virulence-associated traits arise from
867 exapted ecological adaptations [70]. With respect to thermotolerance, its strict restriction to clade A
868 pathogens reinforces its status as a critical prerequisite infection and survival in mammalian hosts.
869 Stressors encountered during infection, such as elevated temperature, have been shown to trigger
870 transposable element (TE) mobilization in both *C. deneoformans* and *C. neoformans* resulting in
871 hypermutator-like phenotypes and accelerated antifungal resistance [71-73]. Although none of the
872 nonpathogenic species analyzed here can grow at 37 °C, several retain high TE burdens [27], raising
873 the possibility that stress-induced genomic plasticity may also occur in environmental contexts and
874 could, under the right conditions, prime lineages for future adaptation. Investigating whether similar
875 mutational dynamics operate in these lineages could reveal latent evolutionary capacity for
876 pathogenicity. Further work integrating pan-genomic and pan-phenomic analyses, ecological data, and
877 broader taxon sampling will be key to resolving how such traits shape adaptation and diversification
878 across the *Cryptococcus* genus.

879 The ecological breadth of *Cryptococcus* is now more apparent with the recovery of newly
880 described species from bark beetle galleries, fruit, and soil across diverse geographic regions. Although
881 distinct from the arboreal hollows, avian guano, and forest debris typically associated with *C.*
882 *neoformans* and *C. gattii* senso lato, these substrates also represent organic-rich, microbially dynamic
883 environments that may support long-term persistence and saprobic growth. These findings demonstrate
884 the importance of environmental sampling in revealing hidden diversity and advancing our

885 understanding of the ecological and evolutionary processes shaping the genus. They also suggest that
886 complex environmental niches may serve as incubators for traits involved in host adaptation and
887 pathogenicity. In this context, the formal recognition of phylogenetically distinct lineages is not
888 taxonomic excess but a necessary step toward cataloging fungal diversity, linking genotypes to
889 phenotypes, and establishing the framework needed for consistent identification of novel species and
890 strains, long-term study, surveillance, and diagnostic development, especially as some species may
891 emerge as clinically relevant under shifting ecological, climatic, or antifungal selective pressures.

892

893 Materials and methods

894 Strains, mating assays, and microscopy

895 Strains analyzed in this study were routinely cultured on YPD medium (10 g/L yeast extract, 20
896 g/L Bacto Peptone, 20 g/L dextrose, and 20 g/L agar), unless specified otherwise. *Cryptococcus*
897 *neoformans*, *C. deneoformans*, and *C. gattii* sensu lato strains were incubated at 30 °C, while other
898 *Cryptococcus* species were maintained at room temperature (20–25 °C). A complete list of strains and
899 associated metadata is provided in **Table S1**.

900 Mating assays were performed by mixing equal amounts of cells from each strain on V8 agar
901 media (10 g/L yeast extract, 20 g/L Bacto Peptone, 20 g/L dextrose, 20 g/L agar; pH 5.0) or Murashige
902 and Skoog (MS) agar media (4.328 g/L MS mix, 16 g/L agar, 1X vitamin mix, pH to 5.8) and incubated
903 in the dark at room temperature (conditions known to induce mating in *C. neoformans*) for up to one
904 month [74]. In the case of *C. porticicola*, incubation was extended due to delayed and sporadic hyphal
905 development, which was first observed around the one-month mark. Plates were regularly monitored
906 microscopically for the development of mating structures. Solo cultures of each strain were also
907 evaluated on V8 pH 5 and/or MS without a compatible partner. Assays were repeated three independent
908 times to ensure consistent phenotypic observations.

To assess hyphal growth, basidia, and spore formation, the edges of yeast colonies and mating patches were examined microscopically and photographed. Specifically for *C. porticicola*, putative sexual structures embedded within the mating media at the patch margins were imaged by excising small agar blocks and gently compressing them between a microscope slide and coverslip. Light microscopy was performed using a Zeiss Axio Scope.A1 microscope equipped with an Axiocam Color camera and ZEN Lite V3.4 software. Samples of sporulating colonies were prepared for scanning electron microscopy (SEM) as previously described [75] and imaged using the Apreo S SEM with an EDS detector (ThermoFisher, USA) at the Shared Materials Instrumentation Facility at Duke University.

Genomic DNA extraction

High-molecular-weight (HMW) genomic DNA was isolated using a cetyltrimethylammonium bromide (CTAB)-based extraction method as previously described [36], with careful handling to reduce mechanical shearing. DNA purity was assessed with a NanoDrop spectrophotometer (Thermo) by measuring A260/A280 and A260/A230 ratios. DNA integrity and fragment length were evaluated by clamped homogeneous electric field (CHEF) electrophoresis [27]. For Illumina short-read whole-genome sequencing, genomic DNA was extracted using a phenol:chloroform-based procedure [27]. Final concentrations were quantified using the Qubit dsDNA Assay Kit (Invitrogen) on a Qubit fluorometer.

Genome sequencing

Whole-genome sequencing was performed with Oxford Nanopore and Illumina technologies. Illumina sequencing for strains DSM111204, DSM111205 and DSM108352 was conducted at the Duke University Sequencing and Genomic Technologies (DUSGT) Core. Libraries were prepared with the Kapa HyperPlus kit and sequenced as paired-end 2 × 150 bp reads on Illumina Novaseq. Nanopore sequencing for strains DSM111204 and DSM108352 was performed in-house. Samples were barcoded

using the SQK-LSK109 ligation sequencing kit in combination with EXP-NBD104 native barcoding kits. Library preparation followed the manufacturer's protocols, and sequencing was carried out on R9.4.1 flow cells (FLO-MIN106) for 48–72 hours under default voltage on a MinION Mk1B device. The corresponding version of MinION software available at the time of each run was applied. Additional details on sequencing platforms, basecalling, and demultiplexing for each sample are provided in **Table S5**.

Genome assembly, gene prediction, and functional annotation

Complete nuclear genomes of *C. porticicola* strains DSM111204 and DSM108352 were assembled using Canu [76] with Nanopore long-read data and polished with both Nanopore and Illumina reads, following a previously established workflow [25]. This process included iterative error correction, removal of mitochondrial and rDNA-only contigs, assessment of assembly completeness and telomeric regions, and manual inspection of read coverage using IGV. The draft genome assembly of *C. porticicola* strain DSM111205 was generated using SPAdes v3.15.3 [77] with default parameters. All assemblies and corresponding raw sequencing data have been deposited in DDBJ/EMBL/GenBank under the BioProject numbers listed in **Table S5**. Gene models were predicted on repeat-masked assemblies using Funannotate v1.8.9 (https://github.com/nextgenusfs/funannotate), as previously described [25]. Manual inspection and curation were limited to the mating-type loci, including the annotation of short mating pheromone precursor genes. Functional annotation was performed using the Funannotate "*annotate*" module and incorporated PFAM and InterPro domains, Gene Ontology (GO) terms, COGs, CAZymes, fungal transcription factors, MEROPS proteases, secreted proteins, secondary metabolites, and BUSCO groups, as previously detailed [25].

959 **Ortholog identification, phylogenetic analyses, and topological support**

960 To construct the phylogenomic data matrix, single-copy orthologs (SC-OGs) were identified
961 across all *Cryptococcus* species and strains, along with the outgroup *Kwoniella shandongensis*
962 (GCA_008629635.2), using OrthoFinder v3.0.1b1, as previously outlined [25]. A total of 3,510 SC-OGs
963 shared among all species were identified. To ensure quality and comparability, only proteins from the
964 *C. neoformans* H99 reference strain that were ≥100 amino acids in length and whose orthologs
965 exceeded 50% of the H99 protein length were retained. These filtering steps yielded 3,321 SC-OGs,
966 which were individually aligned with MAFFT (L-INS-i mode) [78] and trimmed with TrimAl (gappyout)
967 [79]. To enrich phylogenetic signal, only alignments with >30% parsimony-informative sites were
968 retained using PhyKIT (`pis`) [80], resulting in a final dataset of 2,687 SC-OGs. Maximum likelihood (ML)
969 and coalescence-based species trees were inferred from this dataset. For ML inference, SC-OG
970 alignments were concatenated into a partitioned supermatrix (39 taxa, 2,687 partitions, and 1,625,459
971 sites) and analyzed with IQ-TREE using the "`-p`" option [81]. For coalescence-based inference,
972 individual gene trees were inferred with IQ-TREE using the "`-S`" option. Branches with <70% UFBoot
973 support were collapsed using Newick utilities (`nw_ed "i & b <= 70" o`), and a species tree was then
974 estimated with ASTRAL v5.7.8. Branch support and genealogical concordance were assessed using
975 SH-aLRT, UFBoot, gene concordance factors (gCF), and site concordance factors (sCF) for ML trees,
976 and local posterior probabilities (LPP) and quartet support for the ASTRAL tree. Phylogenetic trees were
977 visualized and annotated with iTOL v7. Detailed scripts containing software versions, and model
978 parameters used in these analyses, including visualization of gCF/sCF and ASTRAL support, are
979 accessible at https://doi.org/10.5281/zenodo.15686462.

980

981 **ANI, AAI and dDDH**

982 Pairwise average nucleotide identity (ANI) values were computed using OrthoANIu [82]. Custom
983 Python scripts were used to parse and reformat the resulting matrix, convert ANI values to distance

scores (100 – ANI), and perform average linkage clustering to generate dendrograms, and Newick-format trees visualized with iTOL v7. Minimum, maximum, and mean ANI values between species or clade groupings were also calculated using a metadata file mapping strains to their respective groups. Average amino acid identity (AAI) was computed using EzAAI v1.2.3 [83] with the DIAMOND algorithm and thresholds of 0.4 identity and 0.5 coverage. Pairwise AAI values were parsed using Python scripts and similarly used for clustering and for computing summary statistics across species or clades. Correlation analysis between pairwise ANI and AAI values was assessed using a custom Python script and by computing Pearson and Spearman correlation coefficients. Pairwise digital DNA–DNA hybridization (dDDH) values were estimated through the Genome-to-Genome Distance Calculator (GGDC 3.0; https://ggdc.dsmz.de/ggdc.php#), using Formula 2, which is recommended for incomplete genome assemblies [32]. Due to the requirement for online submission, representative intra- and interspecies genome comparisons were selected and submitted manually (**Table S2**). The output files were parsed with a custom Python script, which extracted Formula 2 dDDH values and the associated probability of exceeding the 70% species threshold. Based on these values, comparisons were categorized as intra- or interspecies, and further classified into confidence levels. This analysis was used to assess congruence with ANI and AAI-based classifications and evaluate species boundaries among closely related lineages. Detailed scripts containing software versions and parameters used in these analyses are available at https://doi.org/10.5281/zenodo.15686462.

Synteny analyses and centromere identification

Synteny relationships were assessed using three complementary approaches. First, for full karyotype comparisons across selected species, we applied the ntSynt toolkit [84] to identify both strain- and species-specific chromosomal rearrangements (**Figures 3A**, **4H**, **S3A**, and **S4A**). For clade A comparisons, ntSynt was run with the following parameters: *-d 5, -k 21, -w 200, --block_size 5000, --merge 20000, --indel 5000, and --w_rounds 200 100*. For clade D comparisons, the

following parameters were used: *-d 20, -k 24, -w 200, --block_size 1000, --merge 10000, --indel 5000, and --w_rounds 200 100*. Resulting synteny blocks were visualized using ntSynt-viz [85], which combines ribbon-style plotting with chromosome painting. A pruned species tree in Newick format was supplied for each comparison to guide genome assembly order and improve visual coherence.

