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

Marco A. Coelho[1,§,*], Márcia David-Palma[1,§], Aleksey V. Kachalkin[2,3], Miroslav Kolařík[4], Benedetta Turchetti[5], José Paulo Sampaio[6], Michael J. Wingfield[7], Matthew C. Fisher[8], Andrey M. Yurkov[9], and Joseph Heitman[1*]

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

[2]M.V. Lomonosov Moscow State University, Moscow, Russia

[3]G.K. Skryabin Institute of Biochemistry and Physiology of Microorganisms, PSCBR RAS, Pushchino, Russia

[4]Laboratory of Fungal Genetics and Metabolism, Institute of Microbiology of the Czech Academy of Sciences, Vídeňská 1083, Prague, Czechia

[5]Department of Agricultural, Food and Environmental Sciences, University of Perugia, Perugia, Italy

[6]UCIBIO, i4HB, Departamento de Ciências da Vida, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Caparica, Portugal

[7]Department of Biochemistry, Genetics and Microbiology, Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa

[8]Department of Infectious Disease Epidemiology, Imperial College London, London, United Kingdom

[9]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. 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 accompanied by genomic signatures consistent with centromere inactivation or loss of centromeric identity. 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.

Author Summary

Cryptococcus is a genus of fungi that includes both pathogenic species capable of causing life-threatening infections in humans and many environmental species that inhabit soil, fruit, decaying wood, and insect-associated environments with no known link to disease. Here, we formally describe six previously unnamed species based on genome-wide comparisons, genetic divergence metrics, chromosome structure, reproductive behavior, and metabolic profiling. One of these species belongs to the pathogenic lineage, while the others are distinct environmental taxa, including several recovered from bark beetles or their galleries, indicating repeated occurrence in this ecological niche. For one of these insect-associated species, represented by multiple isolates, we detected signatures of sexual reproduction and genetic exchange in natural populations. This study expands the framework for investigating how certain *Cryptococcus* lineages adapt to new environments or evolve traits relevant to pathogenicity. It also emphasizes that continuous environmental sampling remains essential for capturing fungal diversity and illustrates the power of combining genomic and phenotypic data to define species boundaries and uncover lineages that may otherwise remain undetected.

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]. Early molecular studies revealed that the genus *Cryptococcus* is polyphyletic, encompassing distantly related yeasts within the class *Tremellomycetes* [8, 9]. Subsequent phylogenetic revisions resolved this taxonomic heterogeneity, transferring taxa unrelated to the pathogenic lineage to other genera such as *Naganishia*, *Papiliotrema*, and *Cutaneotrichosporon* [10, 11].

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, 12]. 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 [13]. 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*) [14]. 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 [15, 16]. In contrast, *C. neoformans*, *C. deneoformans*, *C. bacillisporus*, *C. tetragattii*, and *C. decagattii* are more often associated with infections in immunocompromised hosts [14, 17, 18]. 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 [19]. Although these species are genetically distinct, hybrids have been documented, particularly between *C. neoformans* and *C. deneoformans* (referred to as AD hybrids, alluding to their different serotypes), and more rarely between *C. neoformans* and members of the *C. gattii* complex. However, these hybrids are typically diploid or aneuploid and exhibit reduced fertility and low spore germination rates (<20%), reflecting postzygotic barriers [14, 20, 21].

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 [22], 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) [14, 17, 18], reinforcing the need for accurate species-level distinctions in both public health and evolutionary research.

Beyond the pathogenic species, *Cryptococcus* harbors several nonpathogenic taxa. Some have been recognized for decades—including *C. wingfieldii*, *C. amylolentus*, *C. depauperatus*, and *C. luteus*

[11, 23-25]—while others, such as *C. floricola* [26], have been more recently discovered. 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 [27-29]. 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* (*Tremellomycetes*) [11, 30].

Building on this foundation, recent work and sequence database mining have revealed additional deeply divergent *Cryptococcus* lineages that remain undescribed [30-32]. 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 [30, 32]. 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 [30, 32]. 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 (**Fig 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. Strains were chosen based on the availability of complete genome assemblies, inclusion of opposite mating types where possible, and representation of divergent subclades within each lineage as previously defined [33]. 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 (**Fig 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 (**Figs 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 [34]. 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 occurrence in 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 ficicola* 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 **S1 Table** 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 **Fig 1A**; node N32 in **Fig 1B**) exhibited comparatively lower gene and site concordance factors (gCF = 29.9%, sCF = 33.05%; **Fig 1D**), and lower quartet support (35.64%; **Fig 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 (**S1A–B Figs**). 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%) (**S1C Fig**). Likewise, site concordance factors (sCF) only modestly favored the main topology. ASTRAL quartet support for the corresponding nodes (N28, N23, N16; **S1D Fig**) 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 [35, 36].

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. [19], 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 [14], establishing a precedent for detailed taxonomic recognition within the complex. Because a subsequent publication expressed caution regarding the stability and interpretation of species boundaries in this group [22], the VGV lineage remained unnamed despite

clear evidence of genetic divergence. To maintain consistency with the framework proposed by Hagen et al. [14, 17], and 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) (**Figs 1A–B** and **S1A–B**). The two species were resolved as sister groups in the ASTRAL tree with moderate quartet support (58.3%, node N23; **S1B** and **S1D Figs**), and gene concordance fell just below the 33% threshold for inferring a dominant phylogenetic signal (gCF = 32.7%, node N52; **S1A** and **S1C Figs**), 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 [37], are increasingly applied to yeast systematics under the taxogenomic framework to quantify genetic relatedness beyond traditional barcode sequences [38]. 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 [39]. However, the thresholds developed for prokaryotes are not universally applicable to fungi without prior benchmarking, and their application in yeast remains underdeveloped. Broader application of these and other genomic indices 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 [38, 40].

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 [33]. 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% (**Fig 2A–C** and **S2 Table**). 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 [14, 18] 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 (**Fig 2C** and **S2 Table**), 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 (**Fig 2A–B** and **S2 Table**). The ANI vs. AAI scatterplot (**Fig 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

286 probabilities of exceeding the 70% species threshold between 59.1% and 61.3% (**Fig 2D–F** and **S2**
287 **Table**). For comparison, *C. bacillisporus* and *C. decagattii* yielded a dDDH value of approximately
288 69.3%, just below the conventional cutoff, but with a greater than 75% probability that the true value
289 exceeds the 70% threshold. Nonetheless, these interspecific values remain significantly lower than
290 those observed between *C. neoformans* strains, where dDDH ranges from 79% (probability ≥70% =
291 89.75%) to 99.1% (probability ≥70% = 98.07) (**Fig 2E** and **S2 Table**). This suggests that *C. bacillisporus*
292 and *C. decagattii* lie near the boundary of species-level divergence. In contrast, *C. hyracis* and *C.*
293 *tetragattii* show a more pronounced level of genomic separation. Together, these results support the
294 recognition of *C. hyracis* as a distinct species, building on previous work by Farrer et al. [19] that first
295 identified it as a divergent lineage within the *C. gattii* complex.

296 Although all currently available strains of *C. hyracis* are of the *MAT*α mating type and no sexual
297 cycle has been observed thus far, we tested whether *C. hyracis* could engage in mating with strains
298 from other *C. gattii* (sensu lato) species. Our assays showed that *C. hyracis* is capable of forming sexual
299 structures in pairings with *MAT***a** strains of *C. bacillisporus*, *C. decagattii*, *C. tetragattii,* and *C.*
300 *deuterogattii* (**S2 Fig**). Notably, strain MF13 consistently exhibited stronger mating responses than
301 MF34, indicating possible strain-level differences in mating competence, consistent with prior
302 observations of intraspecific variation and their placement in distinct subclades [19]. While the fertility
303 and ploidy status of resulting progeny were not assessed here, these preliminary results suggest that
304 *C. hyracis* retains the ability to mate with other species in the complex. A more systematic investigation
305 of post-zygotic outcomes will be needed to assess reproductive compatibility and isolation across the
306 *C. gattii* complex and to test species boundaries under the biological species concept.

307

308 **Chromosomal rearrangements and karyotypic diversification in clade A species**

309 Structural variation at the chromosomal level provides a complementary axis of genome
310 divergence and may offer important insights into reproductive isolation and the emergence of species

boundaries. All pathogenic *Cryptococcus* species (clade A) maintain a karyotype of 14 chromosomes (**Fig 1A**) [32], and previous studies based on earlier genome assemblies have identified chromosomal rearrangements between *C. neoformans*, *C. deneoformans*, and members of the *C. gattii* complex [19, 41-43]. However, these analyses were limited to a smaller set of species or relied on lower-resolution assemblies, leaving the extent and pattern of large-scale chromosomal rearrangements across these species not fully characterized. 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 [44] and therefore does not represent the typical karyotype of the species (**S3A–B Figs**). Interspecies differences are summarized in **Fig 3** with one representative strain per species. **S3 Fig** presents additional karyotype comparisons among *C. neoformans* and *C. deneoformans* strains, while **S4 Fig** 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 (**Figs 3A–B**), consistent with intercentromeric recombination (ICR) as a known mechanism of chromosome reshuffling in *Cryptococcus* [43, 45]. 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 (**S4 Fig**). 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) (**Figs 3A** and **S4B**, event a). Two further rearrangements involving anc12–anc13 and anc8–anc10 define the common ancestor of *C. bacillisporus* and *C. decagattii* (**S4B Fig**, 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* (**S4B Fig**, event d). We also detected two strain-specific translocations: one in *C. decagattii* CBS11687 (between anc1 and anc11; **S4B Fig**, event e) and another in *C. tetragattii* CBS11718 (between chrs. anc6 and anc13; **S4B Fig**, 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; **Fig 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

362 explore these aspects, we first conducted pairwise genome comparisons among the six sequenced
363 strains revealing high overall similarity, with ANI values ranging from 99.29% to 99.94% and dDDH
364 values from 94.6% to 97.2% (**Fig 2E** and **S2 Table**), confirming that all isolates belong to the same
365 species despite some geographic separation.

366 Next, we conducted mating assays between strain pairs predicted to be compatible based on
367 sequence differences at both the pheromone/receptor (*STE3*) and homeodomain (*HD*) mating-type loci
368 (**Fig 4A**). These two unlinked *MAT* loci were previously characterized and shown to follow a tetrapolar
369 configuration in *C. porticicola* and other clade D species [30]. Among the tested combinations (**S5 Fig**),
370 only the cross between strains DSM108352 (*a1b3*) and NCYC1536 (*a2b2*) sporadically produced visible
371 hyphae and putative sexual structures embedded in the agar, resembling basidia (**Fig 4B–C**). These
372 structures emerged only after extended incubation (~1 month) and were not consistently observed
373 across replicates. Although their identity could not be definitively confirmed and may represent
374 chlamydospore-like cells, their morphology is more consistent with sexual development observed in
375 related *Kwoniella* species [46-48]. These structures were observed on V8 medium pH 5.0, whereas
376 growth on MS medium led to hyphal formation but without basidial-like clusters (**Fig S5D**). Notably,
377 some hyphal segments appeared to contain unfused clamp connections (**Fig 4C**), potentially indicating
378 a monokaryotic mycelium.

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

NCYC1536—the only cross producing putative sexual structures—showing the greatest separation along PC1, which accounts for the largest proportion of genetic variance (**Fig 4E**).

To assess patterns of historical recombination more directly, we examined linkage disequilibrium (LD). 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 gradually breaking down associations between alleles (**Fig 4F**). Based on the smoothed decay curve, r^2 dropped to 50% of its initial maximum (LD50) at approximately 13.9 kb. Taken together, and with the caveat of limited sample size, these findings indicate that *C. porticicola* undergoes some degree of recombination, 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 (**Figs 4A** and **4G**). 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 (**Fig 4G**). 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 (**Fig 4H**). 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) could contribute to postzygotic isolation through segregation defects, potentially explaining the rarity and incompleteness of sexual development. These findings suggest that *C. porticicola* may comprise

distinct populations (e.g. geographic or host-related) with established genomic and structural divergence.

Chromosome number reduction and centromere dynamics in clade D species

The chromosomal variation identified within *C. porticicola* prompted us to investigate broader trends in karyotype evolution across clade D. All three newly described species in this clade exhibit marked reductions in chromosome number relative to *Cryptococcus* species of clades A and C, which maintain the ancestral 14-chromosome karyotype inferred for the genus [32]. Specifically, *C. porticicola* has 9 chromosomes, while *C. ficicola* and *C. jaspeensis* have only 8 and 7 chromosomes, respectively (**Fig 5A**).

To elucidate the genomic changes underlying these reductions, we compared chromosome-scale assemblies using *C. porticicola* DSM108351 as a reference and mapped centromere positions across all available strains of clade D. In *Cryptococcus*, centromeres are regional rather than point-like, usually spanning tens of kilobases and characterized by a lack of coding sequences, enrichment in transposable elements—particularly long-terminal repeat (LTR) retrotransposons—and high levels of 5-methylcytosine (5mCG) DNA methylation [41, 45, 49-51]. DNA methylation calls were generated from Oxford Nanopore sequencing data obtained either in this work or from previously published datasets [30, 32]. In *C. jaspeensis*, the region syntenic to *CEN7* of *C. porticicola* and *C. ficicola* lacks hallmark centromeric features and likely no longer functions as a centromere (**Figs 5B** and **S6F**). Similarly, both *C. ficicola* and *C. jaspeensis* contain additional presumably inactivated centromeres (*iCENs*) corresponding to *CEN9* and *CEN8* of *C. porticicola*. These *iCENs* retain conserved flanking genes but are depleted of LTR retrotransposons and show only residual 5mCG methylation, consistent with erosion of centromeric identity [52, 53] while preserving local gene order (**Figs 5D**, panel III, and **S6G**).

Notably, the number of *iCENs* in *C. ficicola* (two) and *C. jaspeensis* (three) exceeds the reduction in chromosome number relative to *C. porticicola*, implying that centromere inactivation alone does not

fully explain their current karyotypes. A conserved region on chr. 4 in *C. ficicola* and chr. 5 in *C. jaspeensis* exhibits clear centromeric hallmarks, including TE accumulation and 5mCG enrichment, while the syntenic region in *C. porticicola* displays only spatially restricted 5mCG enrichment and a small number of TEs but does not coincide with any predicted centromere (**Figs 5C** and **5D**, panel II). These findings point to a centromere repositioning event in the ancestor of *C. ficicola* and *C. jaspeensis*, although whether this region represents a newly co-opted centromere or the retention of an ancestral one that became inactive in *C. porticicola* remains uncertain. Direct functional validation will be required to confirm centromeric activity. Importantly, gene order is preserved across these chromosomes, suggesting that centromere repositioning, if it occurred, did so without major local rearrangement. Together, these findings suggest that chromosome number reduction in clade D species, potentially contributing to species diversification, is linked to structural rearrangements followed by centromere inactivation.