Second, to examine chromosomal structural conservation and intercentromeric translocations in greater detail, we performed nucleotide-level synteny analyses using a custom Python workflow (*0_synteny_karyoplotter.py*). Genome-wide alignments were generated with minimap2 (v2.17-r941; *-x asm20*), and synteny blocks were parsed from the resulting PAF files. Analyses were conducted in reference-based mode, with all assemblies aligned to a single genome: *C. neoformans* strain 125.91 for clade A comparisons (**Figure 3B**), and *C. porticicola* DSM108351 for clade D (**Figure 5A**). Centromere coordinates were provided in BED format and overlaid on the genome plots to highlight centromeric regions and infer intercentromeric rearrangements. All visualizations were generated in Python 3 using Biopython, pandas, matplotlib, and numpy, and the final SVG outputs were manually refined in Adobe Illustrator for publication.

Third, linear synteny plots comparing chromosomes and selected genomic regions in clade D species, particularly around centromeres, were generated with EasyFig v2.2.2 [86] using BLASTN for alignment. Centromere positions and genome-wide GC content were determined as previously described [25, 27]. Centromeres were defined as the longest ORF-free regions enriched in centromere-associated LTR retrotransposons and exhibiting elevated 5mCG methylation, with boundaries set by conserved flanking centromeric genes. GC content and 5mCG methylation profiles were visualized in IGV and precisely overlaid at scale with EasyFig-derived synteny plots in Adobe Illustrator. Scripts are available at https://doi.org/10.5281/zenodo.15686462.

DNA methylation detection

DNA methylation analysis was performed using a Dorado–Modkit workflow to detect 5-methylcytosine at CpG dinucleotides (5mCG) from Oxford Nanopore sequencing data. Raw signal data in POD5 format were basecalled with Dorado v0.5.2 using the combined model `sup@latest,5mCG_5hmCG@latest,6mA@latest` to simultaneously call canonical bases and modifications. Output BAM files were then realigned to the genome assemblies with the Dorado aligner and sorted with SAMtools v1.10. Methylation signals were extracted and quantified using Modkit v0.2.5-rc1 (pileup mode, `--preset traditional`) to generate both pileup and bedGraph outputs. Centromere-associated methylation profiles were visualized in IGV and plotted to scale in Adobe Illustrator for integration with genome structure and synteny plots.

rDNA analysis

A unified Python-based pipeline was developed for identifying and analyzing the distribution of rRNA genes across *Cryptococcus* genomes. For each genome, rRNA genes were annotated using Barrnap v0.9 (https://github.com/tseemann/barrnap) with the *"--kingdom euk"* setting, and coordinates for 5S, 5.8S, 18S, and 28S rRNA genes were extracted from the GFF output. Canonical rDNA clusters were defined as colocalized 28S–5.8S–18S genes on the same strand within a 5-kb window, optionally including an adjacent 5S gene. Per-genome chromosome plots were generated to display scaled chromosomes, centromeres, and rRNA gene positions. To test whether 5S genes were non-randomly distributed across chromosomes, the expected number of 5S genes per contig was estimated by normalizing contig lengths relative to the total genome size , assuming uniform probability of 5S placement. Observed versus expected values were compared using a chi-squared goodness-of-fit test, implemented with scipy.stats.chisquare. Standardized residuals were calculated to identify contigs with over- or under-represented 5S copy numbers, and results were tabulated per genome. Additionally, to assess 5S enrichment in subtelomeric regions, subtelomeric windows were defined as

the first and last 2.5% of each contig, corresponding to 50 kb at each end of a 2 Mb chromosome. The midpoint of each 5S gene was used to determine overlap with these regions. A 2×2 contingency table was generated for each genome comparing the number of 5S genes located inside versus outside subtelomeric windows, relative to the corresponding base pair content. Significance was evaluated using a one-sided Fisher's exact test (scipy.stats.fisher_exact, alternative="greater") to test for enrichment. All gene coordinates, cluster counts, statistical results, and graphical outputs were exported as structured TSV and SVG files for each genome. To enhance figure clarity and readability, plots were refined by adjusting color schemes, modifying labels, and manually adding additional features using Adobe Illustrator. Scripts and raw data files are available at https://doi.org/10.5281/zenodo.15686462 and in **Table S3**.

Variant analysis and phylogenetic network in *C. porticicola*

Variant analysis within *C. porticicola*, was performed with Snippy v4.6.0 (https://github.com/tseemann/snippy), mapping reads to the strain DSM108351 reference genome (`--cpus 28, --unmapped, --mincov 10, --minfrac 0.9`). The resulting core SNP alignment (core.aln) served as input for Neighbor-net network reconstruction in SplitsTree v6.4.14. Recombination was assessed with the Pairwise Homoplasy Index (Phi) test, a permutation-based method implemented in SplitsTree that detects recombination by identifying statistically significant incompatibilities among adjacent sites in an alignment. The test compares observed homoplasy to that expected under a model of clonal evolution, with p-values < 0.05 interpreted as evidence for recombination [87].

Population structure and recombination analyses in *C. porticicola*

To investigate recombination within *C. porticicola*, we developed a series of analysis scripts (with filenames indicated in parentheses and available at https://doi.org/10.5281/zenodo.15686462) beginning with the conversion of the multi-sample VCF generated by Snippy to PLINK binary format

using PLINK v1.9 with the `--vcf` and `--allow-extra-chr` options (`0_convert_vcf_to_plink.sh`). Principal Component Analysis (PCA) was conducted in PLINK (`1_run_pca.sh`), and the resulting eigenvectors were plotted using Python (`2_plot_pca.py`) with pandas, seaborn, and matplotlib. Linkage disequilibrium (LD) decay was assessed using PopLDdecay (`3_run_ld_decay_compare.sh`) with a maximum inter-SNP distance of 50 kb. To reduce the influence of rare alleles on LD estimates, we applied a minor allele frequency (MAF) threshold of 0.167, corresponding to a 1/3 threshold for biallelic variants in haploid genomes (i.e., at least 2 of 6 strains carrying the minor allele). LD decay was also calculated without MAF filtering for comparison. The resulting files were parsed and plotted using Python (`4_plot_ld_decay.py`), applying a centered rolling mean and LOWESS smoothing to generate smoothed r^2 curves. LD50 values were estimated as the distance at which smoothed r^2 values dropped to 50% of the initial maximum, providing a quantitative summary of LD decay. To further examine signatures of historical recombination, a custom Python implementation of the four-gamete test was applied to all biallelic SNP pairs within 10 kb windows (`5_four_gamete_test.py`). SNP pairs were binned by physical distance (0–1 kb, 1–5 kb, and 5–10 kb), and the proportion of pairs exhibiting all four possible haplotypes, indicating historical recombination or recurrent mutation, was computed for each bin. Results were visualized as a barplot summarizing violation frequencies per bin.

Biolog phenotypic assays and clustering analyses

Phenotypic MicroArray tests were performed using Biolog Gen III, FF, and YT microplates. Fresh 3–7-day-old cultures grown on PDA were used for plate inoculations. Yeast cell density in sterile water (for YT and FF microplates) and Inoculation Fluid B (for Biolog Gen III microplates) was adjusted to the values recommended by the manufacturer for YT microplates. Microplates were inoculated with 100 µl of cell suspension per well and incubated in the dark at 25 °C. Plates were read immediately after inoculation using a microplate reader (Varioskan LUX Multimode Microplate Reader, Thermo Fisher Scientific) at 490 nm, 590 nm, and 750 nm, following the manufacturer's instructions. Subsequent

readings were taken every 3–7 days for up to 21 days. Plates were shaken before each reading to ensure a more homogeneous suspension.

Biolog microplates are an end-point measurement platform. Tests that showed a positive signal during the period were scored as positive. Tests showing no signal and values close to the negative control were scored as negative. In cases where color development and turbidity did not provide a clear result, tests were evaluated by comparison with carbon sources commonly utilized by the genus *Cryptococcus*, as previously demonstrated by auxanographic tube tests. Phenotypic characterization of cultures followed the methods described in Kurtzman et al. (2011) [88]. For standard species descriptions, auxanographic tests were performed in liquid media using both tubes and microplates. Tube tests were conducted in 2 ml volumes in 5- and 7.5-ml plastic tubes following the method described by Kurtzman et al. (2011) [88]. Microplate tests were performed in 200 µl volumes using the same media composition as the tube tests. Tubes and plates were shaken every 1–2 days. Growth in tubes was evaluated visually, while growth in microplates was measured at 600 nm using the Varioskan LUX microplate reader. Results were scored as positive or weak when there was clear or reduced growth compared to glucose, and as negative when growth was minimal or absent relative to both the negative control and glucose.

Phenotypic data from Biolog substrate utilization assays were analyzed with a custom Python pipeline (`0_biolog_analysis_pipeline.py`). Input consisted of a manually curated Excel file (**Table S4**) containing binary growth data across multiple *Cryptococcus* strains. Dimensionality reduction and clustering analyses were performed to explore global patterns of phenotypic similarity. Principal component analysis (PCA) was conducted on the standardized dataset, and the first two components were visualized with 95% confidence ellipses grouped by clade. A binary clustermap was generated using hierarchical clustering based on Euclidean distance and average linkage. To identify traits most informative for clade-level classification, a Random Forest classifier was trained on the binary matrix treating clade membership as the target variable, and the top 15 features ranked by importance were retained. Clade-wise presence frequencies for these phenotypes were then calculated and visualized.

1135 Redundant phenotypes, defined as those with identical binary profiles across all strains, were
1136 automatically flagged to reduce overrepresentation. All analyses and visualizations were performed
1137 using Python with pandas, scikit-learn, seaborn, matplotlib, and related packages.

1138

1139 **Temperature and melanin production assays**

1140 For temperature assays, pre-inocula of each strain were grown in 4 ml of YPD media overnight
1141 at room temperature. New tubes containing 4 ml YPD media each were inoculated with 50 μL of each
1142 of the overnight-grown pre-inocula and incubated at room temperature for 48 hours. Cells were then
1143 pelleted, washed with sterile distilled water, and adjusted to an optical density of 2.0 at 600 nm. Serial
1144 10-fold dilutions were prepared from these normalized cultures, and 3 µL of each dilution was spotted
1145 onto YPD agar plates. Plates were incubated at 23 °C, 30 °C, 35 °C, and 37 °C, and photographed after
1146 3 days. For melanin production assays, 3 µL of each dilution was also spotted onto two melanin-inducing
1147 media: L-DOPA agar (7.6 mM L-asparagine, 5.6 mM glucose, 22 mM KH_2PO_4, 1mM of $MgSO_4.7H_2O$,
1148 0.5 mM L-DOPA, 0.3 μM thiamine, 20 nM biotin and 20 g/L of agar) and Bird seed agar (7% bird seed
1149 extract, 0.1% glucose, and 20 g/L agar) [75, 89]. These plates were incubated at room temperature in
1150 the dark for 5 days before imaging. All assays were performed in three independent biological replicates
1151 to ensure reproducibility of phenotypic results.

1152

1153 **Funding Statement**

1154 This study was supported by the National Institute of Allergy and Infectious Diseases of the National
1155 Institutes of Health under awards R01 AI050113-20 (J.H.), R01 AI039115-27 (J.H.), and R01 AI133654-
1156 08 (J.H.). JH is Co-Director and Fellow of the CIFAR program Fungal Kingdom: Threats & Opportunities.
1157 The funders had no role in study design, data collection and analysis, decision to publish, or preparation
1158 of the manuscript.

Acknowledgements

We thank Prof. Matthew C. Fisher (Imperial College London) for the isolation of *Cryptococcus hyracis* strains; Dr. Aleksey Kachalkin (Lomonosov Moscow State University and VKM collection) and Dr. Miroslav Kolarik (Czech Academy of Sciences, Prague, CZ) for the isolation of multiple *C. porticicola* strains (DSM111204, DSM111205, DSM108351, DSM108352, and DSM117830); Dr. Benedetta Turchetti (University of Perugia) for isolating strain DSM109063 (*C. fructiofici*); Prof. Mike Wingfield and his team for the isolation of strain CMW60451 (*C. cederbergensis*); and Oliver Röhl and Prof. José Paulo Sampaio (Universidade Nova de Lisboa) for the isolation of strains OR849 (*C. maquisensis*) and OR918 (*C. jaspeensis*). We are also grateful to Dr. Konstanze Bensch for her suggestions and assistance with the Latin naming of the new species. Finally, we thank Fred Dietrich for providing computational resources.

Author contributions

Conceptualization: Marco A. Coelho, Márcia David-Palma, Joseph Heitman

Data Curation: Marco A. Coelho, Márcia David-Palma, Andrey M. Yurkov

Formal analysis: Marco A. Coelho

Funding acquisition: Andrey M. Yurkov, Joseph Heitman

Investigation: Marco A. Coelho, Márcia David-Palma, Andrey M. Yurkov

Project Administration: Joseph Heitman

Resources: Andrey M. Yurkov, Joseph Heitman

Software: Marco A. Coelho

Supervision: Joseph Heitman

Visualization: Marco A. Coelho, Márcia David-Palma

Writing – original draft: Marco A. Coelho, Márcia David-Palma

Writing – review & editing: Marco A. Coelho, Márcia David-Palma, Andrey M. Yurkov, Joseph Heitman

References

1. Vuillemin P. Les blastomycètes pathogènes. Rev Gen Sci Pures Appl. 1901;12:732-51.

2. Barnett JA. A history of research on yeasts 14: medical yeasts part 2, *Cryptococcus neoformans*. Yeast. 2010;27(11):875-904. doi: 10.1002/yea.1786. PubMed PMID: 20641025.

3. Busse O. Über parasitäre zelleinschlüsse und ihre züchtung. Zentralbl Bakteriol. 1894;16:175-80.

4. Sanfelice F. Contributo alla morfologia e biologia dei blastomiceti che si sviluppano nei succhi di alcuni frutti. Ann Igien. 1894;4:463-95.

5. Rajasingham R, Govender NP, Jordan A, Loyse A, Shroufi A, Denning DW, et al. The global burden of HIV-associated cryptococcal infection in adults in 2020: a modelling analysis. Lancet Infect Dis. 2022;22(12):1748-55. Epub 20220829. doi: 10.1016/S1473-3099(22)00499-6. PubMed PMID: 36049486; PubMed Central PMCID: PMC9701154.

6. Kwon-Chung KJ. A new genus, *Filobasidiella*, the perfect state of *Cryptococcus neoformans*. Mycologia. 1975;67:1197-200.

7. Velagapudi R, Hsueh YP, Geunes-Boyer S, Wright JR, Heitman J. Spores as infectious propagules of *Cryptococcus neoformans*. Infect Immun. 2009;77(10):4345-55. Epub 2009/07/22. doi: 10.1128/IAI.00542-09. PubMed PMID: 19620339; PubMed Central PMCID: PMC2747963.

8. Tugume L, Ssebambulidde K, Kasibante J, Ellis J, Wake RM, Gakuru J, et al. Cryptococcal meningitis. Nat Rev Dis Primers. 2023;9(1):62. Epub 2023/11/10. doi: 10.1038/s41572-023-00472-z. PubMed PMID: 37945681.

9. Casalini G, Giacomelli A, Antinori S. The WHO fungal priority pathogens list: a crucial reappraisal to review the prioritisation. Lancet Microbe. 2024;5(7):717-24. Epub 20240409. doi: 10.1016/S2666-5247(24)00042-9. PubMed PMID: 38608682.

10. Hagen F, Khayhan K, Theelen B, Kolecka A, Polacheck I, Sionov E, et al. Recognition of seven species in the *Cryptococcus gattii*/*Cryptococcus neoformans* species complex. Fungal Genetics and Biology. 2015;78:16-48. doi: https://doi.org/10.1016/j.fgb.2015.02.009.

11. Fraser JA, Giles SS, Wenink EC, Geunes-Boyer SG, Wright JR, Diezmann S, et al. Same-sex mating and the origin of the Vancouver Island *Cryptococcus gattii* outbreak. Nature. 2005;437(7063):1360-4. doi: 10.1038/nature04220.

12. Byrnes EJ, 3rd, Li W, Ren P, Lewit Y, Voelz K, Fraser JA, et al. A diverse population of *Cryptococcus gattii* molecular type VGIII in southern Californian HIV/AIDS patients. PLOS Pathog. 2011;7(9):e1002205. Epub 2011/09/13. doi: 10.1371/journal.ppat.1002205. PubMed PMID: 21909264; PubMed Central PMCID: PMC3164645.

13. Hagen F, Lumbsch HT, Arsic Arsenijevic V, Badali H, Bertout S, Billmyre RB, et al. Importance of resolving fungal nomenclature: the case of multiple pathogenic species in the *Cryptococcus* genus. mSphere. 2017;2(4). Epub 2017/09/07. doi: 10.1128/mSphere.00238-17. PubMed PMID: 28875175; PubMed Central PMCID: PMC5577652.

14. Poplin V, Smith C, Caceres DH, Herkert PF, Jegede O, Thompson GR, 3rd, et al. Geographical distribution of the *Cryptococcus gattii* species complex: a systematic review. Lancet Microbe. 2024;5(12):100921. Epub 20240824. doi: 10.1016/S2666-5247(24)00161-7. PubMed PMID: 39191262; PubMed Central PMCID: PMC11995670.

15. Farrer RA, Chang M, Davis MJ, van Dorp L, Yang D-H, Shea T, et al. A new lineage of *Cryptococcus gattii* (VGV) discovered in the Central Zambezian Miombo Woodlands. mBio. 2019;10(6):e02306-19. doi: 10.1128/mBio.02306-19.

16. Kwon-Chung KJ, Bennett JE, Wickes BL, Meyer W, Cuomo CA, Wollenburg KR, et al. The case for adopting the "species complex" nomenclature for the etiologic agents of Cryptococcosis. mSphere. 2017;2(1). Epub 2017/01/20. doi: 10.1128/mSphere.00357-16. PubMed PMID: 28101535; PubMed Central PMCID: PMC5227069.

17. Petch T. Notes on entomogenous fungi. Trans Brit Mycol Soc. 1931;16:209-45.

18. Rodriguez-Carres M, Findley K, Sun S, Dietrich FS, Heitman J. Morphological and genomic characterization of *Filobasidiella depauperata*: a homothallic sibling species of the pathogenic *Cryptococcus* species complex. PLOS ONE. 2010;5(3):e9620. doi: 10.1371/journal.pone.0009620.

19. Liu XZ, Wang QM, Goker M, Groenewald M, Kachalkin AV, Lumbsch HT, et al. Towards an integrated phylogenetic classification of the Tremellomycetes. Stud Mycol. 2015;81:85-147. doi: 10.1016/j.simyco.2015.12.001. PubMed PMID: 26955199; PubMed Central PMCID: PMC4777781.

20. Passer AR, Clancey SA, Shea T, David-Palma M, Averette AF, Boekhout T, et al. Obligate sexual reproduction of a homothallic fungus closely related to the *Cryptococcus* pathogenic species complex. eLife. 2022;11. Epub 2022/06/18. doi: 10.7554/eLife.79114. PubMed PMID: 35713948; PubMed Central PMCID: PMC9296135.

21. Passer AR, Coelho MA, Billmyre RB, Nowrousian M, Mittelbach M, Yurkov AM, et al. Genetic and genomic analyses reveal boundaries between species closely related to *Cryptococcus* pathogens. mBio. 2019;10(3):e00764-19. doi: 10.1128/mBio.00764-19. PubMed PMID: 31186317; PubMed Central PMCID: PMC6561019.

22. Begerow D, Kemler M, Feige A, Yurkov A. Parasitism in yeasts. In: Buzzini P, Lachance M-A, Yurkov A, editors. Yeasts in Natural Ecosystems: Ecology. Cham: Springer International Publishing; 2017. p. 179-210.

23. Guterres DC, Ndacnou MK, Saavedra-Tobar LM, Salcedo-Sarmiento S, Colman AA, Evans HC, et al. *Cryptococcus depauperatus*, a close relative of the human-pathogen *C. neoformans*, associated with coffee leaf rust (*Hemileia vastatrix*) in Cameroon. Braz J Microbiol. 2021;52:2205-14. Epub 2021/08/22. doi: 10.1007/s42770-021-00592-2. PubMed PMID: 34417998.

24. Coelho MA, David-Palma M, Aylward J, Pham NQ, Visagie CM, Fuchs T, et al. Decoding *Cryptococcus*: from African biodiversity to worldwide prevalence. PLOS Pathogens. 2025;21(2):e1012876. doi: 10.1371/journal.ppat.1012876.

25. Coelho MA, David-Palma M, Marincowitz S, Aylward J, Pham NQ, Yurkov AM, et al. Tracing the evolution and genomic dynamics of mating-type loci in *Cryptococcus* pathogens and closely related species. bioRxiv. 2025. Epub 2025/02/24. doi: 10.1101/2025.02.12.637874. PubMed PMID: 39990455; PubMed Central PMCID: PMC11844451.

26. Kachalkin AV, Turchetti B, Inácio J, Carvalho C, Mašínová T, Pontes A, et al. Rare and undersampled dimorphic basidiomycetes. Mycol Prog. 2019. doi: 10.1007/s11557-019-01491-5.

27. Coelho MA, David-Palma M, Shea T, Bowers K, McGinley-Smith S, Mohammad AW, et al. Comparative genomics of the closely related fungal genera *Cryptococcus* and *Kwoniella* reveals karyotype dynamics and suggests evolutionary mechanisms of pathogenesis. PLOS Biol. 2024;22(6):e3002682. Epub 2024/06/06. doi: 10.1371/journal.pbio.3002682. PubMed PMID: 38843310; PubMed Central PMCID: PMC11185503 declared that no competing interests exist.

28. Yurkov AM, Rohl O, Pontes A, Carvalho C, Maldonado C, Sampaio JP. Local climatic conditions constrain soil yeast diversity patterns in Mediterranean forests, woodlands and scrub biome. FEMS Yeast Res. 2016;16(1):fov103. doi: 10.1093/femsyr/fov103. PubMed PMID: 26568202.

29. Steenkamp ET, Wingfield MJ, McTaggart AR, Wingfield BD. Fungal species and their boundaries matter – Definitions, mechanisms and practical implications. Fungal Biol Rev. 2018;32(2):104-16. doi: https://doi.org/10.1016/j.fbr.2017.11.002.

30. James TY, Stajich JE, Hittinger CT, Rokas A. Toward a Fully Resolved Fungal Tree of Life. Annu Rev Microbiol. 2020;74:291-313. Epub 2020/07/15. doi: 10.1146/annurev-micro-022020-051835. PubMed PMID: 32660385.

31. Libkind D, Cadez N, Opulente DA, Langdon QK, Rosa CA, Sampaio JP, et al. Towards yeast taxogenomics: lessons from novel species descriptions based on complete genome sequences. FEMS Yeast Res. 2020;20(6). Epub 2020/07/28. doi: 10.1093/femsyr/foaa042. PubMed PMID: 32710773.

32. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics. 2013;14:60. Epub 2013/02/26. doi: 10.1186/1471-2105-14-60. PubMed PMID: 23432962; PubMed Central PMCID: PMC3665452.

33. Liu F, Hu ZD, Yurkov A, Chen XH, Bao WJ, Ma Q, et al. *Saccharomycetaceae*: delineation of fungal genera based on phylogenomic analyses, genomic relatedness indices and genomics-based synapomorphies. Persoonia. 2024;52:1-21. Epub 20240406. doi: 10.3767/persoonia.2024.52.01. PubMed PMID: 39161631; PubMed Central PMCID: PMC11319838.

34. Desjardins CA, Giamberardino C, Sykes SM, Yu CH, Tenor JL, Chen Y, et al. Population genomics and the evolution of virulence in the fungal pathogen *Cryptococcus neoformans*. Genome Res. 2017;27(7):1207-19. Epub 2017/06/15. doi: 10.1101/gr.218727.116. PubMed PMID: 28611159; PubMed Central PMCID: PMC5495072.