Variation in rRNA gene organization across *Cryptococcus* species

In addition to chromosome number changes, variation in ribosomal RNA (rRNA) gene organization represents another axis of genome structural divergence in *Cryptococcus* [32]. To explore this in more detail, we applied a unified pipeline that integrates rRNA gene prediction, clustering detection, and statistical enrichment tests (see Material and Methods) to survey the chromosomal arrangement of 5S, 5.8S, 18S, and 28S rRNA genes across 20 chromosome-level genome assemblies representing all major clades and species (**Figs 6** and **S7**).

Fully clustered rDNA arrays (defined as tandem repeats of the canonical 28S–5.8S–18S–5S gene units) were observed exclusively in pathogenic species from clade A. In contrast, other species exhibited unclustered architectures, with the 5S gene physically separated from the 28S–5.8S–18S unit and exhibiting lineage-specific variation in chromosomal distribution. Among clade C species, the 5S genes remained significantly enriched on a single chromosome (chi-squared test, $P < 0.0001$ for all

species; **Figs 6B** and **S7**, and **Table 3**), suggesting that despite dispersion, the ancestral chromosomal context of the rDNA array may be partly retained. For instance, in *C. floricola*, over 50% of all 5S genes are located on a single chromosome (standardized residual > +11; **S3 Table**), which also harbors the canonical rDNA array. By contrast, in other clade C species (*C. wingfieldii* and *C. maquisensis*), the 28S–18S–5.8S array appears relocated to alternate chromosomal ends, consistent with interchromosomal mobility potentially mediated by subtelomeric recombination (**S7 Fig**).

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

Phenotypic differentiation among *Cryptococcus* clades

To examine whether genomic divergence across the *Cryptococcus* genus is accompanied by phenotypic differentiation, we performed comparative growth and metabolic profiling across 16 of the 17 recognized species. Eighteen strains were tested, including duplicate representatives for *C. porticicola* and *C. amylolentus* (**Fig 7** and **S4 Table**). *C. depauperatus* was excluded due to its 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 (see Methods for details). Of these, 112 yielded variable

488 responses across the 18 strains and were retained for comparative analyses (**S4 Table**). Hierarchical
489 clustering of the binarized profiles (i.e., positive or negative growth) revealed clade-level groupings,
490 indicating that species within the same clade tend to share more similar metabolic profiles (**Fig 7A**).
491 However, the branching order in the dendrogram does not mirror the species phylogeny (**Fig 1**),
492 suggesting that while combination of metabolic traits can discriminate clades, they may reflect
493 ecological adaptation or functional convergence rather than shared evolutionary history. Principal
494 component analysis (PCA) of the same dataset (**Fig 7B**) supported broad clade separation along PC1
495 (23.7%) and PC2 (10.9%). The only exception was *C. cederbergensis* (clade B), which clustered within
496 the 95% confidence ellipse of clade C, suggesting metabolic similarity despite phylogenetic distance.
497 This pattern may, however, change with the discovery and characterization of additional clade B species
498 in the future.

499 To further explore clade-specific metabolic differentiation, we applied a Random Forest classifier
500 to the binarized Biolog dataset. This analysis identified a ranked set of informative features, with
501 glycerol, glucuronamide, L-serine, xylitol, and D-galacturonic acid among the top substrates contributing
502 to clade discrimination (**S8A Fig**). Visualization of their average presence across clades revealed clear
503 clade-specific patterns (**S8B Fig**). For example, glycerol and xylitol were utilized by all clades except
504 clade A; glucuronamide utilization was largely restricted to clade A; and clades B and C showed
505 impaired growth in 4% and 8% NaCl, consistent with heightened osmotic stress sensitivity (see also
506 Taxonomy section). Conversely, D-melibiose utilization was restricted to clade C species. These trends
507 suggest that substrate utilization profiles may reflect underlying phylogenetic structure and offer
508 phenotypic markers for clade discrimination. To ensure reliability, a subset of Biolog results was cross-
509 validated against standard assimilation tests from classical yeast taxonomy and physiological
510 databases (see Methods for details). As an additional validation of these findings, we focused on
511 glycerol and xylitol, the two most informative polyols identified by the Random Forest analysis, and
512 performed spot growth assays on defined minimal media. The results confirmed the Biolog patterns:
513 pathogenic clade A species exhibited little or no growth when glycerol or xylitol were provided as the

514 sole carbon sources, whereas non-pathogenic species (clades B–D) grew robustly under the same
515 conditions (**S9 Fig**).

516 In parallel, we evaluated strains from all recognized *Cryptococcus* species (excluding *C.*
517 *depauperatus*) for two traits broadly associated with pathogenicity: thermotolerance and melanin
518 production. Growth at 37°C was tested on rich medium, and melanin synthesis was assessed on Bird
519 seed agar and L-DOPA (**Fig 7C** and **S8C**). Thermotolerance was sharply restricted to clade A
520 pathogens: all clade A strains grew robustly at 35°C and 37°C, while none of the clade B, C, or D strains
521 did. Melanin production showed a similar distribution: all clade A strains produced some level of melanin
522 on both media, whereas all nonpathogenic species lacked detectable production. This pattern was not
523 explained by variation in laccase gene copy number. Laccases are oxidoreductase enzymes that
524 catalyze melanin biosynthesis by oxidizing diphenolic compounds such as L-DOPA [54, 55]. For
525 instance, nonpathogenic clade C species such as *C. amylolentus* encode up to four predicted laccase
526 genes but failed to produce melanin under the tested conditions indicating that gene copy number alone
527 does not account for trait expression. Together, these results align with prior studies linking
528 thermotolerance and melanin biosynthesis to virulence in *Cryptococcus* and suggest that these traits
529 contribute to both ecological adaptation and the evolution of pathogenic potential.

530

531 Environmental occurrence and limits of ITS-based detection of newly described species

532 Having established the genomic and phenotypic distinctness of the newly delineated
533 *Cryptococcus* species, we next investigated their presence in environmental sequencing datasets to
534 place them in a broader ecological context. Previous studies that uncovered some of these taxa
535 reported their extremely rare occurrence in the environment [31, 34]. As ITS-based metabarcoding is
536 the most widely used approach in fungal biodiversity surveys [56], it holds promise for detecting rare
537 fungi, including *Cryptococcus*. The broad availability of ITS metabarcoding data also offers an

opportunity to assess how effectively current barcoding methods detect and identify these species in nature.

To this end, we performed exact-match searches of ITS1 and ITS2 sequences against the GlobalFungi database, which compiles over 80,000 fungal metabarcoding samples from diverse ecosystems [57]. Matches were considered only when query sequences were identical in both length and nucleotide composition. *Cryptococcus porticicola* was detected in nine samples from Europe, New Caledonia, and the United States, primarily from soil and decaying wood. Notably, four samples from northern Europe (Estonia, Finland, and Russia) were from coniferous forests dominated by Norway spruce (*Picea abies*), the typical habitat of *Ips typographus* (European spruce bark beetle), the main source of *C. porticicola* (**S1 and S5 Tables**). It was particularly abundant in dead bark samples from the Karelia region [58]. *Cryptococcus maquisensis* was detected in two samples from young *Populus deltoides* xylem collected in Knoxville, Tennessee (USA) [59]. No exact matches were found for *C. cederbergensis*, possibly suggesting a more restricted distribution.

While *C. cederbergensis*, *C. maquisensis*, and *C. porticicola* possess distinct ITS1 and ITS2 sequences, other species exhibit more limited marker resolution. For example, *C. ficicola* and *C. jaspeensis* differ in ITS1 but share an identical ITS2 sequence. ITS1 of *C. ficicola* matched a single air sample from Italy [60], while ITS1 of *C. jaspeensis* returned no matches. The shared ITS2 sequence was recovered from 20 soil samples in Italy, Switzerland [61, 62], and New Caledonia [63]. Similarly, *C. hyracis* shares ITS1 identity with *C. tetragattii*, *C. deuterogattii*, and *C. bacillisporus*, and ITS2 identity with *C. deneoformans*, precluding species-specific identification. It should be noted that the GlobalFungi database is limited in insect-derived samples and lacks bark beetles entirely, providing little insight into insect-associated lineages. A full summary of GlobalFungi matches and ITS1/ITS2 resolution is provided in **S5 Table**, and an interactive map of matching samples is available as **S2 File**.

Taxonomic description of novel *Cryptococcus* species

Description of *Cryptococcus hyracis* M.C. Fisher, M. Vanhove, R.A. Farrer, M. David-Palma, M.A. Coelho, and J. Heitman sp. nov. (Index Fungorum: IF 904433)

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

After 1 week at 25°C on PD and GPY streak is white to cream, mucoid with a smooth glistening 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, as currently available strains are of the same mating type (*MAT*α).

Glucose is not fermented. Positive growth on D-Glucose, D-Galactose, L-Sorbose, D-Xylose, L-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.

Molecular characteristics (holotype): nucleotide sequences of ITS and LSU (D1/D2 domains) are available under GenBank accession number PV828371.

Description of *Cryptococcus cederbergensis* S. Marincowitz, N.Q. Pham, M. David-Palma, M.A. Coelho, J. Heitman, and M.J. Wingfield, sp. nov (Index Fungorum: IF 904434)

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.

Glucose is not fermented. Positive growth on D-Glucose, D-Galactose, 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, Melibiose, Lactose, Raffinose, Melezitose, Inulin, Starch, Glycerol, Erythritol, Ribitol, Xylitol, Arabinitol, D-Glucitol, D-Mannitol, Galactitol, myo-Inositol, 2-Keto-D-Gluconate, D-Gluconate, D-Glucuronate, Succinate, Citrate, Methanol, Ethanol, D-glucarate, Galactonic Acid Lactone, Palatinose, L-Malic acid, Galactaric acid, Gentobiose, and Tween 80. No growth on L-Sorbose, D-Galacturonate, DL-Lactate, Propane 1,2 diol, Quinic acid, L-Tartaric acid, D-Tartaric acid, and Tween 40. No growth in the presence of 10% sodium chloride, but not on 50% D-Glucose. Urea hydrolysis and Diazonium Blue B reaction are positive. Starch-like compounds are produced. Maximum growth temperature: 30°C (weak, slow).

Type: holotype, dried specimen PRU(M) 4637, isolated from a bark beetle (*Lanurgus* sp.) infesting twigs of the endangered conifer *Widdringtonia cedarbergensis*, collected on 17 October 2022 in the Cederberg Mountains of South Africa, preserved in the H.G.W.L. Schweickerdt herbarium (PRU).

Ex-holotype CMW 60451, preserved in a metabolically inactive state in the culture collection of the Forestry and Agricultural Biotechnology Institute (FABI) at the University of Pretoria, South Africa, as CMW-IA 7236, and in the German Collection of Microorganisms and Cell Cultures (DSMZ), Braunschweig, Germany, as DSM 117695.

Molecular characteristics (holotype): nucleotide sequences of ITS and LSU (D1/D2 domains) are available under GenBank accession number PV828378.

Description of *Cryptococcus maquisensis* M.A. Coelho, M. David-Palma, O. Röhl, J.P. Sampaio, J. Heitman, and A.M. Yurkov, sp. nov. (Index Fungorum: IF 904435)

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

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. Short pseudohyphae observed in old cultures. No true hyphae were 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, Cellobiose, Salicin, Arbutin spliting, Melibiose, Raffinose, Melezitose, Starch, Glycerol, Erythritol, Ribitol, Xylitol, Arabinitol, D-Glucitol, D-Mannitol, Galactitol, myo-Inositol, 2-Keto-D-Gluconate, D-Gluconate, D-Glucuronate, Succinate, Citrate, Ethanol, D-glucarate, Galactonic Acid Lactone, Palatinose, L-Malic acid, L-Tartaric acid, Galactaric acid, and Gentobiose. No growth on L-Sorbose, L-Arabinose, Me a-D-Glucoside, Lactose, Inulin, D-Galacturonate, DL-Lactate, Methanol, Propane 1,2 diol, Quinic acid, D-Tartaric acid, Tween 40, and Tween 80. Growth in the presence of 10% sodium chloride (weak), and on 50% and

637 60% D-Glucose. Urea hydrolysis and Diazonium Blue B reaction are positive. Starch-like compounds
638 are produced. Maximum growth temperature: 30°C.

639 Type: strain OR849, isolated from soil, collected in September 2013 in Serra da Arrábida,
640 Setúbal, Portugal; preserved in a metabolically inactive state in the Portuguese Yeast Culture Collection
641 (PYCC), Caparica, Portugal (PYCC 10407, holotype), and in the German Collection of Microorganisms
642 and Cell Cultures (DSMZ), Braunschweig, Germany (DSM 104548, isotype).

643 Molecular characteristics (holotype): nucleotide sequences of ITS and LSU (D1/D2 domains)
644 are available under GenBank accession number PV828377.

645

646 **Description of *Cryptococcus ficicola* B. Turchetti, M. David-Palma, M.A. Coelho, A.M. Yurkov,**
647 **and J. Heitman, sp. nov. (Index Fungorum: IF 904436)**

648 Etymology: from Latin *ficus* (fig), referring to the fig fruit, the source of isolation.

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

654 Glucose is not fermented. Positive growth on D-Glucose, D-Galactose, D-Ribose, D-Xylose, D-
655 Arabinose, L-Rhamnose, Sucrose, Maltose, a,a-Trehalose, Cellobiose, Salicin, Arbutin spliting,
656 Raffinose, Melezitose, Starch, Glycerol, Erythritol, Ribitol, Xylitol, Arabinitol, D-Glucitol, D-Mannitol,
657 Galactitol, myo-Inositol, 2-Keto-D-Gluconate, D-Gluconate, D-Glucuronate, D-Galacturonate,
658 Succinate, Ethanol, Palatinose, L-Malic acid, Gentobiose, and Tween 80. No growth on L-Sorbose, D-
659 Glucosamine, L-Arabinose, Me a-D-Glucoside, Melibiose, Lactose, Inulin, DL-Lactate, Citrate,
660 Methanol, Propane 1,2 diol, Quinic acid, D-glucarate, Galactonic Acid Lactone, L-Tartaric acid, D-
661 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.

Molecular characteristics (holotype): nucleotide sequences of ITS and LSU (D1/D2 domains) are available under GenBank accession number PV828380.

Description of *Cryptococcus jaspeensis* M.A. Coelho, M. David-Palma, O. Röhl, J.P. Sampaio, J. Heitman, and A.M. Yurkov, sp. nov. (Index Fungorum: IF 904437)

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: strain OR918, isolated from soil collected in September 2013 in Serra da Arrábida, Setúbal, Portugal, preserved in a metabolically inactive state in the Portuguese Yeast Culture Collection (PYCC), Caparica, Portugal (PYCC 10408, Holotype), and in the German Collection of Microorganisms and Cell Cultures (DSMZ), Braunschweig, Germany (DSM 104549, isotype).