35. Sun S, Yadav V, Billmyre RB, Cuomo CA, Nowrousian M, Wang L, et al. Fungal genome and mating system transitions facilitated by chromosomal translocations involving intercentromeric recombination. PLOS Biol. 2017;15(8):e2002527. Epub 2017/08/12. doi: 10.1371/journal.pbio.2002527. PubMed PMID: 28800596; PubMed Central PMCID: PMC5568439.

36. Yadav V, Sun S, Coelho MA, Heitman J. Centromere scission drives chromosome shuffling and reproductive isolation. Proc Natl Acad Sci U S A. 2020;117(14):7917-28. Epub 2020/03/21. doi: 10.1073/pnas.1918659117. PubMed PMID: 32193338; PubMed Central PMCID: PMC7149388.

37. Statzell-Tallman A, Belloch C, Fell JW. *Kwoniella mangroviensis* gen. nov., sp.nov. (*Tremellales*, *Basidiomycota*), a teleomorphic yeast from mangrove habitats in the Florida Everglades and Bahamas. FEMS Yeast Res. 2008;8(1):103-13. Epub 2007/10/27. doi: 10.1111/j.1567-1364.2007.00314.x. PubMed PMID: 17961172.

38. Metin B, Findley K, Heitman J. The mating type locus (*MAT*) and sexual reproduction of *Cryptococcus heveanensis*: insights into the evolution of sex and sex-determining chromosomal regions in fungi. PLOS Genet. 2010;6(5):e1000961. Epub 2010/05/27. doi: 10.1371/journal.pgen.1000961. PubMed PMID: 20502678; PubMed Central PMCID: PMC2873909.

39. Guerreiro MA, Springer DJ, Rodrigues JA, Rusche LN, Findley K, Heitman J, et al. Molecular and genetic evidence for a tetrapolar mating system in the basidiomycetous yeast *Kwoniella mangrovensis* and two novel sibling species. Eukaryot Cell. 2013;12(5):746-60. Epub 2013/03/26. doi: 10.1128/ec.00065-13. PubMed PMID: 23524993; PubMed Central PMCID: Pmc3647770.

40. Janbon G, Ormerod KL, Paulet D, Byrnes EJ, 3rd, Yadav V, Chatterjee G, et al. Analysis of the genome and transcriptome of *Cryptococcus neoformans* var. *grubii* reveals complex RNA expression and microevolution

leading to virulence attenuation. PLOS Genet. 2014;10(4):e1004261. Epub 2014/04/20. doi: 10.1371/journal.pgen.1004261. PubMed PMID: 24743168; PubMed Central PMCID: PMC3990503.

41. Yadav V, Sun S, Billmyre RB, Thimmappa BC, Shea T, Lintner R, et al. RNAi is a critical determinant of centromere evolution in closely related fungi. Proc Natl Acad Sci U S A. 2018. Epub 2018/03/07. doi: 10.1073/pnas.1713725115. PubMed PMID: 29507212.

42. Schotanus K, Heitman J. Centromere deletion in *Cryptococcus deuterogattii* leads to neocentromere formation and chromosome fusions. eLife. 2020;9:e56026. Epub 2020/04/21. doi: 10.7554/eLife.56026. PubMed PMID: 32310085; PubMed Central PMCID: PMC7188483.

43. Schotanus K, Yadav V, Heitman J. Epigenetic dynamics of centromeres and neocentromeres in *Cryptococcus deuterogattii*. PLOS Genet. 2021;17(8):e1009743. Epub 2021/09/01. doi: 10.1371/journal.pgen.1009743. PubMed PMID: 34464380; PubMed Central PMCID: PMC8407549.

44. Sankaranarayanan SR, Ianiri G, Coelho MA, Reza MH, Thimmappa BC, Ganguly P, et al. Loss of centromere function drives karyotype evolution in closely related *Malassezia* species. eLife. 2020;9:e53944. Epub 2020/01/21. doi: 10.7554/eLife.53944. PubMed PMID: 31958060; PubMed Central PMCID: PMC7025860.

45. Coelho MA, Ianiri G, David-Palma M, Theelen B, Goyal R, Narayanan A, et al. Frequent transitions in mating-type locus chromosomal organization in *Malassezia* and early steps in sexual reproduction. Proc Natl Acad Sci U S A. 2023;120(32):e2305094120. Epub 2023/07/31. doi: 10.1073/pnas.2305094120. PubMed PMID: 37523560; PubMed Central PMCID: PMC10410736.

46. Narayanan A, Reza MH, Sanyal K. Behind the scenes: Centromere-driven genomic innovations in fungal pathogens. PLOS Pathog. 2024;20(3):e1012080. Epub 2024/03/28. doi: 10.1371/journal.ppat.1012080. PubMed PMID: 38547101; PubMed Central PMCID: PMC10977804.

47. Zhu X, Williamson PR. Role of laccase in the biology and virulence of *Cryptococcus neoformans*. FEMS Yeast Res. 2004;5(1):1-10. doi: 10.1016/j.femsyr.2004.04.004. PubMed PMID: 15381117.

48. Williamson PR. Biochemical and molecular characterization of the diphenol oxidase of *Cryptococcus neoformans*: identification as a laccase. J Bacteriol. 1994;176(3):656-64. doi: 10.1128/jb.176.3.656-664.1994. PubMed PMID: 8300520; PubMed Central PMCID: PMC205102.

49. Gostincar C. Towards genomic criteria for delineating fungal species. J Fungi (Basel). 2020;6(4). Epub 2020/10/30. doi: 10.3390/jof6040246. PubMed PMID: 33114441; PubMed Central PMCID: PMC7711752.

50. Lachance MA, Lee DK, Hsiang T. Delineating yeast species with genome average nucleotide identity: a calibration of ANI with haplontic, heterothallic *Metschnikowia* species. Antonie Van Leeuwenhoek. 2020;113(12):2097-106. Epub 20201013. doi: 10.1007/s10482-020-01480-9. PubMed PMID: 33048250.

51. Lucking R, Aime MC, Robbertse B, Miller AN, Aoki T, Ariyawansa HA, et al. Fungal taxonomy and sequence-based nomenclature. Nat Microbiol. 2021;6(5):540-8. Epub 20210426. doi: 10.1038/s41564-021-00888-x. PubMed PMID: 33903746; PubMed Central PMCID: PMC10116568.

52. Boekhout T, Aime MC, Begerow D, Gabaldon T, Heitman J, Kemler M, et al. The evolving species concepts used for yeasts: from phenotypes and genomes to speciation networks. Fungal Divers. 2021;109(1):27-55. Epub 20210626. doi: 10.1007/s13225-021-00475-9. PubMed PMID: 34720775; PubMed Central PMCID: PMC8550739.

53. Billmyre RB, Croll D, Li W, Mieczkowski P, Carter DA, Cuomo CA, et al. Highly recombinant VGII *Cryptococcus gattii* population develops clonal outbreak clusters through both sexual macroevolution and asexual microevolution. mBio. 2014;5(4):e01494-14. Epub 2014/07/31. doi: 10.1128/mBio.01494-14. PubMed PMID: 25073643; PubMed Central PMCID: PMC4128362.

54. Kobayashi Y, Kayamori A, Aoki K, Shiwa Y, Matsutani M, Fujita N, et al. Chromosome-level genome assemblies of *Cutaneotrichosporon* spp. (Trichosporonales, Basidiomycota) reveal imbalanced evolution between nucleotide sequences and chromosome synteny. BMC Genomics. 2023;24(1):609. Epub 2023/10/12. doi: 10.1186/s12864-023-09718-2. PubMed PMID: 37821828; PubMed Central PMCID: PMC10568926.

55. Greig D, Travisano M, Louis EJ, Borts RH. A role for the mismatch repair system during incipient speciation in *Saccharomyces*. J Evol Biol. 2003;16(3):429-37. Epub 2003/11/26. PubMed PMID: 14635842.

56. Liti G, Barton DB, Louis EJ. Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. Genetics. 2006;174(2):839-50. Epub 2006/09/05. doi: 10.1534/genetics.106.062166. PubMed PMID: 16951060; PubMed Central PMCID: PMC1602076.

57. Priest SJ, Coelho MA, Mixao V, Clancey SA, Xu Y, Sun S, et al. Factors enforcing the species boundary between the human pathogens *Cryptococcus neoformans* and *Cryptococcus deneoformans*. PLOS Genet. 2021;17(1):e1008871. Epub 2021/01/20. doi: 10.1371/journal.pgen.1008871. PubMed PMID: 33465111; PubMed Central PMCID: PMC7846113.

58. Priest SJ, Yadav V, Roth C, Dahlmann TA, Kück U, Magwene PM, et al. Uncontrolled transposition following RNAi loss causes hypermutation and antifungal drug resistance in clinical isolates of *Cryptococcus neoformans*. Nat Microbiol. 2022;7(8):1239-51. Epub 2022/08/03. doi: 10.1038/s41564-022-01183-z. PubMed PMID: 35918426.

59. Seidl MF, Kramer HM, Cook DE, Fiorin GL, van den Berg GCM, Faino L, et al. Repetitive elements contribute to the diversity and evolution of centromeres in the fungal genus *Verticillium*. mBio. 2020;11(5). Epub 2020/09/10. doi: 10.1128/mBio.01714-20. PubMed PMID: 32900804; PubMed Central PMCID: PMC7482064.

60. Ola M, O'Brien CE, Coughlan AY, Ma Q, Donovan PD, Wolfe KH, et al. Polymorphic centromere locations in the pathogenic yeast *Candida parapsilosis*. Genome Res. 2020;30(5):684-96. Epub 2020/05/20. doi: 10.1101/gr.257816.119. PubMed PMID: 32424070; PubMed Central PMCID: PMC7263194.

61. Bergeron J, Drouin G. The evolution of 5S ribosomal RNA genes linked to the rDNA units of fungal species. Curr Genet. 2008;54(3):123-31. Epub 2008/08/16. doi: 10.1007/s00294-008-0201-2. PubMed PMID: 18704425.

62. McStay B, Grummt I. The epigenetics of rRNA genes: from molecular to chromosome biology. Annu Rev Cell Dev Biol. 2008;24:131-57. Epub 2008/07/12. doi: 10.1146/annurev.cellbio.24.110707.175259. PubMed PMID: 18616426.

63. Rooney AP, Ward TJ. Evolution of a large ribosomal RNA multigene family in filamentous fungi: birth and death of a concerted evolution paradigm. Proc Natl Acad Sci U S A. 2005;102(14):5084-9. doi: 10.1073/pnas.0409689102. PubMed PMID: 15784739; PubMed Central PMCID: PMC555991.

64. Kapitonov VV, Jurka J. A novel class of SINE elements derived from 5S rRNA. Mol Biol Evol. 2003;20(5):694-702. Epub 2003/04/08. doi: 10.1093/molbev/msg075. PubMed PMID: 12679554.

65. Feng B, Li Y, Xu B, Liu H, Steenwyk JL, David KT, et al. Unique trajectory of gene family evolution from genomic analysis of nearly all known species in an ancient yeast lineage. Mol Syst Biol. 2025. Epub 2025/05/28. doi: 10.1038/s44320-025-00118-0. PubMed PMID: 40425814.

66. David KT, Schraiber JG, Crandall JG, Labella AL, Opulente DA, Harrison MC, et al. Convergent expansions of keystone gene families drive metabolic innovation in Saccharomycotina yeasts. Proc Natl Acad Sci U S A. 2025;122(23):e2500165122. Epub 2025/06/03. doi: 10.1073/pnas.2500165122. PubMed PMID: 40460114.

67. Missall TA, Moran JM, Corbett JA, Lodge JK. Distinct stress responses of two functional laccases in *Cryptococcus neoformans* are revealed in the absence of the thiol-specific antioxidant Tsa1. Eukaryot Cell.

2005;4(1):202-8. Epub 2005/01/12. doi: 10.1128/EC.4.1.202-208.2005. PubMed PMID: 15643075; PubMed Central PMCID: PMC544170.

68. Waterman SR, Hacham M, Panepinto J, Hu G, Shin S, Williamson PR. Cell wall targeting of laccase of *Cryptococcus neoformans* during infection of mice. Infect Immun. 2007;75(2):714-22. Epub 2006/11/15. doi: 10.1128/IAI.01351-06. PubMed PMID: 17101662; PubMed Central PMCID: PMC1828480.

69. Pukkila-Worley R, Gerrald QD, Kraus PR, Boily MJ, Davis MJ, Giles SS, et al. Transcriptional network of multiple capsule and melanin genes governed by the *Cryptococcus neoformans* cyclic AMP cascade. Eukaryot Cell. 2005;4(1):190-201. Epub 2005/01/12. doi: 10.1128/EC.4.1.190-201.2005. PubMed PMID: 15643074; PubMed Central PMCID: PMC544166.

70. Rokas A. Evolution of the human pathogenic lifestyle in fungi. Nat Microbiol. 2022;7(5):607-19. Epub 20220504. doi: 10.1038/s41564-022-01112-0. PubMed PMID: 35508719; PubMed Central PMCID: PMC9097544.

71. Gusa A, Williams JD, Cho JE, Averette AF, Sun S, Shouse EM, et al. Transposon mobilization in the human fungal pathogen *Cryptococcus* is mutagenic during infection and promotes drug resistance in vitro. Proc Natl Acad Sci U S A. 2020;117(18):9973-80. Epub 2020/04/19. doi: 10.1073/pnas.2001451117. PubMed PMID: 32303657; PubMed Central PMCID: PMC7211991.

72. Gusa A, Yadav V, Roth C, Williams JD, Shouse EM, Magwene P, et al. Genome-wide analysis of heat stress-stimulated transposon mobility in the human fungal pathogen *Cryptococcus deneoformans*. Proc Natl Acad Sci U S A. 2023;120(4):e2209831120. Epub 2023/01/21. doi: 10.1073/pnas.2209831120. PubMed PMID: 36669112; PubMed Central PMCID: PMC9942834.

73. Mackey AI, Fraunfelter V, Shaltz S, McCormick J, Schroeder C, Perfect JR, et al. Temperature and genetic background drive mobilization of diverse transposable elements in a critical human fungal pathogen. bioRxiv. 2025. Epub 2025/06/06. doi: 10.1101/2025.05.19.654958. PubMed PMID: 40475665; PubMed Central PMCID: PMC12139950.

74. Sun S, Priest SJ, Heitman J. *Cryptococcus neoformans* mating and genetic crosses. Curr Protoc Microbiol. 2019:e75. Epub 2019/01/21. doi: 10.1002/cpmc.75. PubMed PMID: 30661293.

75. Peterson PP, Choi JT, Fu C, Cowen LE, Sun S, Bahn YS, et al. The *Cryptococcus neoformans* STRIPAK complex controls genome stability, sexual development, and virulence. PLOS Pathog. 2024;20(11):e1012735. Epub 2024/11/19 22:19. doi: 10.1371/journal.ppat.1012735. PubMed PMID: 39561188; PubMed Central PMCID: PMC11614259.

76. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 2017;27(5):722-36. Epub 2017/03/17. doi: 10.1101/gr.215087.116. PubMed PMID: 28298431; PubMed Central PMCID: PMC5411767.

77. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19(5):455-77. Epub 2012/04/18. doi: 10.1089/cmb.2012.0021. PubMed PMID: 22506599; PubMed Central PMCID: PMC3342519.

78. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30(4):772-80. Epub 2013/01/19. doi: 10.1093/molbev/mst010. PubMed PMID: 23329690; PubMed Central PMCID: PMC3603318.

79. Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics. 2009;25(15):1972-3. Epub 2009/06/10. doi: 10.1093/bioinformatics/btp348. PubMed PMID: 19505945; PubMed Central PMCID: Pmc2712344.

80. Steenwyk JL, Buida TJ, Labella AL, Li Y, Shen XX, Rokas A. PhyKIT: a broadly applicable UNIX shell toolkit for processing and analyzing phylogenomic data. Bioinformatics. 2021;37(16):2325-31. Epub 2021/02/10. doi: 10.1093/bioinformatics/btab096. PubMed PMID: 33560364; PubMed Central PMCID: PMC8388027.

81. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020;37(5):1530-4. Epub 2020/02/06. doi: 10.1093/molbev/msaa015. PubMed PMID: 32011700; PubMed Central PMCID: PMC7182206.

82. Yoon SH, Ha SM, Lim J, Kwon S, Chun J. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek. 2017;110(10):1281-6. Epub 2017/02/17. doi: 10.1007/s10482-017-0844-4. PubMed PMID: 28204908.

83. Kim D, Park S, Chun J. Introducing EzAAI: a pipeline for high throughput calculations of prokaryotic average amino acid identity. J Microbiol. 2021;59(5):476-80. Epub 2021/04/29. doi: 10.1007/s12275-021-1154-0. PubMed PMID: 33907973.

84. Coombe L, Kazemi P, Wong J, Birol I, Warren RL. Multi-genome synteny detection using minimizer graph mappings. bioRxiv. 2024:2024.02.07.579356. doi: 10.1101/2024.02.07.579356.

85. Coombe L, Warren RL, Birol I. ntSynt-viz: Visualizing synteny patterns across multiple genomes. bioRxiv. 2025:2025.01.15.633221. doi: 10.1101/2025.01.15.633221.

86. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics. 2011;27(7):1009-10. Epub 2011/02/01. doi: 10.1093/bioinformatics/btr039. PubMed PMID: 21278367; PubMed Central PMCID: PMC3065679.

87. Bruen TC, Philippe H, Bryant D. A simple and robust statistical test for detecting the presence of recombination. Genetics. 2006;172(4):2665-81. Epub 2006/02/21. doi: 10.1534/genetics.105.048975. PubMed PMID: 16489234; PubMed Central PMCID: PMC1456386.

88. Kurtzman CP, Fell JW, Boekhout T, Robert V. Methods for isolation, phenotypic characterization and maintenance of yeasts. In: Kurtzman CP, Fell JW, Boekhout T, editors. The Yeasts: A Taxonomic Study. 5th ed. London: Elsevier; 2011. p. 87-110.

89. Upadhya R, Probst C, Alspaugh JA, Lodge JK. Measuring Stress Phenotypes in *Cryptococcus neoformans*. Methods Mol Biol. 2024;2775:277-303. Epub 2024/05/17. doi: 10.1007/978-1-0716-3722-7_19. PubMed PMID: 38758325; PubMed Central PMCID: PMC11521573.

Figures

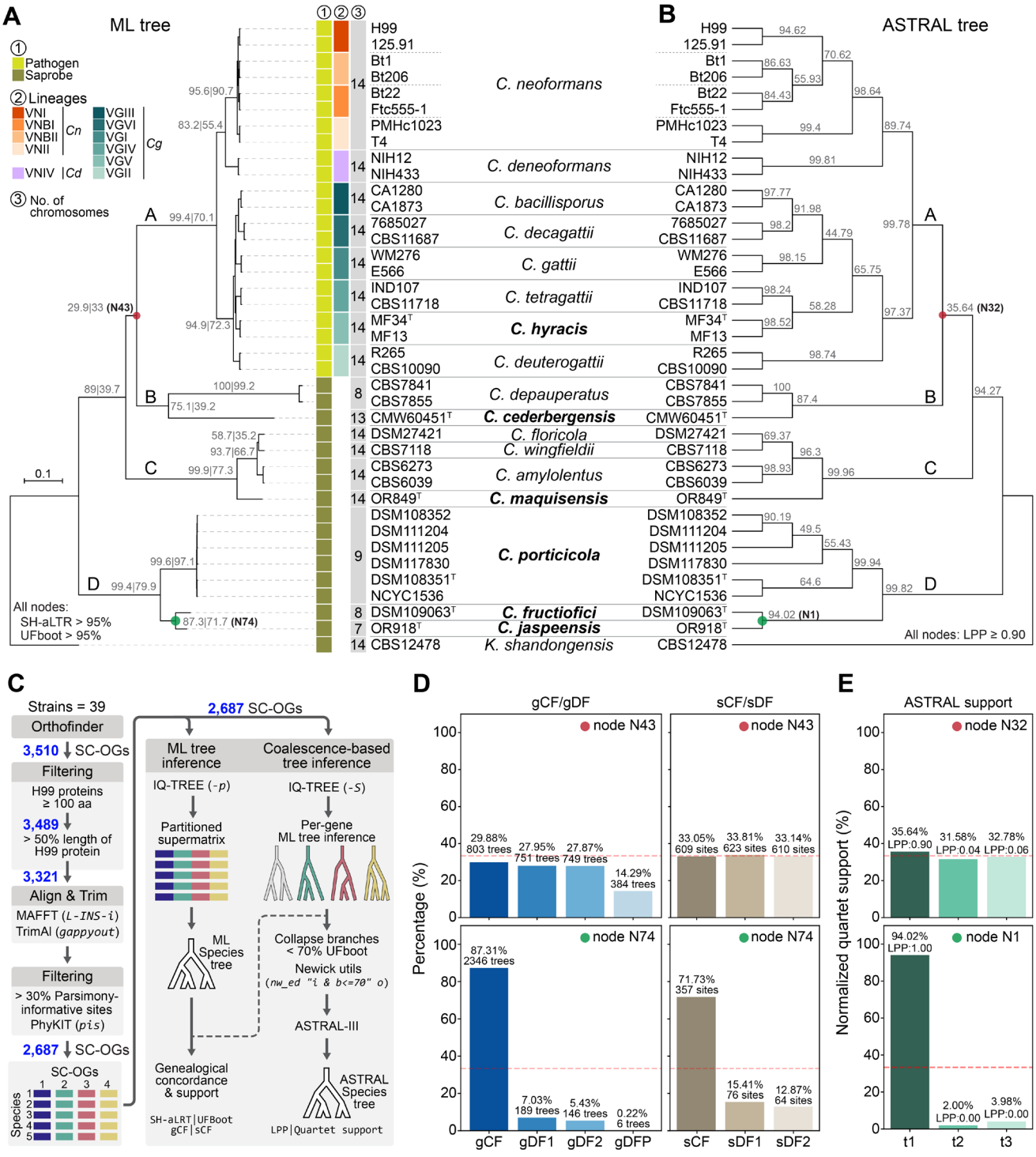
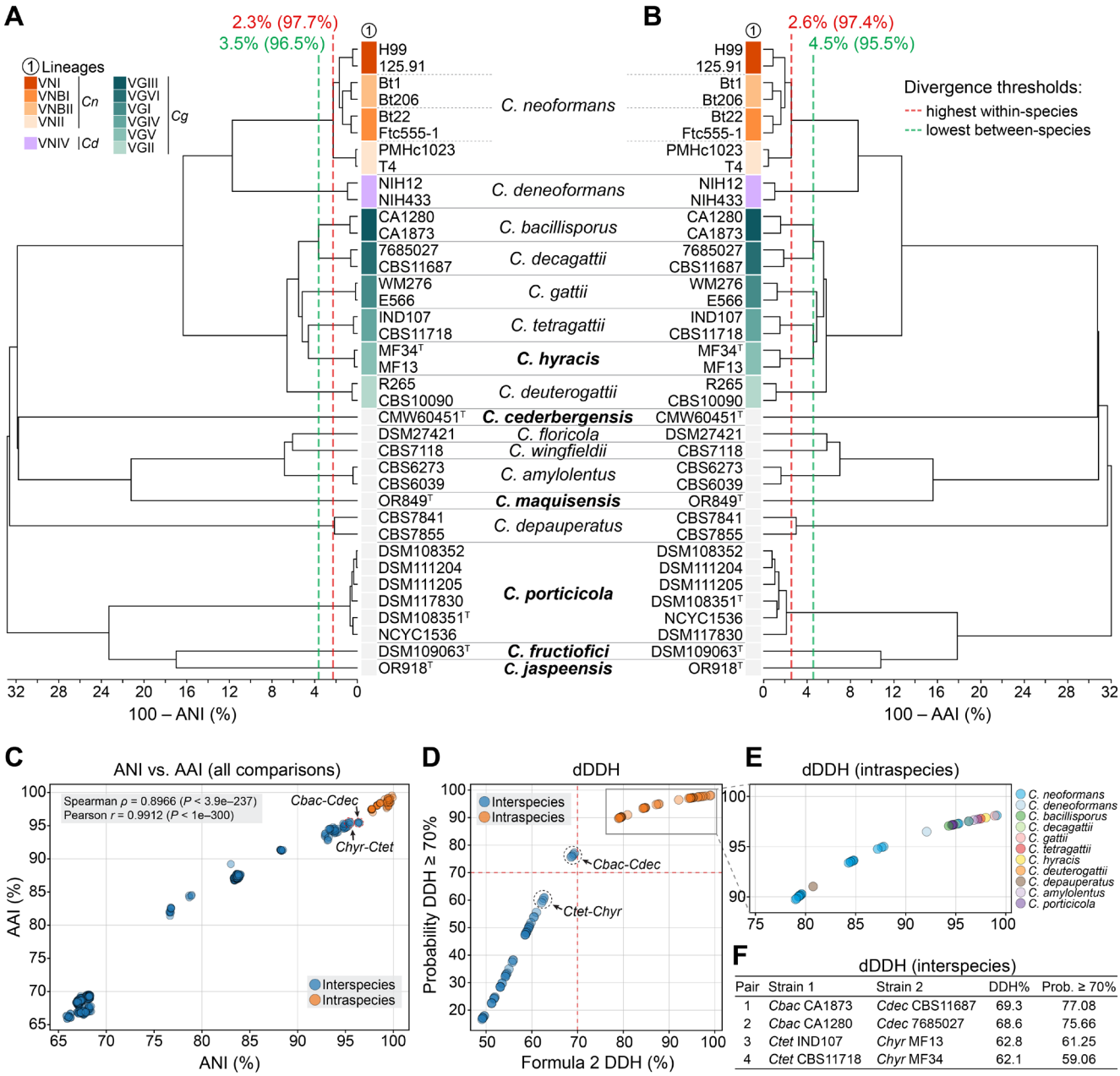