Molecular characteristics (holotype): nucleotide sequences of ITS and LSU (D1/D2 domains) are available under GenBank accession number PV828379.

Description of *Cryptococcus porticicola* M.A. Coelho, M. David-Palma, A.V. Kachalkin, T.A. Kuznetsova, M. Kolarik, A.M. Yurkov, and J. Heitman, sp. nov. (Index Fungorum: IF 904438)

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

711 approximately 8–10 µm in diameter. Their morphology was consistent with basidia, but their identity
712 could not be definitively confirmed. No sexual structures were observed on MS medium.

713 Glucose is not fermented. Positive growth on D-Glucose, D-Galactose, L-Sorbose, D-
714 Glucosamine, D-Ribose, D-Xylose, L-Arabinose, D-Arabinose, L-Rhamnose, Sucrose, Maltose, a,a-
715 Trehalose, Me a-D-Glucoside, Cellobiose, Salicin, Arbutin spliting, Raffinose, Melezitose, Starch,
716 Glycerol, Erythritol, Ribitol, Xylitol, Arabinitol, D-Glucitol, D-Mannitol, Galactitol, myo-Inositol, 2-Keto-D-
717 Gluconate, D-Gluconate, D-Glucuronate, D-Galacturonate, Succinate, Ethanol, D-glucarate, Galactonic
718 Acid Lactone, Palatinose, L-Malic acid, Galactaric acid, Gentobiose, Tween 40, and Tween 80. No
719 growth on Melibiose, Lactose, Inulin, DL-Lactate, Citrate, Methanol, Propane 1,2 diol, Quinic acid, L-
720 Tartaric acid, and D-Tartaric acid. Utilization of nitrogen sources: positive growth on potassium nitrate,
721 sodium nitrite, ethylamine, and lysine. No growth in the presence of 10% sodium chloride (weak), and
722 on 50% and 60%D-Glucose. Urea hydrolysis and Diazonium Blue B reaction are positive. Starch-like
723 compounds are produced. Maximum growth temperature: 30°C.

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

730 Paratypes: KBP Y-6300 (= DSM 108352 = VKM Y-3027) from the gallery of great spruce bark
731 beetle, *Dendroctonus micans* (in the vicinity of Dmitrov town, Moscow region, Russia), KBP Y-6370 (=
732 DSM 111204 = VKM Y-3517) and KBP Y-6556 (= DSM 111205 = VKM Y-3589) from spruce bark beetle
733 gut, *Ips typographus* (in the vicinities of Nakipelovo and Gzhel villages, Moscow region, Russia), CCUG
734 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).

Molecular characteristics (holotype): nucleotide sequences of ITS and LSU (D1/D2 domains) are available under GenBank accession number PV828382.

Discussion

This study expands the known diversity of the genus *Cryptococcus* by formally describing six new species distributed across four distinct clades. 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 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 rapid, reproducible, and evolutionarily grounded basis for species delineation [36, 38]. 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 [64, 65]. Still, metrics such as ANI and AAI must be benchmarked against external data reflecting fungal biology, including ecology and reproduction—traits traditionally employed for species circumscription under the phenotypic and biological species concepts [56, 66].

758 Species boundaries in this study were defined primarily by genome-scale similarity patterns.
759 Lineages were recognized as distinct species when they formed well-supported monophyletic groups
760 and showed genome-wide similarity lower than the minimum observed among conspecifics of the well-
761 circumscribed reference species *C. neoformans*, that is, below the maximum genomic divergence still
762 consistent with being the same species. This within-species range provided an empirical upper bound
763 for intraspecific similarity, which was contrasted with the interspecific similarity observed between the
764 closely related species *C. bacillisporus* and *C. decagattii*, representing the highest ANI values observed
765 between distinct species. Lineages with similarity values falling below this interspecific range were
766 therefore interpreted as distinct species, whereas those with intermediate values may represent
767 borderline or incipient cases. Because absolute genomic thresholds for fungal species remain
768 undefined, these criteria were applied empirically and interpreted within a comparative rather than
769 prescriptive framework. Additional evidence, such as differences in chromosomal structure, sexual
770 compatibility, or ecology, provided supporting context where available. When different datasets were
771 not fully congruent, genome-wide similarity and phylogenetic distinctness were prioritized as the most
772 stable indicators of long-term evolutionary independence, whereas reproductive, structural, and
773 phenotypic traits (with more limited coverage) were considered complementary and often asynchronous
774 indicators of speciation. This integrative framework ensured consistency even when experimental
775 assessments of mating compatibility or progeny viability were limited or variable among strains.
776 Although applied here to *Cryptococcus*, this approach can be extended to other basidiomycete genera,
777 particularly those that can be benchmarked against existing mating-compatibility tests and/or
778 documented host-range data.

779 In our phylogenomic analyses, residual conflict remained at both deep and shallow nodes,
780 although strong overall concordance was observed across inference methods. For instance, while the
781 placement of clades B and C relative to clade A was topologically consistent, reduced gene and site
782 concordance and lower quartet support at this deep node point to lingering uncertainty, likely reflecting
783 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 [19, 67]. 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 dDDH provide a complementary and quantitative framework for evaluating species boundaries, as demonstrated in prokaryotes [37]. Although fungi lack a universally accepted genomic threshold for species delimitation, a 95% ANI cutoff has been proposed based on limited comparisons across a few selected ascomycetous and basidiomycetous yeasts. However, our comparative analyses support emerging evidence that this threshold may not be suitable and underestimate species-level divergence in some basidiomycete lineages [38, 68]. For example, the lowest ANI and dDDH values observed among *C. neoformans* strains (97.7% and 79%, respectively) provide a conservative lower bound for intraspecific genomic similarity, while the highest ANI and dDDH values between *C. bacillisporus* and *C. decagattii* (96.5% ANI; 69.4% dDDH), two recognized species, establish a practical upper bound for interspecific genomic similarity.

The newly described species *C. hyracis* sp. nov. illustrates the utility of this framework. Phylogenomic analyses resolved *C. hyracis* as sister to *C. tetragattii*, with ANI values ranging from 95.32% to 95.51%, and dDDH values from 62.1% to 62.8%. These values fall below both the intraspecific thresholds observed in *C. neoformans* and even the interspecific values seen between *C. bacillisporus* and *C. decagattii*, providing quantitative support for species-level distinction. Although all currently available *C. hyracis* strains are *MAT*α, precluding intraspecific mating assays, interspecific crosses with *MAT***a** strains from other species produced sexual structures, suggesting the lineage

retains sexual competence. Further studies evaluating progeny viability and reproductive compatibility with *C. tetragattii* are needed but are constrained by the current lack of *C. hyracis MAT***a** isolates for control crosses. More broadly, systematic reproductive assessments across the *C. gattii* complex would help refine the interpretive power of genome-based divergence thresholds. When standardized and applied in a clade-specific context, these genomic similarity metrics should provide a scalable and reproducible framework for species delimitation, particularly in lineages where experimental crosses are infeasible. These comparative benchmarks will be refined as broader sampling, including opposite mating types, and additional genome sequencing improve clade-specific calibration. The presence of hybrid strains remains a caveat, underscoring the need to integrate genomic evidence with other lines of data. Together, these findings extend earlier support for the distinctiveness of *C. hyracis* [19] and reinforce the broader utility of empirical, lineage-specific benchmarks for taxonomic resolution in *Cryptococcus* and other fungal taxa.

Chromosomal architecture and genome-wide sequence divergence do not always evolve in concert, and both can independently shape species boundaries. For instance, *C. hyracis* and *C. tetragattii* are resolved as sister species and differ by ~4.5% at the genome level, yet their karyotypes are nearly collinear apart from a single large inversion on chromosome 2, which would only impact fertility in the rare case of crossover within the inverted segment. A similar decoupling of sequence and structural evolution is observed between *C. neoformans* 125.91 and *C. deneoformans* NIH433, which diverge by ~12% but differ structurally by only two large-scale inversions (on chromosomes 1 and 9). This pattern echoes observations in other fungal systems. In *Saccharomyces*, for instance, S. *cerevisiae* and *S. paradoxus* retain collinear genomes but produce sterile hybrids due to anti-recombination effects mediated by mismatch repair [69, 70]. In *Cryptococcus*, deletion of *MSH2* (a key component of the mismatch repair system) enhances hybrid viability between *C. neoformans* and *C. deneoformans*, yet fails to restore normal levels of meiotic recombination, implying the presence of additional post-zygotic barriers [71]. Adding further nuance, some *C. neoformans* strains (e.g., H99, Bt65, Bt81) harbor private rearrangements [43, 72], demonstrating that chromosomal changes can arise

independently within species and may not reflect broader lineage evolution. In contrast, *C. bacillisporus* and *C. decagattii*, the most closely related species pair in our dataset (~3.5% divergence), exhibit up to two fixed reciprocal translocations between their sequenced strains, despite high genome-wide similarity. Whether these rearrangements initiated speciation or arose secondarily remains unresolved. However, experimental work in *C. neoformans* shows that targeted centromere breakage can generate novel karyotypes that reduce meiotic viability in the absence of sequence divergence, supporting a causal role for chromosomal changes in reproductive isolation [43]. Although often underappreciated, these findings draw further attention to the role of chromosomal rearrangements in fungal speciation and illustrate the importance of broad within-species sampling to fully capture karyotypic dynamics and evolutionary trajectories.

In addition to the structural rearrangements observed among closely related pathogenic species, our analyses revealed more extensive karyotypic shifts in clade D, including multiple reductions in chromosome number. The three newly described clade D species (*C. porticicola*, *C. ficicola*, and *C. jaspeensis*) harbor fewer chromosomes (9, 8, and 7, respectively) than the ancestral 14-chromosome karyotype inferred for *Cryptococcus* and its sister genus *Kwoniella* [32]. Comparative synteny analyses of centromeric regions suggest that these chromosome number reductions were primarily driven by structural rearrangements followed by centromere inactivation. Several centromere-syntenic regions retain conserved flanking gene order but lack TE enrichment and display only residual 5mCG, consistent with inactivated centromeres (*iCENs*). This aligns with broader models in which centromere identity erodes through the progressive loss or excision of repeat-rich DNA, as reported in *C. depauperatus* [32] and other fungi such as *Verticillium*, *Malassezia* and *Candida* [52, 53, 73, 74], supporting centromere inactivation as a recurring process in karyotype evolution. We also identified an intercentromeric recombination (ICR) event in *C. porticicola* strain DSM111204, associated with a strain-specific chromosomal rearrangement. ICR has been reported in both pathogenic and nonpathogenic *Cryptococcus* species [32, 43, 45], and its detection here extends the known phylogenetic distribution of this mechanism.

While centromere inactivation and ICR can explain many of the observed rearrangements, the patterns in *C. ficicola* and *C. jaspeensis* point to an additional process. In these species, the total number of inferred *iCENs* surpasses the number of chromosomes lost when compared to *C. porticicola*, suggesting that centromere inactivation alone does not fully account for their current karyotypes. This raises the possibility that centromere repositioning or neocentromere formation may have accompanied or followed inactivation events to maintain chromosome stability. A candidate centromere shared by both species, located on chromosomes 4 and 5, respectively, exhibits hallmark centromeric features, including enrichment for TEs and 5mCG DNA methylation. The syntenic region in *C. porticicola* shows only partial enrichment for these features, suggesting that it may represent either a newly co-opted centromere in the ancestor of *C. ficicola* and *C. jaspeensis*, or an ancestral centromere that was subsequently lost or inactivated in *C. porticicola*. If this region indeed functions as a centromere, its architecture contrasts with neocentromeres described in *C. deuterogattii*, which were experimentally induced by targeted deletion of native centromeres and typically form at gene-rich, repeat-poor, and unmethylated loci and are often epigenetically unstable [50, 51]. This distinction may reflect either species-specific differences or contrasts between newly formed neocentromeres and those that have stabilized over extended evolutionary periods. Future work will aim to functionally test centromere activity once these species become genetically tractable, which will enable targeted validation (e.g., via CENP-A ChIP-seq) and experimental assessment of centromere dynamics. Taken together, our findings suggest that centromere inactivation, intercentromeric recombination, and centromere repositioning are recurrent processes contributing to genome restructuring and karyotype evolution in *Cryptococcus*, with potential implications for lineage diversification and speciation.

Ribosomal RNA (rRNA) genes have long served as core phylogenetic and taxonomic markers in fungal systematics [56, 66]. 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 [75]

and likely maintained by constraints related to nucleolar organization and coordinated rRNA dosage control [76]. 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. ficicola* 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. Interestingly, in *Schizosaccharomyces octosporus* and *S. osmophilus*, rapidly evolving *wtf* poison–antidote genes are frequently positioned adjacent to the highly dynamic 5S rDNA repeats; this genomic proximity suggests that the repeat-rich 5S neighborhood may facilitate duplication and diversification of these selfish elements, accelerating local genome evolution [77]. While no equivalent meiotic drive system has been characterized in *Cryptococcus*, this pattern raises the possibility that analogous repeat-mediated processes could underlie the rapid 5S architecture changes observed in this genus. 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 [78]. 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 [78, 79]. The subtelomeric bias observed in some *Cryptococcus* species is also consistent with known hotspots of recombination and TE activity [72]. 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 differences raise important questions about the functional consequences of rDNA organization, including possible effects on nucleolar structure, rRNA transcriptional regulation, or genome stability. While these impacts remain unclear, the extent of observed divergence suggests that rRNA gene architecture represents an underappreciated axis of genome evolution in *Cryptococcus* and positions

914 this genus as a promising model for investigating rRNA gene dynamics and genome
915 compartmentalization in fungi.