Figure 1. Phylogenomic relationships and species delimitation within *Cryptococcus*. (A) Maximum likelihood (ML) phylogeny inferred from a concatenated data matrix comprising protein alignments of 2,687 single-copy orthologs (SC-OGs) across 39 strains, representing 17 *Cryptococcus* species (including six newly proposed species indicated in boldface) and rooted with the outgroup *Kwoniella shandongensis*. Gene and site concordance

factors (gCF|sCF) are shown in grey at nodes separating species and major clades. A red circle highlights a conflict-prone node (N43), while a green circle marks an example of a well-supported node (N74). Support values from SH-aLRT and ultrafast bootstrap (UFboot) exceed 95% for all nodes and were omitted for clarity but are available in the corresponding tree file (https://doi.org/10.5281/zenodo.15686462). Branch lengths are shown as the number of substitutions per site (scale bar). Major clades are labeled as partitions A–D. For reference, lineages designations are provided for *C. neoformans* (*Cn*: VNI, VNBI, VNBII, VNII), *C. deneoformans* (*Cd*: VNIV) and *C. gattii* complex (*Cg*: VGI–VGVI), and chromosome numbers are indicated for each species. **(B)** Species tree inferred with ASTRAL, which summarizes gene tree topologies under a multispecies coalescent model, using the same set of 2,687 single-copy orthologs (SC-OGs). Colored circles indicate the corresponding nodes highlighted in panel A, with red marking a conflict-prone node (N32) and green marking an example of a well-supported node (N1). Support values in grey represent the quartet support for the main topology (q1). Local posterior probabilities (LPP) are ≥ 0.90 for all nodes and are not shown for simplicity but are available in the corresponding tree file (https://doi.org/10.5281/zenodo.15686462). Major clades are labeled as partitions A–D. **(C)** Overview of the phylogenomic workflow. A total of 3,510 single-copy orthologs (SC-OGs) were initially identified using OrthoFinder across 39 genomes. These were filtered based on alignment length, sequence completeness, and phylogenetic informativeness, resulting in 2,687 high-confidence SC-OGs. Protein alignments were used to infer individual gene trees and a concatenated supermatrix. The resulting datasets were analyzed with maximum likelihood (ML) methods (panel A) and the coalescent-based ASTRAL approach (panel B). **(D)** Gene and site concordance at two representative nodes: N43 (conflict-prone) and N74 (well-supported). Bar plots on the left show gene concordance factors (gCF) and gene discordance factors (gDF1, gDF2, and gDFP), representing the proportion of gene trees that support the main topology (gCF, dark blue), each of the two alternative topologies (gDF1 and gDF2, light blue), or that display conflicting relationships due to paraphyly of one or more clades (gDFP, pale blue). The corresponding site concordance (sCF) and site discordance (sDF1 and sDF2) values are shown on the right, based on site-wise likelihood support for each topology. The red dashed line at 33.3% indicates the expected support under a hard polytomy, where all three resolutions are equally likely due to lack of phylogenetic signal. **(E)** ASTRAL quartet support for the same nodes shown in panel D: N32 (conflict-prone) and N1 (well-supported). Bar plots show the normalized frequencies of the three possible quartet topologies around each node, based on gene tree topologies. The main topology is shown in dark green, with alternatives in lighter shades. Local posterior probabilities (LPP) for each topology are indicated above the bars. The red dashed line (33.3%) reflects the expected distribution under a hard polytomy.



Pair	Strain 1	Strain 2	DDH%	Prob. ≥ 70%
1	*Cbac* CA1873	*Cdec* CBS11687	69.3	77.08
2	*Cbac* CA1280	*Cdec* 7685027	68.6	75.66
3	*Ctet* IND107	*Chyr* MF13	62.8	61.25
4	*Ctet* CBS11718	*Chyr* MF34	62.1	59.06

Figure 2. Genomic divergence and species boundaries among *Cryptococcus* species. (A–B) UPGMA clustering based on pairwise sequence divergence, calculated as 100 – Average Nucleotide Identity (ANI) in panel A and 100 – Average Amino Acid Identity (AAI) in panel B. Vertical dashed lines indicate divergence thresholds: red marks the highest observed divergence within a species (i.e., the lowest ANI or AAI between *C. neoformans* lineages), and green marks the lowest observed divergence between species (i.e., the highest ANI or AAI between *C. bacillisporus* and *C. decagattii*). **(C)** Scatterplot comparing ANI and AAI values across all pairwise genome comparisons. Intraspecies comparisons are shown in orange and interspecies comparisons in blue. The two interspecies pairs with the highest genomic similarity are highlighted: *C. bacillisporus* (*Cbac*) vs. *C. decagattii* (*Cdec*), and *C. hyracis* (*Chyr*) vs. *C. tetragattii* (*Ctet*). Pearson and Spearman correlation coefficients indicate a strong positive relationship between ANI and AAI values across all comparisons. **(D)** Digital DNA–DNA hybridization (dDDH) estimates for pairwise genome comparisons. The x-axis shows values derived from *Formula 2* of the Genome-to-Genome Distance Calculator (GGDC), which estimates the expected wet-lab DDH percentage based on genome BLAST comparisons. The y-axis shows the probability that the true DDH value exceeds 70%,

a widely accepted threshold for species demarcation. Dashed lines indicate the 70% threshold for both axes. Intra-species comparisons are shown in orange and inter-species comparisons in blue. **(E)** Zoomed-in view on intraspecies dDDH comparisons, grouped by species. **(F)** Table of dDDH values for select interspecies comparisons that approach or exceed the 70% threshold, including the genome pairs with the highest interspecific scores. *C. bacillisporus*–*C. decagattii* comparisons fall just below the 70% cutoff but have high probabilities (>75%) of exceeding it, indicating borderline genomic divergence despite their current recognition as separate species. In contrast, *C. tetragattii*–*C. hyracis* comparisons fall well below both thresholds, supporting their recognition as distinct species.

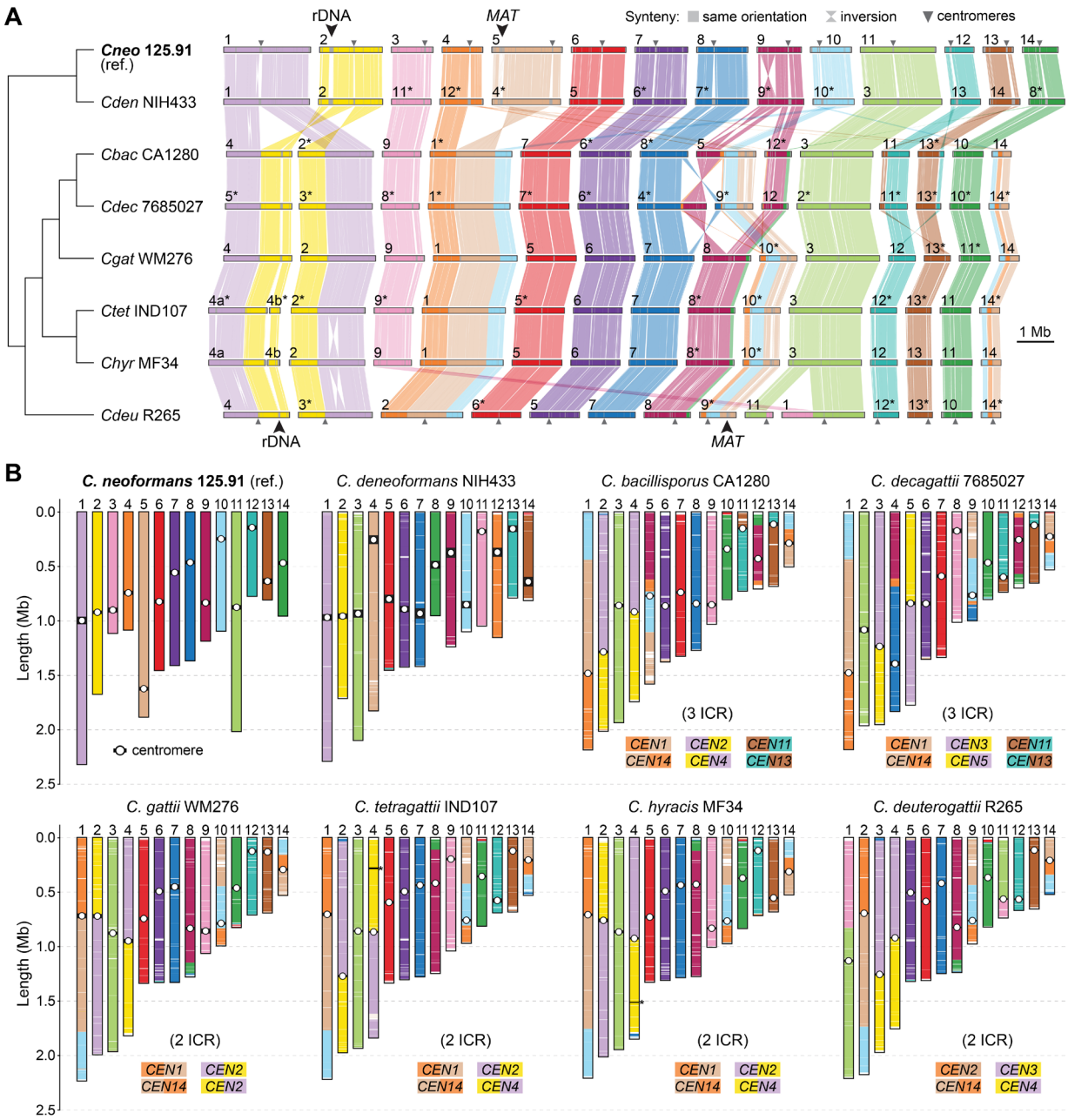


Figure 3. Chromosomal rearrangements among clade A pathogenic *Cryptococcus* species. (A) Synteny comparisons between *C. neoformans* strain 125.91 (reference) and representative strains from 7 other clade A species (8 species total). Rearrangements, including large inversions and translocations, were identified using the ntSynt toolkit and visualized with ntSynt-viz. All species maintain 14 chromosomes, but some differences in chromosomal structure are evident. Chromosomes were reordered to maximize collinearity, and those inverted relative to their original assembly orientations are marked with asterisks. The positions of the mating-type locus (*MAT*) and the rRNA locus are also indicated. **(B)** Schematic comparison of chromosome organization across the same strains, highlighting large-scale translocations and centromere locations. Chromosomes are color-coded according to their correspondence with the *C. neoformans* reference. Putative intercentromeric recombination (ICR) events are indicated where synteny breakpoints coincide with centromeres. Asterisks denote chromosomes comprising two contigs interrupted at the rDNA array. Diagrams were generated using a custom Python (available at https://doi.org/10.5281/zenodo.15686462) based on pairwise alignments from minimap2.