916 Beyond structural genome evolution, we also explored the recombination potential of *C.*
917 *porticicola*, the only newly described nonpathogenic species for which multiple isolates of opposite
918 mating types were available. All six strains were isolated from bark beetle galleries or larval guts,
919 indicating a strong ecological association with conifer-infesting beetles. In mating assays, only one
920 cross (DSM108352 × NCYC1536) produced hyphae and embedded basidia-like structures resembling
921 those of *Kwoniella* rather than the typical aerial basidia of *Cryptococcus*, suggesting a potentially
922 divergent mode of sexual development. Although these structures emerged only sporadically and their
923 identity requires further ultrastructural or genetic confirmation, they appeared morphologically distinct
924 from the chlamydospore-like cells reported in *C. neoformans* under stress or nutrient-limiting conditions,
925 which are typically more darkly refractile and seem to possess thicker cell walls [80]. Notably, the two
926 mating strains differ by two chromosomal translocations, which are likely to reduce progeny viability.
927 Despite limited morphological evidence, genomic analyses based on a small number of isolates support
928 ongoing sexual reproduction in *C. porticicola*. Genome-wide SNP patterns revealed reticulate
929 relationships and measurable LD decay, with LD50 approximately 13.9 kb. This value is broadly
930 comparable to those reported for recombining *C. neoformans* subpopulations (5–7.5 kb when analyzed
931 separately; [33]) and to those observed in *Saccharomyces paradoxus* (LD50 ~9 kb; [81]). Although the
932 number of isolates remains limited, these comparable magnitudes indicate that recombination likely
933 occurs in natural populations of *C. porticicola*, albeit potentially at lower effective frequency or within
934 localized population contexts. While additional sampling will be needed to clarify the extent of sexual
935 reproduction, the combined genomic and morphological evidence suggests that meiotic processes likely
936 contribute to genetic cohesion, evolutionary flexibility, and ecological persistence in this species. Future
937 work will focus on optimizing mating conditions to enable successful crosses between strains with more
938 collinear karyotypes, as all other combinations tested to date have not yielded evidence of mating. In
939 parallel, recovering additional strains from bark beetle-associated substrates may help identify more

940 sexually compatible pairs. If these approaches prove unsuccessful, or if sporulation remains rare or
941 progeny are technically difficult to isolate, introducing selectable markers into compatible strains could
942 provide an alternative means to detect genetic exchange following co-culture, as demonstrated in other
943 non-model *Cryptococcus* lineages [25].

944 Complementing our genomic analyses, we evaluated phenotypic variation across the
945 *Cryptococcus* genus to determine whether clade-level divergence is mirrored by metabolic traits.
946 Comparative growth profiling revealed clear clade-level differentiation: hierarchical clustering and
947 principal component analysis (PCA) based on Biolog substrate utilization patterns, benchmarked
948 against standard assimilation tests and culture-collection databases, showed that species within the
949 same clade tend to group together phenotypically, broadly reflecting phylogenetic relationships. Notably,
950 *C. cederbergensis* (clade B) clustered within the clade C phenotypic space, suggesting functional
951 convergence potentially linked to similar environmental conditions. Despite the complexity of the
952 experimental design and validation framework, the consistent clade-level growth responses and
953 clustering observed support its utility for comparative phenotyping and highlight candidate traits for
954 future genomic dissection. Under the assay conditions used here, glycerol and xylitol, two informative
955 substrates identified by the Random Forest analysis, supported robust growth of saprobic species from
956 clades B, C, and D, but only limited or no growth of pathogenic clade A species, when provided as sole
957 carbon sources.

958 Xylitol originates from xylose, a major constituent of plant hemicellulose, and occurs at low levels
959 in plant saps, fruits, and decaying wood, supporting its biogenic availability in plant-rich environments.
960 In wood-decay and insect-associated microhabitats, xylose/xylitol turnover is common: multiple yeasts
961 isolated from rotting wood actively ferment xylose and produce xylitol [82], and similar xylose-fermenting
962 yeasts have been documented in the guts of wood-feeding beetles [83, 84]. Soil yeasts likewise utilize
963 intermediates of plant degradation [85], which may explain the isolation of *C. jaspeensis* and *C.*
964 *maquisensis* from soils in Portugal [34]. Similar to xylitol, glycerol and related polyols occur in bark-
965 beetle–associated habitats, where they seem to accumulate in insect tissues as cryoprotectants and

966 metabolic intermediates [86]. The prevalence of xylitol- and glycerol-positive species among clades B,
967 C, and D, many of which are bark- or insect-associated, may therefore reflect adaptive responses to
968 polyol-enriched microhabitats. While preliminary, these observations provide testable hypotheses for
969 future work connecting substrate use and occupied niche and motivate targeted assays under
970 ecologically relevant conditions. Similar to findings in *Saccharomycotina* yeasts, where ecological
971 transitions often coincide with extensive remodeling of metabolic gene repertoires [87, 88], our
972 phenotype-first approach in *Cryptococcus* also establishes a foundational framework to connect
973 metabolic traits with ecological strategies and historical adaptation. Future comparative genomic
974 analyses will clarify whether these phenotypic shifts reflect underlying changes in gene content or
975 regulation.

976 Our Biolog screens and validation assays were performed at 25°C for standardization, which
977 may not reflect the optimal conditions for substrate utilization across species, nor account for potential
978 metabolic activity under natural or host-associated environments. For instance, although growth in
979 glycerol was recorded as negative in both our experiments and taxonomic literature, studies have
980 reported glycerol utilization by *C. neoformans* strains at higher temperatures [89-91]. Thus, the apparent
981 clade-level differences in glycerol/xylitol utilization should be interpreted as condition-dependent, and
982 they raise the hypothesis that substrate use in pathogenic clade A species may shift at elevated
983 temperatures relevant to host environments. Systematic testing across temperatures (e.g., 25–37°C),
984 and under host-mimicking conditions, will be necessary to resolve these interactions.

985 Strikingly, we confirmed that two traits historically associated with virulence in *Cryptococcus*—
986 thermotolerance and melanin production—are restricted to clade A pathogens in our dataset, under the
987 tested conditions. Yet, in the case of melanin production, nonpathogenic species retain a full
988 complement of known biosynthetic genes [32], including multiple laccase homologs; several even
989 harbor greater copy numbers than their pathogenic counterparts. In *C. neoformans*, melanin synthesis
990 is largely driven by the cell wall–associated laccase *LAC1*, which plays a central role in oxidative stress
991 resistance and survival within macrophages [92, 93], while its paralog *LAC2* can partially compensate

992 under certain conditions [94]. The retention and potential redundancy of these laccase genes in
993 nonpathogenic lineages therefore likely reflect ancestral ecological functions (such as oxidation of
994 environmental phenolics and plant-derived compounds, or detoxification of reactive metabolites), that
995 were later co-opted to support virulence in pathogenic species. This interpretation is consistent with
996 models of exaptation, in which virulence-associated traits arise from preexisting adaptations to diverse
997 ecological niches [95]. Beyond laccase, classical physiological work in *C. neoformans* has shown that
998 melanization requires active, high-affinity uptake of exogenous catechol and catecholamine precursors
999 (e.g., L-DOPA, dopamine, norepinephrine) [96], although the specific transporter(s) responsible remain
1000 unidentified. With respect to thermotolerance, its strict restriction to clade A pathogens reinforces its
1001 status as a critical prerequisite for infection and survival in mammalian hosts. Stressors encountered
1002 during infection, such as elevated temperature, have been shown to trigger transposable element
1003 mobilization in both *C. deneoformans* and *C. neoformans*, resulting in hypermutator-like phenotypes
1004 and accelerated antifungal resistance [97-99]. Although none of the nonpathogenic species analyzed
1005 here can grow at 37°C, several retain high TE burdens [32], raising the possibility that stress-induced
1006 genomic plasticity may also occur in environmental contexts and could, under the right conditions, prime
1007 lineages for future adaptation. Investigating whether similar mutational dynamics operate in these
1008 lineages could reveal a latent evolutionary capacity for pathogenicity. Further work integrating pan-
1009 genomic and pan-phenomic analyses, ecological data, and broader taxon sampling will be key to
1010 resolving how such traits shape adaptation and diversification across the *Cryptococcus* genus.

1011 The ecological breadth of *Cryptococcus* is now more apparent with the recovery of newly
1012 described species from bark beetle galleries, fruit, and soil across diverse geographic regions. Although
1013 distinct from the arboreal hollows, avian guano, and forest debris typically associated with *C.*
1014 *neoformans* and *C. gattii* senso lato, these substrates also represent organic-rich, microbially dynamic
1015 environments that may support long-term persistence and saprobic growth. These findings demonstrate
1016 the importance of environmental sampling in revealing hidden diversity and advancing our
1017 understanding of the ecological and evolutionary processes shaping the genus. They also suggest that

complex environmental niches may serve as incubators for traits involved in host adaptation and pathogenicity. In this context, the formal recognition of phylogenetically distinct lineages is not taxonomic excess but a necessary step toward cataloging fungal diversity, linking genotypes to phenotypes. While our integrated approach convincingly resolved species boundaries, it also revealed a key challenge for future surveys: ribosomal markers such as ITS and LSU lack the resolution to distinguish closely related *Cryptococcus* species, limiting species-level identification in short-read metabarcoding datasets. This highlights the need for continued development of genome-informed frameworks to enable consistent identification of novel species and strains, which are essential to support long-term study, surveillance, and diagnostic development, especially as some species may emerge as clinically relevant under shifting ecological, climatic, or antifungal selective pressures.

1029 # Materials and methods

1030 ## Strains, mating assays, and microscopy

1031 Strains analyzed in this study were routinely cultured on YPD medium (10 g/L yeast extract, 20
1032 g/L Bacto Peptone, 20 g/L dextrose, and 20 g/L agar), unless specified otherwise. *Cryptococcus*
1033 *neoformans*, *C. deneoformans*, and *C. gattii* sensu lato strains were incubated at 30°C, while other
1034 *Cryptococcus* species were maintained at room temperature (20–25°C). A complete list of strains and
1035 associated metadata is provided in **S1 Table**.

1036 Mating assays were performed by mixing equal amounts of cells from each strain on V8 agar
1037 media (10 g/L yeast extract, 20 g/L Bacto Peptone, 20 g/L dextrose, 20 g/L agar; pH 5.0) or Murashige
1038 and Skoog (MS) agar media (4.328 g/L MS mix, 16 g/L agar, 1X vitamin mix, pH to 5.8) and incubated
1039 in the dark at room temperature (conditions known to induce mating in *C. neoformans*) for up to one
1040 month [100]. In the case of *C. porticicola*, incubation was extended due to delayed and sporadic hyphal
1041 development, which was first observed around the one-month mark. Plates were regularly monitored
1042 microscopically for the development of mating structures. Solo cultures of each strain were also
1043 evaluated on V8 pH 5 and/or MS without a compatible partner. Assays were repeated three independent
1044 times to ensure consistent phenotypic observations.

1045 To assess hyphal growth, basidia, and spore formation, the edges of yeast colonies and mating
1046 patches were examined microscopically and photographed. Specifically for *C. porticicola*, putative
1047 sexual structures embedded within the mating media at the patch margins were imaged by excising
1048 small agar blocks and gently compressing them between a microscope slide and coverslip. Light
1049 microscopy was performed using a Zeiss Axio Scope.A1 microscope equipped with an Axiocam Color
1050 camera and ZEN Lite V3.4 software. Samples of sporulating colonies were prepared for scanning
1051 electron microscopy (SEM) as previously described [101] and imaged using the Apreo S SEM with an
1052 EDS detector (ThermoFisher, USA) at the Shared Materials Instrumentation Facility at Duke University.

1053

Genomic DNA extraction

High-molecular-weight (HMW) genomic DNA was isolated using a cetyltrimethylammonium bromide (CTAB)-based extraction method as previously described [43], 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 [32]. For Illumina short-read whole-genome sequencing, genomic DNA was extracted using a phenol:chloroform-based procedure [32]. 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 **S6 Table**.

Genome assembly, gene prediction, and functional annotation

1078 Complete nuclear genomes of *C. porticicola* strains DSM111204 and DSM108352 were
1079 assembled using Canu [102] with Nanopore long-read data and polished with both Nanopore and
1080 Illumina reads, following a previously established workflow [30]. This process included iterative error
1081 correction, removal of mitochondrial and rDNA-only contigs, assessment of assembly completeness
1082 and telomeric regions, and manual inspection of read coverage using IGV. The draft genome assembly
1083 of *C. porticicola* strain DSM111205 was generated using SPAdes v3.15.3 [103] with default parameters.
1084 Gene models were predicted on repeat-masked assemblies using Funannotate v1.8.9
1085 (https://github.com/nextgenusfs/funannotate), as previously described [30]. Manual inspection and
1086 curation were limited to the mating-type loci, including the annotation of short mating pheromone
1087 precursor genes. Functional annotation was performed using the Funannotate "*annotate*" module and
1088 incorporated PFAM and InterPro domains, Gene Ontology (GO) terms, COGs, CAZymes, fungal
1089 transcription factors, MEROPS proteases, secreted proteins, secondary metabolites, and BUSCO
1090 groups, as previously detailed [30]. All assemblies and corresponding raw sequencing data have been
1091 deposited in DDBJ/EMBL/GenBank under the BioProject numbers listed in **S6 Table**. The assemblies
1092 and annotations are also available at Figshare (https://doi.org/10.6084/m9.figshare.30375169).

1093

1094 Ortholog identification, phylogenetic analyses, and topological support

1095 To construct the phylogenomic data matrix, single-copy orthologs (SC-OGs) were identified
1096 across all *Cryptococcus* species and strains, along with the outgroup *Kwoniella shandongensis*
1097 (GCA_008629635.2), using OrthoFinder v3.0.1b1, as previously outlined [30]. A total of 3,510 SC-OGs
1098 shared among all species were identified. To ensure quality and comparability, only proteins from the
1099 *C. neoformans* H99 reference strain that were ≥100 amino acids in length and whose orthologs
1100 exceeded 50% of the H99 protein length were retained. These filtering steps yielded 3,321 SC-OGs,
1101 which were individually aligned with MAFFT (L-INS-i mode) [104] and trimmed with TrimAl (gappyout)
1102 [105]. To enrich phylogenetic signal, only alignments with >30% parsimony-informative sites were

retained using PhyKIT (*pis*) [106], resulting in a final dataset of 2,687 SC-OGs. Maximum likelihood (ML) and coalescence-based species trees were inferred from this dataset. For ML inference, SC-OG alignments were concatenated into a partitioned supermatrix (39 taxa, 2,687 partitions, and 1,625,459 sites) and analyzed with IQ-TREE using the "*-p*" option [107]. For coalescence-based inference, individual gene trees were inferred with IQ-TREE using the "*-S*" option. Branches with <70% UFBoot support were collapsed using Newick utilities (*nw_ed "i & b <= 70" o*), and a species tree was then estimated with ASTRAL v5.7.8. Branch support and genealogical concordance were assessed using SH-aLRT, UFBoot, gene concordance factors (gCF), and site concordance factors (sCF) for ML trees, and local posterior probabilities (LPP) and quartet support for the ASTRAL tree. Phylogenetic trees were visualized and annotated with iTOL v7. Detailed scripts containing software versions, and model parameters used in these analyses, including visualization of gCF/sCF and ASTRAL support, are accessible at https://doi.org/10.5281/zenodo.15686462. The complete set of ML gene trees that served as input for the different analyses, along with the quartet annotations file containing topology partitions and frequency data, is provided as **S1 File**.

ANI, AAI and dDDH

Pairwise average nucleotide identity (ANI) values were computed using OrthoANIu [108]. Custom 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 [109] 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 [39]. Due to the requirement for online submission, representative intra- and interspecies genome comparisons were selected and submitted manually (**S2 Table**). 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 [110] to identify both strain- and species-specific chromosomal rearrangements (**Figs 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 [111], 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 (**Fig 3B**), and *C. porticicola* DSM108351 for clade D (**Fig 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 [112] using BLASTN for alignment. Centromere positions and genome-wide GC content were determined as previously described [30, 32]. 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. For previously generated Nanopore datasets available only in FAST5 format, raw signals were first converted to POD5 using the ONT *pod5 convert fast5* utility. 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 **S3 Table**.