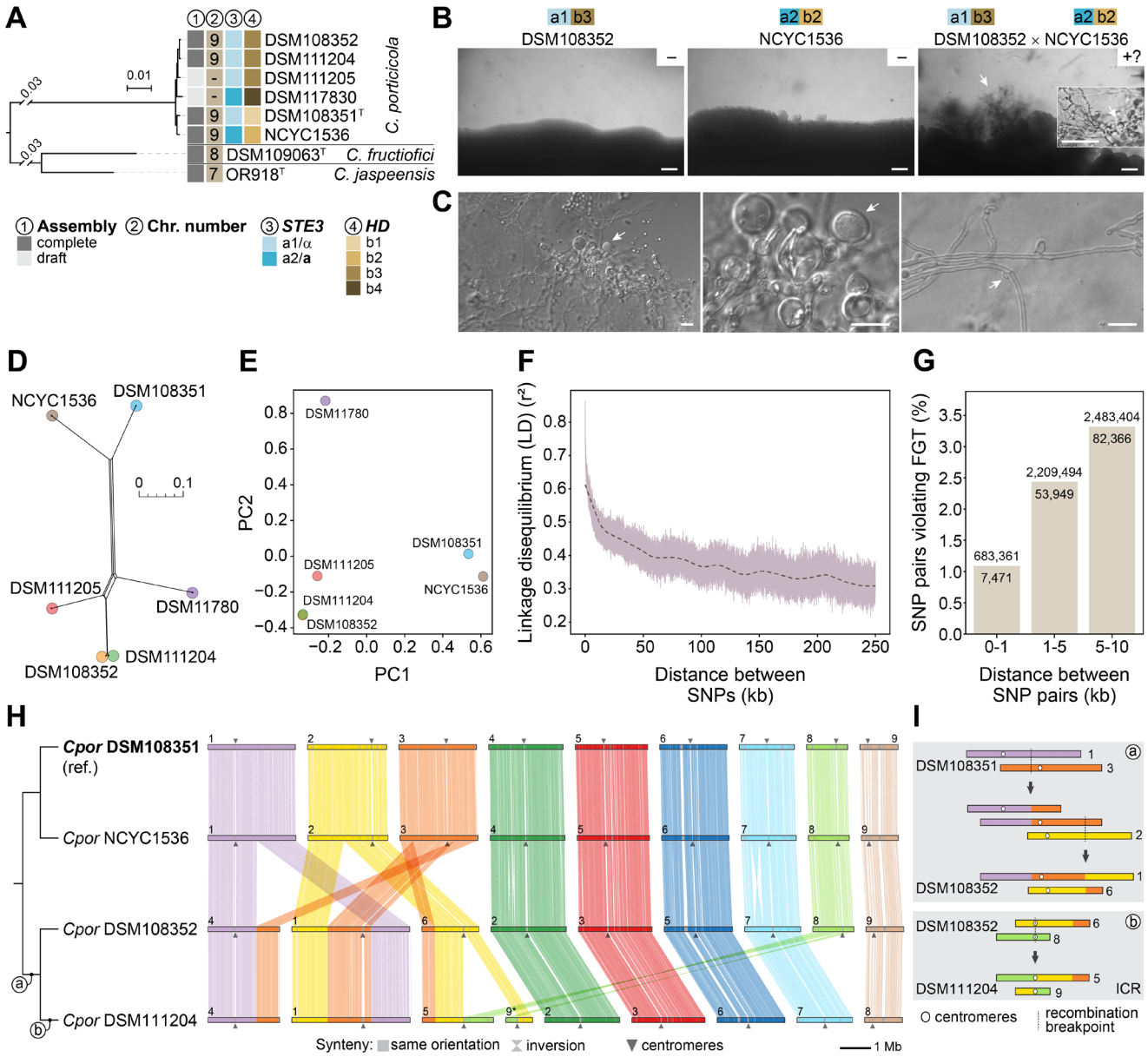


Figure 4. Mating, recombination signatures, and chromosomal structural variation in *Cryptococcus porticicola* (A) Maximum likelihood phylogeny of clade D species (pruned from Fig. 1), highlighting *C. porticicola*, *C. fructiofici*, and *C. jaspeensis*. Labels indicate genome assembly status (complete or draft), chromosome number inferred from telomere-to-telomere assemblies, mating pheromone receptor allele (*STE3* a1/α or *STE3* a2/a), and *HD* allele identity (b1–b4) inferred from sequence comparisons. **(B)** Mating assay between *C. porticicola* strains DSM108352 (*a1b3*) and NCYC1536 (*a2b2*), showing hyphal formation and putative sexual structures. Solo cultures are shown on the left and center panels; the cross is shown on the right, where hyphal formation is observed embedded in the agar. Basidia-like structures are visible and resemble sexual development in *Kwoniella* species. Assays were performed on V8 medium (pH 5.0) at room temperature in the dark for 1 month. Scale bars: 200 µm (50 µm in the inset). **(C)** Light microscopy images of hyphal and basidia-like structures recovered from the mating plate shown in panel B. The left panel shows hyphae and basidia-like cells (arrows), with a zoomed-in view in the middle panel. The right panel highlights a hyphal segment with an apparently unfused clamp connection (arrow). Scale bars: 10 µm. **(D)** Neighbor-Net phylogenetic network constructed from core SNPs showing signals of conflicting phylogenetic relationships (reticulation), consistent with recombination. **(E)** Principal component analysis (PCA) of core SNP data showing genetic structure among the six strains. **(F)** Linkage

disequilibrium (LD) decay curve showing mean pairwise r^2 values (minor allele frequency, MAF ≥ 0.167) as a function of physical distance between SNPs. Dashed line shows lowess-smoothed trend. **(G)** Four-gamete test (FGT) showing increasing proportions of SNP pairs violating the test with physical distance, indicative of recombination. Total SNP pair counts (top numbers) and violating pairs (bottom numbers) are shown for each bin. **(H)** Chromosomal synteny comparisons across *C. porticicola* employing strain DSM108351 as reference. Each block represents a chromosome; ribbons denote orthologous regions as determined by the ntSynt toolkit. Chromosomes were reordered to maximize collinearity, and those inverted relative to their original assembly orientations are marked with asterisks. **(I)** Detailed view of intraspecific chromosomal rearrangements in *C. porticicola*. Centromeres (circles) and synteny breakpoints are shown for each rearrangement. ICR: intercentromeric recombination.

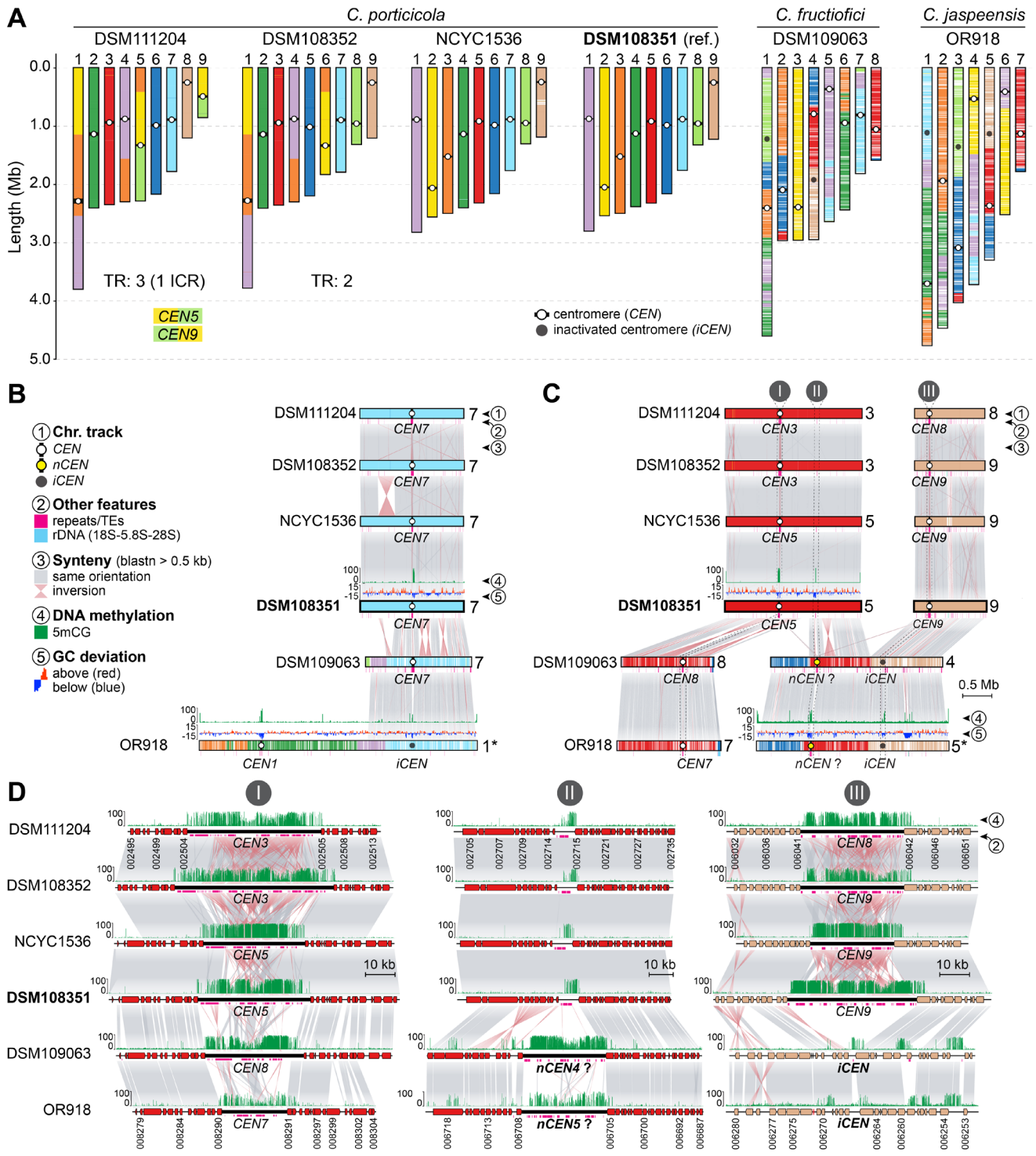


Figure 5. Karyotype evolution and centromere dynamics in clade D *Cryptococcus* species. (A) Chromosome-length plots for six telomere-to-telomere assemblies representing *C. porticicola*, *C. fructiofici*, and *C. jaspeensis*. Colors indicate synteny blocks, with *C. porticicola* DSM108351 as the reference. Chromosome number differs across species (9, 8, and 7, respectively), and multiple structural changes are observed, including species- and strain-specific translocations (TR). Circles mark predicted centromeres: white for active centromeres (*CEN*) and dark gray for putative inactivated centromeres (*iCEN*). For *C. porticicola*, the total number of translocations (including those consistent with intercentromeric recombination, ICR) is noted below each strain.