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 [113].

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. To further examine signatures of historical recombination, linkage disequilibrium (LD) decay was assessed using PopLDdecay (`3_run_ld_decay_compare.sh`) with a maximum inter-SNP distance of 250 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.

Biolog phenotypic assays, scoring, cross-validation, 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. Cell density was adjusted to the manufacturer's recommended values for YT microplates, using sterile water for YT and FF plates and Inoculation Fluid B for Gen III plates. Microplates were inoculated with 100 µl of cell suspension per well and incubated in the dark at 25°C. Absorbance was measured immediately after inoculation and then every 3–7 days for up to 21 days at 490 nm, 590 nm, and 750 nm using a Varioskan LUX Multimode Microplate Reader (Thermo Fisher Scientific). Plates were shaken before each reading to ensure a homogeneous suspension.

Because Biolog assays are end-point measurements, wells showing color development or turbidity within the incubation period were scored as positive, whereas those showing no signal and values close to the negative control were scored as negative. Ambiguous cases were re-evaluated against standard assimilation profiles for *Cryptococcus* species and, when possible, retested in traditional liquid or microplate formats following the methods described in Kurtzman et al. (2011) [114]. Because growth dynamics differ between simple and complex substrates, a key challenge in assessing growth was establishing a reliable baseline to distinguish positive from negative growth responses. To ensure robust interpretation across substrates of differing complexity, we implemented a cross-

validation framework drawing on published physiological data [114] and culture-collection databases (CBS WI, PYCC), and by repeating a subset of tests in tubes or custom microplates. Cross-validation covered 12 of the 17 recognized species (71%), including 50% of clade A (*C. neoformans*, *C. deneoformans*, *C. bacillisporus*, *C. gattii*, and *C. hyracis*) and clade B (*C. cederbergensis*) species, and all species from clades C (*C. floricola*, *C. wingfieldii*, *C. amylolentus*, and *C. maquisensis*) and D (*C. porticicola*, *C. ficicola*, and *C jaspeensis*). In total, reference data were available for 46 of the 153 substrates tested (30% overall, corresponding to 49–77% of a single plate) and included representatives of monosaccharides, disaccharides, polysaccharides, polyols, organic acids, and low-molecular-weight aromatic compounds (**S4C Table**). When possible, members of the same clade were processed in parallel, which helped to flag and resolve occasional inconsistencies. These comparisons informed interpretation thresholds and confirmed that the majority of Biolog calls were reproducible across methods.

Growth on glycerol and xylitol was further assessed by spot assays on solid YNB medium (without amino acids). A 10× YNB stock was prepared by dissolving 6.7 g of Difco™ Yeast Nitrogen Base without amino acids and the appropriate carbon source in 100 ml of water, adjusting the pH to 5.6, and sterilizing by filtration. Each medium was normalized to provide the same total carbon as 5 g of glucose [114]. For each 100 ml of final medium, 10 ml of 10× YNB + carbon source was mixed with 90 ml of a 2% agar solution (autoclaved and cooled to ~50°C). Glycerol and xylitol were tested alongside glucose (positive control). YNB without added carbon served as a control for potential nutrient carry-over (negative control). Inocula were prepared by growing the cells for 48 h at room temperature (25 ± 1°C) in 5 ml of YNB + 0.1% glucose (carbon-starvation medium) at 70 rpm [114]. Cells were pelleted, washed twice, adjusted to an optical density of 2 at 600 nm, serially diluted 10-fold, and 3 µl of each dilution was spotted onto the plates. Growth was assessed after 7 days at room temperature (25 ± 1°C).

Phenotypic data from Biolog assays were analyzed with a custom Python pipeline (`0_biolog_analysis_pipeline.py`). Input consisted of a manually curated Excel file (**S4 Table**) 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. Redundant phenotypes, defined as those with identical binary profiles across all strains, were automatically flagged to reduce overrepresentation. All analyses and visualizations were performed using Python with pandas, scikit-learn, seaborn, matplotlib, and related packages.

Standard physiological tests

For standard species descriptions, growth tests were also 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) [114]. 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. The custom microplate format also enabled cross-validating of some substrates (e.g., L-sorbose, D-glucosamine, and D-saccharic acid) or testing of additional substrates (e.g., dulcitol, soluble starch, and aldaric acids such as D-tartaric, and L-tartaric), as well as growth under osmotic stress (NaCl 5%, 8%, and 10%) and in high-glucose conditions (50% and 60%).

Temperature and melanin production assays

For temperature assays, pre-inocula were grown in 4 ml of YPD medium overnight at room temperature. New tubes containing 4 ml of YPD were inoculated with 50 µl of each overnight-grown pre-inocula and incubated at room temperature for 48 hours. Cells were then pelleted, washed with sterile distilled water, and adjusted to an optical density of 2.0 at 600 nm. Serial 10-fold dilutions were prepared from these normalized cultures, and 3 µl of each dilution was spotted onto YPD agar plates. Plates were incubated at 23°C, 30°C, 35°C, and 37°C, and photographed after 3 days. For melanin production assays, 3 µl of each dilution was also spotted onto two melanin-inducing media: L-DOPA agar (7.6 mM L-asparagine, 5.6 mM glucose, 22 mM KH_2PO_4, 1 mM of $MgSO_4.7H_2O$, 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 extract, 0.1% glucose, and 20 g/L agar) [101, 115]. Bird seed agar provides a complex, plant-derived substrate rich in natural phenolic precursors (e.g., caffeic acid and related catechols) that are oxidized by fungal phenoloxidase (laccase), producing the characteristic brown-black pigmentation. In contrast, L-DOPA medium supplies a single defined precursor, allowing a direct test of laccase-dependent melanin synthesis under controlled conditions. A *C. neoformans* H99 *lac1*Δ mutant was used as a negative control. Plates were incubated at room temperature in the dark for 5 days before imaging. All assays were performed in three independent biological replicates to ensure reproducibility of phenotypic results.

Environmental detection via GlobalFungi database

To screen for environmental presence of the newly described *Cryptococcus* species, we performed exact-match searches against the GlobalFungi release 5 database [57], which compiles ITS1 and ITS2 amplicon data from 84,972 samples across 846 studies. Full-length ITS1 and ITS2 regions were extracted from each genome using ITSx [116] implemented in SEED 2 [117]. Unique haplotypes were identified through dereplication in SEED 2, and only sequences identical in both length and

1328 nucleotide composition were considered matches. ITS sequences from additional *Cryptococcus*
1329 species included in the phylogenetic analysis (**Fig 1**) were queried for comparison. Full metadata and
1330 hit summaries are available in **S5 Table**.

Funding Statement

This study was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under awards R01 AI050113-20 (J.H.), R01 AI039115-27 (J.H.), and R01 AI133654-08 (J.H.). We note that these collaborative studies were concluded prior to June 27th, 2025. J.H. is Co-Director and Fellow, and M.C.F. is a Fellow, of the CIFAR program Fungal Kingdom: Threats & Opportunities. Collections leading to the isolation of *C. cederbergensis* were supported entirely and independently by the Harry Oppenheimer Fellowship Award of The Oppenheimer Memorial Trust (to M.J.W.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements

We thank Dr. Seonju Marincowitz and Dr. Nam Pham for their assistance in isolating *Cryptococcus cederbergensis*, Dr. Tatiana A. Kuznetsova for assistance in isolating *C. porticicola* strains, and Dr. Oliver Röhl for assistance in isolating strains OR849 (*C. maquisensis*) and OR918 (*C. jaspeensis*). We also thank Dr. K.J. Kwon-Chung for her foundational contributions to the study of *Cryptococcus* species and for laying the groundwork for the description of the VGV lineage now formally named in this study. We are grateful to Dr. Konstanze Bensch for her valuable suggestions and help with the Latin naming of the new species. We also thank Dr. 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, Miroslav Kolařík

Funding acquisition: Andrey M. Yurkov, Michael J. Wingfield, Joseph Heitman

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

Project Administration: Joseph Heitman

Resources: Andrey M. Yurkov, Aleksey V. Kachalkin, Miroslav Kolařík, Benedetta Turchetti, José Paulo Sampaio, Michael J. Wingfield, Matthew C. Fisher, 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, Aleksey V. Kachalkin, Miroslav Kolařík, Benedetta Turchetti, José Paulo Sampaio, Michael J. Wingfield, Matthew C. Fisher, 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. Fell JW, Boekhout T, Fonseca A, Scorzetti G, Statzell-Tallman A. Biodiversity and systematics of basidiomycetous yeasts as determined by large-subunit rDNA D1/D2 domain sequence analysis. Int J Syst Evol Microbiol. 2000;50 Pt 3:1351-71. doi: 10.1099/00207713-50-3-1351. PubMed PMID: 10843082.

9. Scorzetti G, Fell JW, Fonseca A, Statzell-Tallman A. Systematics of basidiomycetous yeasts: a comparison of large subunit D1/D2 and internal transcribed spacer rDNA regions. FEMS Yeast Res. 2002;2(4):495-517. doi: 10.1111/j.1567-1364.2002.tb00117.x. PubMed PMID: 12702266.

10. Liu XZ, Wang QM, Theelen B, Groenewald M, Bai FY, Boekhout T. Phylogeny of tremellomycetous yeasts and related dimorphic and filamentous basidiomycetes reconstructed from multiple gene sequence analyses. Stud Mycol. 2015;81:1-26. Epub 20151002. doi: 10.1016/j.simyco.2015.08.001. PubMed PMID: 26955196; PubMed Central PMCID: PMC4777771.

11. 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.

12. 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 20231109. doi: 10.1038/s41572-023-00472-z. PubMed PMID: 37945681.

13. 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.

14. 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. PubMed PMID: 25721988.

15. 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. PubMed PMID: 16222245.

16. 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.

17. 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.

18. 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.

19. 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.

20. Lengeler KB, Cox GM, Heitman J. Serotype AD strains of *Cryptococcus neoformans* are diploid or aneuploid and are heterozygous at the mating-type locus. Infect Immun. 2001;69(1):115-22. doi: 10.1128/IAI.69.1.115-122.2001. PubMed PMID: 11119496; PubMed Central PMCID: PMC97862.

21. Vogan AA, Khankhet J, Xu J. Evidence for mitotic recombination within the basidia of a hybrid cross of *Cryptococcus neoformans*. PLOS ONE. 2013;8(5):e62790. Epub 20130514. doi: 10.1371/journal.pone.0062790. PubMed PMID: 23690954; PubMed Central PMCID: PMC3653895.

22. 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.

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

24. 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. PubMed PMID: 20224779; PubMed Central PMCID: PMC2835752.

25. 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.

26. 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.

27. 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.

28. 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.

29. 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. PubMed PMID: 39899605; PubMed Central PMCID: PMC11790112.

30. Coelho MA, David-Palma M, Marincowitz S, Aylward J, Pham NQ, Yurkov AM, et al. The complex evolution and genomic dynamics of mating-type loci in *Cryptococcus* and *Kwoniella*. PLOS Biol. 2025;23(10):e3003417. Epub 20251003. doi: 10.1371/journal.pbio.3003417. PubMed PMID: 41042803; PubMed Central PMCID: PMC12510652.

31. 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.

32. 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.

33. 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.

34. 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.

35. 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.

36. 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.

37. Riesco R, Trujillo ME. Update on the proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. Int J Syst Evol Microbiol. 2024;74(3). doi: 10.1099/ijsem.0.006300. PubMed PMID: 38512750; PubMed Central PMCID: PMC10963913.

38. 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.

39. 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.

40. 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.

41. 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.

42. Farrer RA, Desjardins CA, Sakthikumar S, Gujja S, Saif S, Zeng Q, et al. Genome evolution and innovation across the four major lineages of *Cryptococcus gattii*. mBio. 2015;6(5):e00868-15. Epub 2015/09/04. doi: 10.1128/mBio.00868-15. PubMed PMID: 26330512; PubMed Central PMCID: PMC4556806.

43. 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.

44. Morrow CA, Lee IR, Chow EW, Ormerod KL, Goldinger A, Byrnes EJ, 3rd, et al. A unique chromosomal rearrangement in the *Cryptococcus neoformans* var. *grubii* type strain enhances key phenotypes associated with virulence. MBio. 2012;3(2). doi: 10.1128/mBio.00310-11. PubMed PMID: 22375073; PubMed Central PMCID: PMC3302566.

45. 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.

46. 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.

47. 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.

48. 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.

49. 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.

50. 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.

51. 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.

52. 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.

53. 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.

54. 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.

55. 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.

56. 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.

57. Větrovský T, Morais D, Kohout P, Lepinay C, Algora C, Awokunle Holla S, et al. GlobalFungi, a global database of fungal occurrences from high-throughput-sequencing metabarcoding studies. Sci Data. 2020;7(1):228. Epub 20200713. doi: 10.1038/s41597-020-0567-7. PubMed PMID: 32661237; PubMed Central PMCID: PMC7359306.

58. Kazartsev I, Shorohova E, Kapitsa E, Kushnevskaya H. Decaying *Picea abies* log bark hosts diverse fungal communities. Fungal Ecol. 2018;33:1-12. doi: https://doi.org/10.1016/j.funeco.2017.12.005.

59. Cregger MA, Veach AM, Yang ZK, Crouch MJ, Vilgalys R, Tuskan GA, et al. The *Populus* holobiont: dissecting the effects of plant niches and genotype on the microbiome. Microbiome. 2018;6(1):31. Epub 20180212. doi: 10.1186/s40168-018-0413-8. PubMed PMID: 29433554; PubMed Central PMCID: PMC5810025.

60. Franco Ortega S, Ferrocino I, Adams I, Silvestri S, Spadaro D, Gullino ML, et al. Monitoring and surveillance of aerial mycobiota of rice paddy through DNA metabarcoding and qPCR. J Fungi (Basel). 2020;6(4). Epub 20201217. doi: 10.3390/jof6040372. PubMed PMID: 33348656; PubMed Central PMCID: PMC7766667.

61. Gschwend F, Hartmann M, Mayerhofer J, Hug AS, Enkerli J, Gubler A, et al. Site and land-use associations of soil bacteria and fungi define core and indicative taxa. FEMS Microbiol Ecol. 2022;97(12). doi: 10.1093/femsec/fiab165. PubMed PMID: 34940884; PubMed Central PMCID: PMC8752248.

62. Sannino C, Borruso L, Mezzasoma A, Battistel D, Ponti S, Turchetti B, et al. Abiotic factors affecting the bacterial and fungal diversity of permafrost in a rock glacier in the Stelvio Pass (Italian Central Alps). Appl Soil Ecol. 2021;166:104079. doi: https://doi.org/10.1016/j.apsoil.2021.104079.