(B) Example of centromere loss in *C. jaspeensis*. A region syntenic to *CEN7* in *C. porticicola* and *C. fructiofici* appears inactivated in *C. jaspeensis* OR918 (*iCEN* on chr. 1), but still shows 5mCG methylation, (see **Figure S6F** for a zoomed-in view of this region). Track descriptions are provided on the key on the left. **(C–D)** Synteny comparisons showing centromere inactivation and possible neocentromere formation in *C. fructiofici* and *C. jaspeensis*. In panel C, *CEN5* of *C. porticicola* DSM108351 aligns with *CEN8* in *C. fructiofici* and *CEN7* in *C. jaspeensis*, suggesting a shared ancestral origin (zoomed view in panel D, region I). By contrast, *CEN9* in *C. porticicola* appears to have been inactivated (*iCEN*) in both *C. fructiofici* and *C. jaspeensis*, coinciding with reduced transposable element content and diminished 5mCG methylation signal, and conservation of flanking genes aside from a minor inversion (panel D, region III). Notably, the current centromere locations in chr. 4 of *C. fructiofici* and chr. 5 of *C. jaspeensis* lack orthology with any of *C. porticicola* centromeres, but the corresponding region in *C. porticicola* exhibits low levels of transposons and 5mCG methylation (panel D, region II). This raises the possibility that a neocentromere (*nCEN*) emerged at this shared position in the common ancestor of *C. fructiofici* and *C. jaspeensis*, or alternatively, that this region harbored an ancestral centromere that was subsequently lost or inactivated in *C. porticicola*.

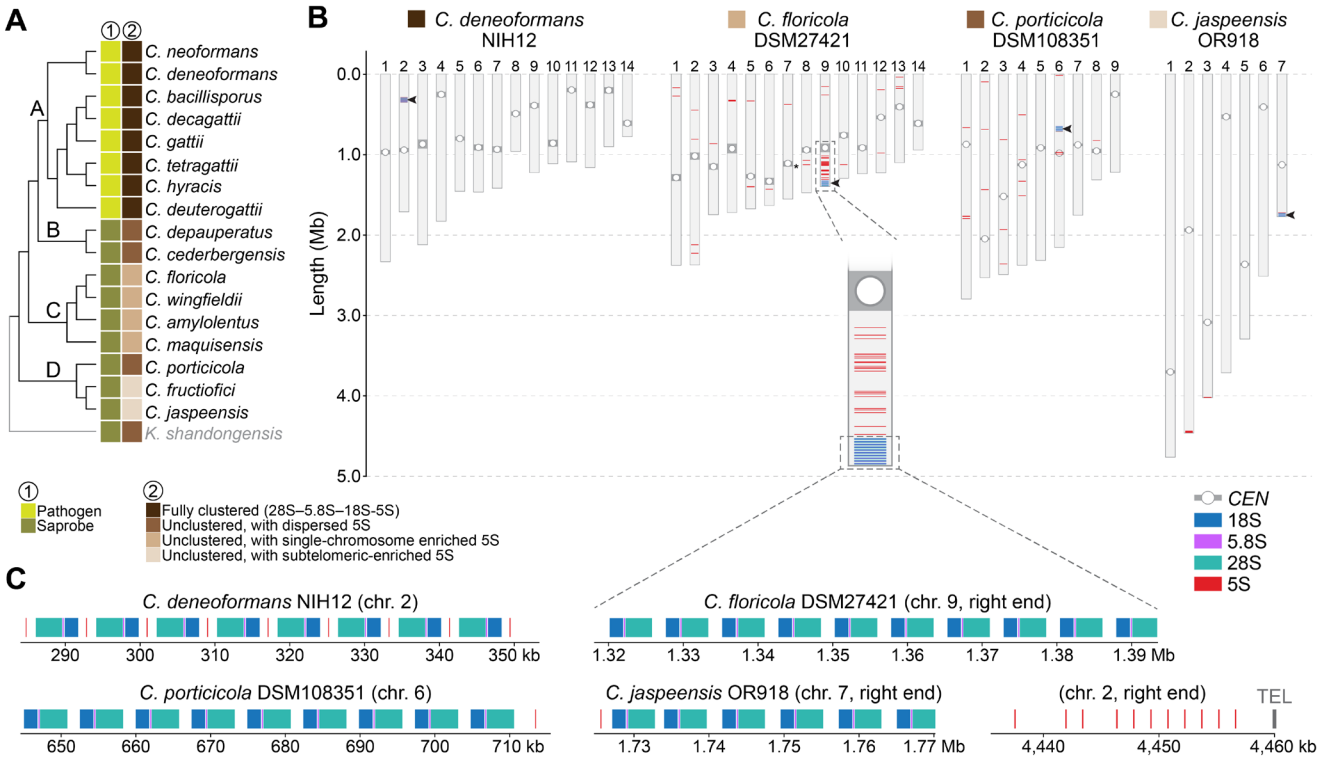


Figure 6. Variation in ribosomal RNA (rRNA) gene organization across *Cryptococcus* species. (A) Cladogram of *Cryptococcus* species (adapted from Figure 1), annotated with rRNA gene organization types. All pathogenic species from clade A exhibit a fully clustered rRNA array (28S–5.8S–18S–5S). In contrast, saprobic species display unclustered configurations with lineage-specific patterns: dispersed 5S genes (e.g., clade B), 5S enrichment on a single chromosome (e.g., clade C), or subtelomeric 5S enrichment (e.g., *C. fructiofici*, *C. jaspeensis* in clade D). **(B)** Genome-wide distribution of rRNA genes (28S, 5.8S, 18S, 5S) across four representative species: *C. deneoformans* (clade A), *C. floricola* (clade C), *C. porticicola*, and *C. jaspeensis* (both clade D). Arrowheads indicate the position of the core rDNA array, and predicted centromeres are marked with open circles. In clade C, 5S genes are significantly enriched on a single chromosome (e.g., chr. 10 in *C. floricola*, see inset), whereas in clade D species, 5S genes are predominantly subtelomeric (e.g., chr. 2 in *C. jaspeensis*). See **Figure S7** for additional species and strains. **(C)** Zoomed views of representative rDNA clusters and subtelomeric 5S regions. TEL, telomere.

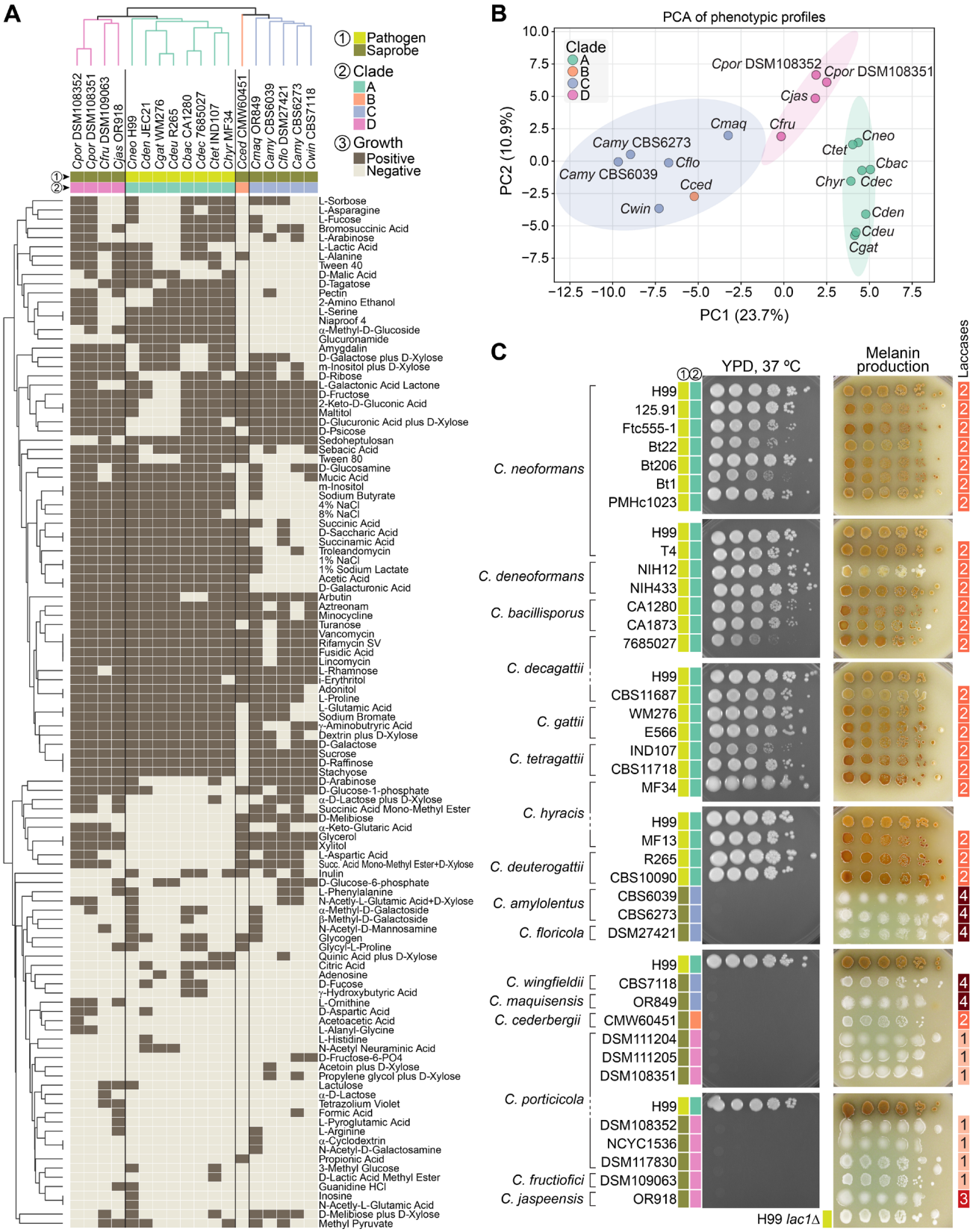


Figure 7. Phenotypic profiling distinguishes *Cryptococcus* clades and pathogenic species. (A) Clustermap showing hierarchical clustering of strains (columns) and informative phenotypic tests (rows), based on binary

growth/no-growth responses in Biolog assays. Of the 153 substrates tested, 113 showed variable responses across strains and were retained for clustering. Growth was assessed using Biolog YT, FF, and GEN III MicroPlates. Binary values indicate positive growth (brown) or absence of growth (beige). Strains are colored by phylogenetic clade, and clustering was performed using average linkage and Euclidean distance. Most strains cluster according to clade, with metabolic profiles broadly reflecting phylogenetic structure. **(B)** Principal component analysis (PCA) of Biolog growth profiles. The first two components explain 34.6% of the total variance. Ellipses represent 95% confidence intervals for each clade, based on a multivariate normal distribution. Most strains separate by clade, with the exception of *C. cederbergensis* (clade B), which clusters within the clade C ellipse, suggesting phenotypic similarity despite phylogenetic divergence. **(C)** Growth at 37 °C (after 4 days) and melanin production on Bird seed agar (after 5 days at room temperature) across all tested strains. All strains from pathogenic species in clade A exhibited robust thermotolerance and melanin synthesis, whereas strains from saprobic species in clades B, C, and D showed neither trait. The rightmost column indicates the predicted number of laccase genes. Several nonpathogenic strains harbor multiple laccase genes, yet failed to produce melanin, indicating that gene copy number alone does not account for trait expression. See **Figure S8C** for growth at other temperatures and results on alternative melanin-inducing media (L-DOPA).