63. Luis P, Saint-Genis G, Vallon L, Bourgeois C, Bruto M, Marchand C, et al. Contrasted ecological niches shape fungal and prokaryotic community structure in mangroves sediments. Environ Microbiol. 2019;21(4):1407-24. Epub 20190319. doi: 10.1111/1462-2920.14571. PubMed PMID: 30807675.

64. 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.

65. 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.

66. 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.

67. 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.

68. 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.

69. 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.

70. 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.

71. 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.

72. 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.

73. 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.

74. 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.

75. 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 20080815. doi: 10.1007/s00294-008-0201-2. PubMed PMID: 18704425.

76. 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.

77. De Carvalho M, Jia GS, Nidamangala Srinivasa A, Billmyre RB, Xu YH, Lange JJ, et al. The *wtf* meiotic driver gene family has unexpectedly persisted for over 100 million years. Elife. 2022;11. Epub 20221013. doi: 10.7554/eLife.81149. PubMed PMID: 36227631; PubMed Central PMCID: PMC9562144.

78. 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.

79. 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.

80. Lin X, Heitman J. Chlamydospore formation during hyphal growth in *Cryptococcus neoformans*. Eukaryot Cell. 2005;4(10):1746-54. doi: 10.1128/EC.4.10.1746-1754.2005. PubMed PMID: 16215181; PubMed Central PMCID: PMC1265899.

81. Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, et al. Population genomics of domestic and wild yeasts. Nature. 2009;458(7236):337-41. Epub 20090211. doi: 10.1038/nature07743. PubMed PMID: 19212322; PubMed Central PMCID: PMC2659681.

82. Cadete RM, Melo MA, Dussan KJ, Rodrigues RC, Silva SS, Zilli JE, et al. Diversity and physiological characterization of D-xylose-fermenting yeasts isolated from the Brazilian Amazonian Forest. PLOS ONE. 2012;7(8):e43135. Epub 2012/08/23. doi: 10.1371/journal.pone.0043135. PubMed PMID: 22912807; PubMed Central PMCID: PMC3418277.

83. Suh SO, Marshall CJ, McHugh JV, Blackwell M. Wood ingestion by passalid beetles in the presence of xylose-fermenting gut yeasts. Mol Ecol. 2003;12(11):3137-45. Epub 2003/11/25. doi: 10.1046/j.1365-294x.2003.01973.x. PubMed PMID: 14629392.

84. Blackwell M. Made for each other: ascomycete yeasts and insects. Microbiol Spectr. 2017;5(3). Epub 2017/06/10. doi: 10.1128/microbiolspec.FUNK-0081-2016. PubMed PMID: 28597823; PubMed Central PMCID: PMC11687504.

85. Yurkov AM. Yeasts of the soil - obscure but precious. Yeast. 2018;35(5):369-78. Epub 20180302. doi: 10.1002/yea.3310. PubMed PMID: 29365211; PubMed Central PMCID: PMC5969094.

86. Kostál V, Zahradníčková H, Simek P, Zelený J. Multiple component system of sugars and polyols in the overwintering spruce bark beetle, *Ips typographus*. J Insect Physiol. 2007;53(6):580-6. Epub 2007/05/08. doi: 10.1016/j.jinsphys.2007.02.009. PubMed PMID: 17481650.

87. 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.

88. 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.

89. Hu G, Cheng PY, Sham A, Perfect JR, Kronstad JW. Metabolic adaptation in *Cryptococcus neoformans* during early murine pulmonary infection. Mol Microbiol. 2008;69(6):1456-75. Epub 2008/08/05. doi: 10.1111/j.1365-2958.2008.06374.x. PubMed PMID: 18673460; PubMed Central PMCID: PMC2730461.

90. Price MS, Betancourt-Quiroz M, Price JL, Toffaletti DL, Vora H, Hu G, et al. *Cryptococcus neoformans* requires a functional glycolytic pathway for disease but not persistence in the host. mBio. 2011;2(3):e00103-11. Epub 2011/06/10. doi: 10.1128/mBio.00103-11. PubMed PMID: 21652778; PubMed Central PMCID: PMC3110414.

91. Zhao Y, Lin X. A PAS protein directs metabolic reprogramming during cryptococcal adaptation to hypoxia. mBio. 2021;12(2). Epub 2021/03/18. doi: 10.1128/mBio.03602-20. PubMed PMID: 33727360; PubMed Central PMCID: PMC8092316.

92. 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. doi: 10.1128/EC.4.1.202-208.2005. PubMed PMID: 15643075; PubMed Central PMCID: PMC544170.

93. 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 20061113. doi: 10.1128/IAI.01351-06. PubMed PMID: 17101662; PubMed Central PMCID: PMC1828480.

94. 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. doi: 10.1128/EC.4.1.190-201.2005. PubMed PMID: 15643074; PubMed Central PMCID: PMC544166.

95. 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.

96. Polacheck I, Hearing VJ, Kwon-Chung KJ. Biochemical studies of phenoloxidase and utilization of catecholamines in *Cryptococcus neoformans*. J Bacteriol. 1982;150(3):1212-20. doi: 10.1128/jb.150.3.1212-1220.1982. PubMed PMID: 6804439; PubMed Central PMCID: PMC216342.

97. 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 20200417. doi: 10.1073/pnas.2001451117. PubMed PMID: 32303657; PubMed Central PMCID: PMC7211991.

98. 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 20230120. doi: 10.1073/pnas.2209831120. PubMed PMID: 36669112; PubMed Central PMCID: PMC9942834.

99. 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 20250523. doi: 10.1101/2025.05.19.654958. PubMed PMID: 40475665; PubMed Central PMCID: PMC12139950.

100. 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.

101. 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 20241119. doi: 10.1371/journal.ppat.1012735. PubMed PMID: 39561188; PubMed Central PMCID: PMC11614259.

102. 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.

103. 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.

104. 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.

105. 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.

106. 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.

107. 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.

108. 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 20170215. doi: 10.1007/s10482-017-0844-4. PubMed PMID: 28204908.

109. 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 20210428. doi: 10.1007/s12275-021-1154-0. PubMed PMID: 33907973.

110. 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.

111. 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.

112. 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.

113. 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.

114. 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.

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

116. Bengtsson-Palme J, Ryberg M, Hartmann M, Branco S, Wang Z, Godhe A, et al. Improved software detection and extraction of ITS1 and ITS2 from ribosomal ITS sequences of fungi and other eukaryotes for analysis of environmental sequencing data. Methods Ecol Evol. 2013;4(10):914-9. doi: https://doi.org/10.1111/2041-210X.12073.

117. Vetrovský T, Baldrian P, Morais D. SEED 2: a user-friendly platform for amplicon high-throughput sequencing data analyses. Bioinformatics. 2018;34(13):2292-4. doi: 10.1093/bioinformatics/bty071. PubMed PMID: 29452334; PubMed Central PMCID: PMC6022770.

1741 ## Figure legends

1742

1743 **Fig 1. Phylogenomic relationships and species delimitation within *Cryptococcus*. (A)** Maximum
1744 likelihood (ML) phylogeny inferred from a concatenated data matrix comprising protein alignments of
1745 2,687 single-copy orthologs (SC-OGs) across 39 strains, representing 17 *Cryptococcus* species
1746 (including six newly proposed species indicated in boldface) and rooted with the outgroup *Kwoniella*
1747 *shandongensis*. Gene and site concordance factors (gCF|sCF) are shown in grey at nodes separating
1748 species and major clades. A red circle highlights a conflict-prone node (N43), while a green circle marks
1749 an example of a well-supported node (N74). Support values from SH-aLRT and ultrafast bootstrap
1750 (UFboot) exceed 95% for all nodes and were omitted for clarity but are available in the corresponding
1751 tree file (https://doi.org/10.5281/zenodo.15686462). Branch lengths are shown as the number of
1752 substitutions per site (scale bar). Major clades are labeled as partitions A–D. For reference, lineages
1753 designations are provided for *C. neoformans* (*Cn*: VNI, VNBI, VNBII, VNII), *C. deneoformans* (*Cd*: VNIV)
1754 and *C. gattii* complex (*Cg*: VGI–VGVI), and chromosome numbers are indicated for each species.
1755 Asterisks next to *C. hyracis* strains denote that this species has not yet been associated with human
1756 infections. **(B)** Species tree inferred with ASTRAL, which summarizes gene tree topologies under a
1757 multispecies coalescent model, using the same set of 2,687 single-copy orthologs (SC-OGs). Colored
1758 circles indicate the corresponding nodes highlighted in panel A, with red marking a conflict-prone node
1759 (N32) and green marking an example of a well-supported node (N1). Support values in grey represent
1760 the quartet support for the main topology (q1). Local posterior probabilities (LPP) are ≥ 0.90 for all nodes
1761 and are not shown for simplicity but are available in the corresponding tree file
1762 (https://doi.org/10.5281/zenodo.15686462). Major clades are labeled as partitions A–D. **(C)** Overview
1763 of the phylogenomic workflow. A total of 3,510 single-copy orthologs (SC-OGs) were initially identified
1764 using OrthoFinder across 39 genomes. These were filtered based on alignment length, sequence
1765 completeness, and phylogenetic informativeness, resulting in 2,687 high-confidence SC-OGs. Protein
1766 alignments were used to infer individual gene trees and a concatenated supermatrix. The resulting

1767 datasets were analyzed with maximum likelihood (ML) methods (panel A) and the coalescent-based
1768 ASTRAL approach (panel B). **(D)** Gene and site concordance at two representative nodes: N43 (conflict-
1769 prone) and N74 (well-supported). Bar plots on the left show gene concordance factors (gCF) and gene
1770 discordance factors (gDF1, gDF2, and gDFP), representing the proportion of gene trees that support
1771 the main topology (gCF, dark blue), each of the two alternative topologies (gDF1 and gDF2, light blue),
1772 or that display conflicting relationships due to paraphyly of one or more clades (gDFP, pale blue). The
1773 corresponding site concordance (sCF) and site discordance (sDF1 and sDF2) values are shown on the
1774 right, based on site-wise likelihood support for each topology. The red dashed line at 33.3% indicates
1775 the expected support under a hard polytomy, where all three resolutions are equally likely due to lack
1776 of phylogenetic signal. **(E)** ASTRAL quartet support for the same nodes shown in panel D: N32 (conflict-
1777 prone) and N1 (well-supported). Bar plots show the normalized frequencies of the three possible quartet
1778 topologies around each node, based on gene tree topologies. The main topology is shown in dark green,
1779 with alternatives in lighter shades. Local posterior probabilities (LPP) for each topology are indicated
1780 above the bars. The red dashed line (33.3%) reflects the expected distribution under a hard polytomy.

1781

1782 **Fig 2. Genomic divergence and species boundaries among *Cryptococcus* species. (A–B)**
1783 UPGMA clustering based on pairwise sequence divergence, calculated as 100 – Average Nucleotide
1784 Identity (ANI) in panel A and 100 – Average Amino Acid Identity (AAI) in panel B. Vertical dashed lines
1785 indicate divergence thresholds: red marks the highest observed divergence within a species (i.e., the
1786 lowest ANI or AAI between *C. neoformans* lineages), and green marks the lowest observed divergence
1787 between species (i.e., the highest ANI or AAI between *C. bacillisporus* and *C. decagattii*). **(C)** Scatterplot
1788 comparing ANI and AAI values across all pairwise genome comparisons. Intraspecies comparisons are
1789 shown in orange and interspecies comparisons in blue. The two interspecies pairs with the highest
1790 genomic similarity are highlighted: *C. bacillisporus* (*Cbac*) vs. *C. decagattii* (*Cdec*), and *C. hyracis* (*Chyr*)
1791 vs. *C. tetragattii* (*Ctet*). Pearson and Spearman correlation coefficients indicate a strong positive
1792 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.

Fig 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

1818 generated using a custom Python (available at https://doi.org/10.5281/zenodo.15686462) based on
1819 pairwise alignments from minimap2.

1820

1821 **Fig 4. Mating, recombination signatures, and chromosomal structural variation in *Cryptococcus***
1822 ***porticicola* (A)** Maximum likelihood phylogeny of clade D species (pruned from Fig. 1), highlighting *C.*
1823 *porticicola*, *C. ficicola*, and *C. jaspeensis*. Labels indicate genome assembly status (complete or draft),
1824 chromosome number inferred from telomere-to-telomere assemblies, mating pheromone receptor allele
1825 (*STE3* a1/α or *STE3* a2/a), and *HD* allele identity (b1–b4) inferred from sequence comparisons. **(B)**
1826 Mating assay between *C. porticicola* strains DSM108352 (*a1b3*) and NCYC1536 (*a2b2*), showing
1827 hyphal formation and putative sexual structures. Solo cultures are shown on the left and center panels;
1828 the cross is shown on the right, where hyphal formation is observed embedded in the agar. Basidia-like
1829 structures are visible and resemble sexual development in *Kwoniella* species. Assays were performed
1830 on V8 medium (pH 5.0) at room temperature in the dark for 1 month. Scale bars: 200 μm (50 μm in the
1831 inset). **(C)** Light microscopy images of hyphal and putative sexual structures recovered from the mating
1832 plate shown in panel B. The left panel shows hyphae and rounded cells resembling basidia (arrows),
1833 with a zoomed-in view in the middle panel. The right panel highlights a hyphal segment with an
1834 apparently unfused clamp connection (arrow). Scale bars: 10 μm. **(D)** Neighbor-Net phylogenetic
1835 network constructed from core SNPs showing signals of conflicting phylogenetic relationships
1836 (reticulation), consistent with recombination. **(E)** Principal component analysis (PCA) of core SNP data
1837 showing genetic structure among the six strains. The first two principal components explain 46.5% and
1838 21.3% of the total variance, respectively. **(F)** Linkage disequilibrium (LD) decay curve showing mean
1839 pairwise r^2 values (minor allele frequency, MAF ≥ 0.167) as a function of physical distance between
1840 SNPs. Dashed line shows lowess-smoothed trend. **(G)** Chromosomal synteny comparisons across *C.*
1841 *porticicola* employing strain DSM108351 as reference. Each block represents a chromosome; ribbons
1842 denote orthologous regions as determined by the ntSynt toolkit. Chromosomes were reordered to
1843 maximize collinearity, and those inverted relative to their original assembly orientations are marked with

1844 asterisks. **(H)** Detailed view of intraspecific chromosomal rearrangements in *C. porticicola*. Centromeres
1845 (circles) and synteny breakpoints are shown for each rearrangement. ICR: intercentromeric
1846 recombination.

1847

1848 **Fig 5. Karyotype evolution and centromere dynamics in clade D *Cryptococcus* species. (A)**
1849 Chromosome-length plots for six telomere-to-telomere assemblies representing *C. porticicola*, *C.*
1850 *ficicola*, and *C. jaspeensis*. Colors indicate synteny blocks, with *C. porticicola* DSM108351 as the
1851 reference. Chromosome number differs across species (9, 8, and 7, respectively), and multiple
1852 structural changes are observed, including species- and strain-specific translocations (TR). Circles
1853 mark predicted centromeres: white for active centromeres (*CEN*) and dark gray for putative inactivated
1854 centromeres (*iCEN*). For *C. porticicola*, the total number of translocations (including those consistent
1855 with intercentromeric recombination, ICR) is noted below each strain. **(B)** Example of centromere loss
1856 in *C. jaspeensis*. A region syntenic to *CEN7* in *C. porticicola* and *C. ficicola* appears inactivated in *C.*
1857 *jaspeensis* OR918 (*iCEN* on chr. 1), but still shows 5mCG methylation, (see **S6F Fig** for a zoomed-in
1858 view of this region). Track descriptions are provided on the key on the left. **(C–D)** Synteny comparisons
1859 showing putative centromere inactivation and repositioning in *C. ficicola* and *C. jaspeensis*. In panel C,
1860 *CEN5* of *C. porticicola* DSM108351 aligns with *CEN8* in *C. ficicola* and *CEN7* in *C. jaspeensis*,
1861 suggesting a shared ancestral origin (zoomed view in panel D, region I). By contrast, *CEN9* in *C.*
1862 *porticicola* appears to have been inactivated (*iCEN*) in both *C. ficicola* and *C. jaspeensis*, coinciding
1863 with reduced transposable element content and diminished 5mCG methylation signal, and conservation
1864 of flanking genes aside from a minor inversion (panel D, region III). The predicted centromere locations
1865 in chr. 4 of *C. ficicola* and chr. 5 of *C. jaspeensis* lack orthology with any of *C. porticicola* centromeres,
1866 but the corresponding region in *C. porticicola* exhibits low levels of transposons and 5mCG methylation
1867 (panel D, region II). This region may represent a repositioned or newly co-opted centromere (labeled
1868 *nCEN*) in the common ancestor of *C. ficicola* and *C. jaspeensis*, or alternatively may have harbored an
1869 ancestral centromere that was subsequently lost or inactivated in *C. porticicola*.

1870

1871 **Fig 6. Variation in ribosomal RNA (rRNA) gene organization across *Cryptococcus* species. (A)**
1872 Cladogram of *Cryptococcus* species (adapted from Figure 1), annotated with rRNA gene organization
1873 types. All pathogenic species from clade A exhibit a fully clustered rRNA array (28S–5.8S–18S–5S). In
1874 contrast, saprobic species display unclustered configurations with lineage-specific patterns: dispersed
1875 5S genes (e.g., clade B), 5S enrichment on a single chromosome (e.g., clade C), or subtelomeric 5S
1876 enrichment (e.g., *C. ficicola*, *C. jaspeensis* in clade D). **(B)** Genome-wide distribution of rRNA genes
1877 (28S, 5.8S, 18S, 5S) across four representative species: *C. deneoformans* (clade A), *C. floricola* (clade
1878 C), *C. porticicola*, and *C. jaspeensis* (both clade D). Arrowheads indicate the position of the core rDNA
1879 array, and predicted centromeres are marked with open circles. In clade C, 5S genes are significantly
1880 enriched on a single chromosome (e.g., chr. 10 in *C. floricola*, see inset), whereas in clade D species,
1881 5S genes are predominantly subtelomeric (e.g., chr. 2 in *C. jaspeensis*). See **S7 Fig** for additional
1882 species and strains. **(C)** Zoomed views of representative rDNA clusters and subtelomeric 5S regions.
1883 TEL, telomere.

1884

1885 **Fig 7. Phenotypic profiling distinguishes *Cryptococcus* clades and pathogenic species. (A)**
1886 Clustermap showing hierarchical clustering of strains (columns) and informative phenotypic tests
1887 (rows), based on binary growth/no-growth responses in Biolog assays. Of the 153 substrates tested,
1888 112 showed variable responses across strains and were retained for clustering. Growth was assessed
1889 using Biolog YT, FF, and GEN III MicroPlates. Binary values indicate positive growth (brown) or absence
1890 of growth (beige). Strains are colored by phylogenetic clade, and clustering was performed using
1891 average linkage and Euclidean distance. Most strains cluster according to clade, with metabolic profiles
1892 broadly reflecting phylogenetic structure. **(B)** Principal component analysis (PCA) of Biolog growth
1893 profiles. The first two components explain 34.6% of the total variance. Ellipses represent 95%
1894 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. Strains were grown for 48 hours in YPD, washed, and spotted in 10-fold dilutions starting from an OD600 of 2.0. 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. In *C. wingfieldii*, one laccase-like gene is truncated and was therefore excluded from the count (marked with an asterisk). See **S8C Fig** for growth at other temperatures and results on alternative melanin-inducing medium (L-DOPA).

Supporting information

S1 Fig. Expanded view of phylogenomic support and conflict among clade A *Cryptococcus* species. (A) Pruned maximum likelihood (ML) phylogeny (from Figure 1A), showing only clade A species. The tree was inferred from a concatenated alignment of 2,687 single-copy orthologs (SC-OGs). Conflict-prone nodes are marked in red and labeled by node ID. Concordance factors based on gene trees (gCF) and site patterns (sCF) are shown in grey. All nodes are fully supported by SH-aLRT and ultrafast bootstrap (UFboot = 100%). **(B)** Pruned ASTRAL species tree (from Figure 1B), based on the same SC-OGs. All nodes have local posterior probability (LPP) = 1.0. Quartet support for the main topology is shown in grey. Nodes corresponding to those in panel A are labeled by their ASTRAL node IDs. **(C)** Gene and site concordance at three conflict-prone nodes (N47, N52, N61). Gene concordance factors (gCF) and discordance factors (gDF1, gDF2, and gDFP) are shown on the left. The main topology is shown in dark blue; alternatives and paraphyly are shown in lighter blue shades. Site concordance (sCF) and discordance (sDF1, sDF2) are shown on the right. The red dashed line at 33.3% represents the expected support under a hard polytomy. **(D)** ASTRAL quartet support for the corresponding nodes (N28, N23, N16). Bar plots show the normalized frequency of each of the three

possible quartet topologies, with the main topology shown in dark green and the two alternatives in lighter shades. Local posterior probabilities (LPP) are indicated above each bar. In all cases, ASTRAL strongly supports the main topology (LPP = 1.0), despite the presence of gene tree discordance in the ML-based analyses (panel C). This highlights how coalescent-based methods can recover consistent species relationships even in the presence of substantial incomplete lineage sorting, which likely underlies the low gCF and high gDFP values observed at these nodes.

S2 Fig. Mating assays demonstrate the ability of *Cryptococcus hyracis* to undergo sexual reproduction. (A) Pairwise mating assays using two *C. hyracis MAT*α strains (MF34 and MF13) crossed with *MAT***a** strains from other species in the *C. gattii* complex. Crosses were performed on V8 agar and examined by light microscopy after 2–3 weeks. Hyphal growth with basidia and spores (inset) indicated successful mating; positive and negative interactions are marked with "+" and "–", respectively. Strain MF13 generally showed stronger mating responses than MF34. For *C. bacillisporus*, we tested both the wild-type strain B4546 and a derived *crg1*Δ mutant. *CRG1* encodes a negative regulator of pheromone signaling that acts via G-protein signaling by stimulating GTP hydrolysis to GDP, thereby extinguishing Gα-GTP signaling. Thus, deletion mutants are hypersensitive to mating signals and serve as sensitive indicators of mating potential. Scale bars = 200 μm (10 μm in insets). **(B)** Positive control cross between *C. neoformans MAT*α (KN99α) and *MAT***a** (KN99**a**) showing robust filamentation and basidia formation. **(C–E)** Scanning electron microscopy (SEM) of interspecies mating interactions involving *C. hyracis*. Panels show examples of basidia and basidiospore formation, as well as fused clamp connections, observed in crosses with *C. bacillisporus* (C) and *C. tetragattii* (D–E). These structures confirm the ability of *C. hyracis* to engage in sexual development.

S3 Fig. Karyotype comparisons among *C. neoformans* and *C. deneoformans* strains. (A) Synteny analysis of 14 chromosomes across multiple *C. neoformans* strains representing the four major lineages

(VNI, VNII, VNBI, VNBII) and two *C. deneoformans* (VNIV) strains. Chromosomes were ordered to maximize collinearity with the *C. neoformans* strain 125.91 reference genome, and those inverted relative to their original orientations are indicated with asterisks. Overall synteny is well conserved across *C. neoformans* strains, except for strain H99 (VNI), which harbors a known translocation between chrs. 3 and 11. *C. deneoformans* strains are largely syntenic to *C. neoformans*, except for some large inversions. The locations of the rDNA cluster and MAT locus are indicated. Centromere positions are marked by grey triangles. Rearrangements, including large inversions and translocations, were identified using the ntSynt toolkit and visualized with ntSynt-viz. **(B)** Schematic of the previously characterized reciprocal translocation in *C. neoformans* H99 between chrs. 3 and 11. Dashed lines indicate the breakpoint positions, and grey triangles mark centromere locations. This rearrangement distinguishes H99 from other VNI strains.

S4 Fig. Karyotype comparisons among *C. gattii* (sensu lato) species. (A) Synteny comparisons between *C. gattii* strain WM276 (VGI, reference) and representative strains from each of the species within the *C. gattii* complex. Chromosomes were reordered to maximize collinearity with the reference, and those inverted relative to their original assembly orientations are marked with asterisks. Synteny blocks were identified using the ntSynt toolkit and visualized with ntSynt-viz. Despite broad conservation of synteny, numerous rearrangements are observed across species, including translocations and inversions. The chromosomal locations of the rDNA cluster and the *MAT* locus are also indicated, as are predicted centromeres (triangles). **(B)** Schematic representations of selected interspecific chromosomal rearrangements. Colored bars represent putative ancestral chromosomes and their rearranged derivatives in extant species, with breakpoints indicated by dashed lines. Centromeres are shown as circles. Panels highlight key rearrangements across different nodes of the phylogeny, including several likely intercentromeric recombination (ICR) events, such as that shared between *C. bacillisporus* and *C. decagattii* (event b) and the derived translocation found in *C. tetragattii* CBS11718 (event f).

1972

1973 **S5 Fig. Mating assays between *C. porticicola* strains under two different media conditions. (A)**
1974 Summary table indicating presence (+?) and absence (–) of hyphal and basidial-like structures. "+?"
1975 denotes putative structures consistent with sexual development, though the identity of basidia-like
1976 structures could not be definitively confirmed. Mating-type alleles (*STE3* and *HD*) are shown for each
1977 strain. **(B)** Representative light microscopy images from V8 pH 5.0 medium assays. Only the
1978 DSM108352 (a1b3) × NCYC1536 (a2b2) cross showed hyphal growth and potential sexual structures
1979 embedded in the agar (arrow); the inset highlights a structure resembling a basidium. Scale bars:
1980 200 μm (50 μm for inset). **(C)** Summary table of mating outcomes on MS medium. **(D)** Representative
1981 light microscopy images from MS medium assays. Pseudohyphal growth, characterized by constricted
1982 septa and chain-like elongation, is observed in the solo culture of NCYC1536 and in the DSM108351 ×
1983 NCYC1536 cross (arrows). True hyphae growth, defined by continuous non-constricted filaments, is
1984 observed in the DSM108352 × NCYC1536 cross (arrow), although no apparent basidial-like structures
1985 were detected throughout the course of the experiment (marked as "–*"). Scale bars: 200 μm unless
1986 otherwise noted (50 μm for insets).

1987

1988 **S6 Fig. Synteny and structural context of conserved and apparently lost centromeres across**
1989 **clade D *Cryptococcus* species. (A–E)** Synteny comparisons across conserved centromeric regions
1990 in *C. porticicola* (strains DSM111204, DSM108352, NCYC1536, and DSM108351), *C. ficicola*
1991 (DSM109063), and *C. jaspeensis* (OR918). Tracks highlight conserved centromere positioning and
1992 surrounding synteny for *CEN1*, *CEN2*, *CEN3*, *CEN4*, and *CEN6* of *C. porticicola* DSM108351 and their
1993 corresponding orthologous regions in the other strains and species. **(F)** Apparent loss of *CEN7* in *C.*
1994 *jaspeensis* OR918. The top panel shows chromosome-scale synteny across strains, highlighting the
1995 conserved position of *CEN7* in all species except *C. jaspeensis*, where the orthologous region lacks
1996 canonical centromeric signatures and is marked as putative inactivated centromere (*iCEN*). The bottom

panel provides a zoomed-in view of this region with tracks showing synteny, 5mCG DNA methylation, and TE content. Although the centromeric repeat landscape appears disrupted in OR918, residual DNA methylation and partial synteny may indicate a recently inactivated centromere. **(G)** Apparent centromere loss in both *C. ficicola* and *C. jaspeensis*. The top panel shows chromosome-scale synteny, highlighting the position of *CEN8* from *C. porticicola* DSM108351. In strain DSM111204, this region underwent intercentromeric recombination, while in *C. ficicola* and *C. jaspeensis* the corresponding region lacks the repeat-rich structure typical of active centromeres. The bottom panel shows a zoomed-in view of this region, indicating that the orthologous loci in DSM109063 and OR918 lack the repeat-rich structure characteristic of active or ancestral centromeres, and only residual 5mCG DNA methylation, consistent with inactivated centromeres (*iCEN*).

S7 Fig. Genome-wide organization of ribosomal RNA (rRNA) genes across *Cryptococcus* species. Chromosomal distribution of 28S, 5.8S, 18S, and 5S rRNA genes across all analyzed *Cryptococcus* strains with chromosome-level assemblies. Arrowheads indicate the position of the core rDNA array, and predicted centromeres are marked with open circles. In *C. floricola*, chr. 7 is represented as two joined contigs with a gap at *CEN7* (marked with an asterisk). The rDNA array (28S–5.8S–18S) is typically clustered on a single chromosome, whereas 5S genes show variable distribution patterns across lineages. **(A)** Clade A (pathogenic) species exhibit fully clustered rRNA arrays, while saprobic species from other clades **(B-D)** display unclustered configurations with lineage-specific features: scattered 5S genes (e.g., *C. depauperatus* and *C. porticicola*), chromosome-specific enrichment (e.g., *C. floricola*), or subtelomeric localization (e.g., *C. jaspeensis*). In *C. amylolentus*, the 28S–5.8S–18S rDNA cluster is absent from the current assembly owing to unresolved repetitive regions near the end of chr. 9. The inferred position, marked with an arrowhead labeled “(a),” is supported by a few long reads from the original sequencing dataset that extend into the rDNA region.

2022 **S8 Fig. Phenotypic traits distinguishing *Cryptococcus* species and clades**.
2023 **(A)** Top 15 phenotypic tests contributing most to clade-level classification based on Random Forest
2024 analysis of binarized Biolog data. Substrates are ranked by feature importance in predicting clade
2025 membership. **(B)** Mean presence frequency of the top 15 substrates from panel A across *Cryptococcus*
2026 clades. Some substrates (e.g., glycerol, xylitol, glucuronamide) display clade-specific patterns, with
2027 utilization largely restricted to or absent from particular clades, highlighting their potential utility for
2028 phenotypic discrimination. **(C)** Growth of all tested strains at 23°C, 30°C, 35°C, and 37°C on YPD
2029 medium for 4 days. Cells were grown for 48 hours in YPD, washed, and spotted in 10-fold dilutions
2030 starting from an OD600 of 2.0. Robust growth at 37°C was observed exclusively in clade A strains,
2031 consistent with thermotolerance as a pathogenicity-associated trait. No strains from saprobic species
2032 in clades B–D showed appreciable growth at this temperature. Among saprobes, *C. cederbergensis*
2033 displayed limited growth above 30°C, whereas *C. maquisensis* OR849 exhibited weak growth at 35°C
2034 (visible in the highest inoculum spot). A colony marked with an "X" represents contamination and should
2035 be disregarded. **(D)** Melanin production on L-DOPA agar plates incubated at room temperature for 5
2036 days. Spotting assay was conducted as described in C. All clade A strains produced visible melanin,
2037 while all nonpathogenic strains from clades B–D lacked pigmentation under these conditions. Among
2038 pathogens, melanin production was somewhat reduced in *C. decagattii* CBS11687 and *C.*
2039 *deneoformans* NIH12 under this growth condition.

2040

2041 **S9 Fig. Validation of glycerol and xylitol assimilation across *Cryptococcus* species.** Spot assays
2042 testing growth on defined minimal medium (YNB without amino acids) containing glucose, glycerol, or
2043 xylitol as the sole carbon source. YNB without added carbon served as a control for potential nutrient
2044 carry-over. Cells were pre-grown for 48 h at room temperature (25 ± 1°C) in YNB + 0.1% glucose
2045 (carbon-starvation medium), washed, and spotted in 10-fold dilutions starting from an OD600 of 2.
2046 Plates were incubated at room temperature for 7 days before imaging. Pathogenic clade A species

exhibited little or no growth on glycerol or xylitol, whereas non-pathogenic species (clades B–D) grew well under the same conditions.

S1 Table. List of strains used in this study, and summary of genome assembly statistics and other genomic features.

S2 Table. ANI, AAI, and dDDH metrics across *Cryptococcus* genomes. **(A)** Full matrix of ANI values computed for all pairwise genome comparisons. **(B)** Summary of pairwise ANI values between species. **(C)** Min., max., and mean ANI values calculated between and within clades, based on grouped pairwise comparisons. **(D)** Full AAI values computed for all pairwise genome comparisons. **(E)** Summary of pairwise AAI values between species. **(F)** Min., max., and Mean AAI values calculated between and within clades, based on grouped pairwise comparisons. **(G)** Closest interspecies genome comparisons based on identity values. **(H)** Digital DNA-DNA hybridization (dDDH) values from Genome-to-Genome Distance Calculator (GGDC). **(I)** dDDH values with species-level summary and strain annotations.

S3 Table. **rRNA gene analysis and statistical tests**. **(A)** Genome-wide summary of rRNA gene counts. **(B)** Genome-wide summary of 5S rRNA gene counts. **(C)** Chromosomal distribution of 5S rRNA genes, observed vs. expected. **(D)** Chromosomal enrichment of 5S rRNA genes and statistical analyses. **(E)** Summary of 5S gene localization relative to subtelomeric regions.

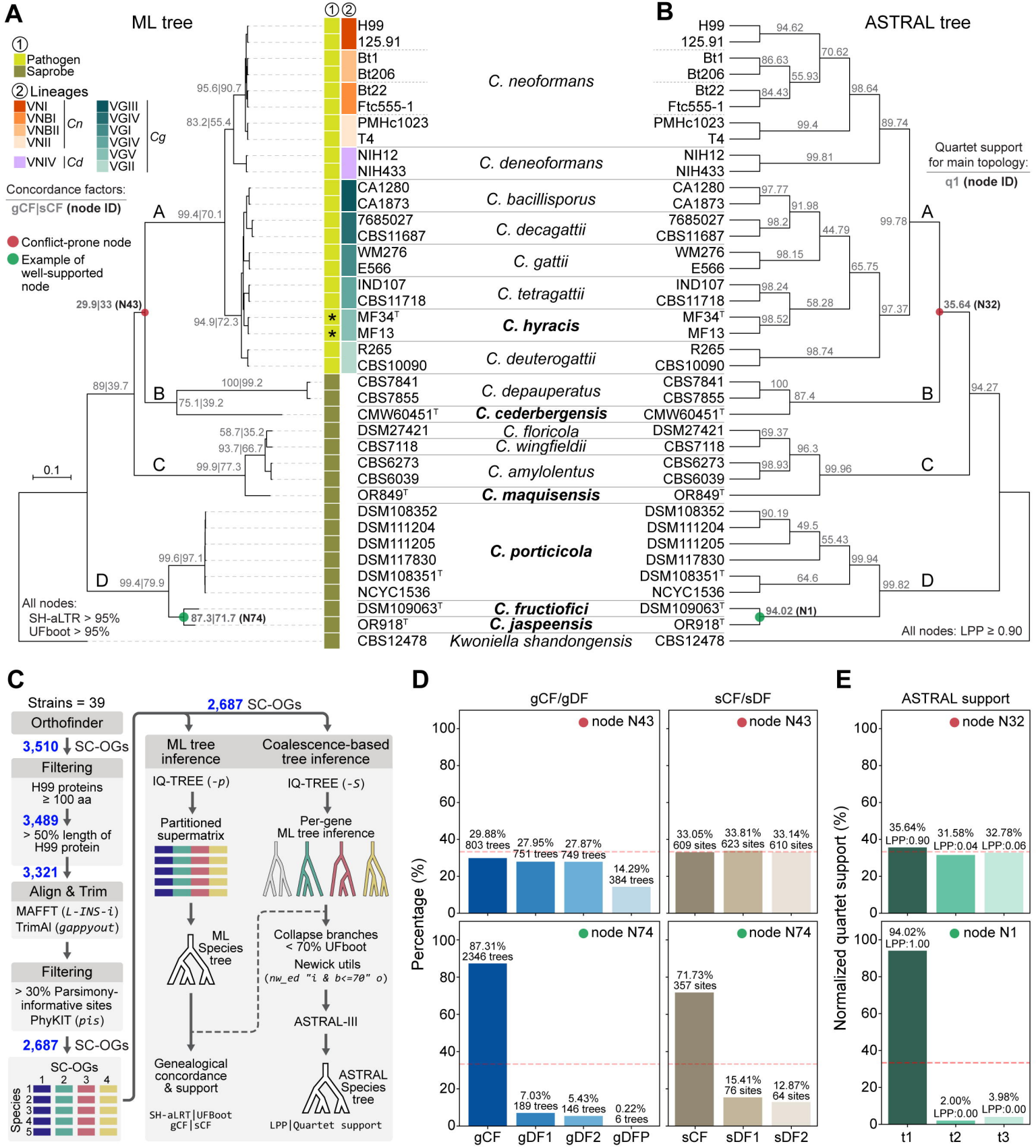
S4 Table. **Phenotypic profiling results**. **(A)** Full list of tests. **(B)** Informative tests showing variable response across strains. **(C)** Cross-validation of assays.

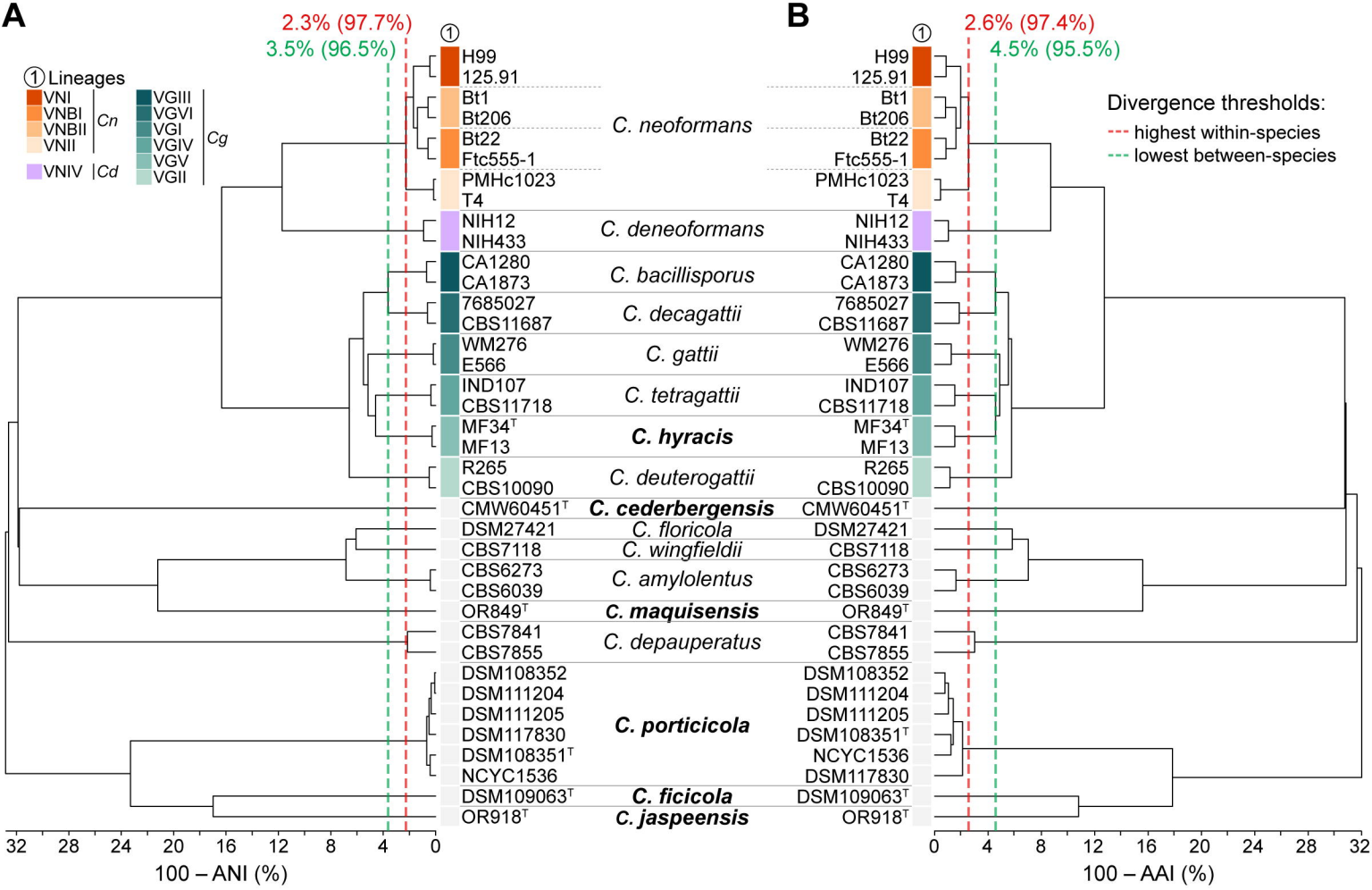
S5 Table. Summary of GlobalFungi ITS matches to newly described *Cryptococcus* species. Table listing exact ITS1 and ITS2 matches identified in the GlobalFungi database for each *Cryptococcus* species. Includes marker resolution (delimitation power) of ITS1 and ITS2 across species, match type (ITS1 or ITS2), geographic origin, substrate, and study metadata. Data correspond to matches visualized in S2 File.

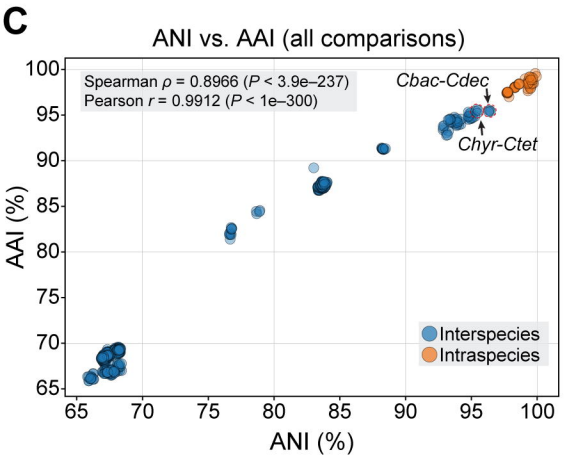
S6 Table. Genome sequencing and information on raw sequencing data generated in this study. (A) Genome sequencing, assembly, polishing, and gene prediction/annotation approaches. **(B)** NCBI accession numbers of each genome and raw read data generated and used in this study.

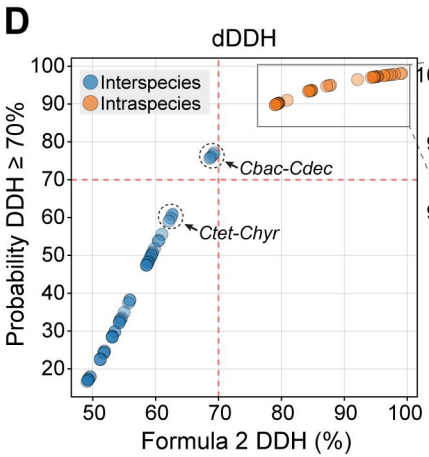
S1 File. Gene tree and quartet topology datasets used for concordance-factor and ASTRAL analyses. Includes ML gene trees (trimal_pergene.treefile), trees with branches <70 % UFboot collapsed for ASTRAL (trimal_pergene.BS70.treefile), and the ASTRAL quartet annotation file with topology partitions and frequencies (trimal_pergene.t16.freqquad.csv).

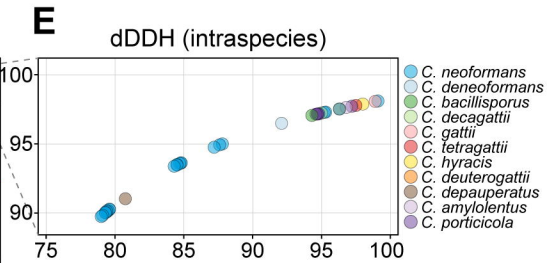
S2 File. Interactive map of GlobalFungi sampling sites. Interactive HTML map showing GlobalFungi sampling sites with exact ITS1 or ITS2 matches to newly described *Cryptococcus* species, as summarized in S5 Table.


A
ML tree
① Pathogen
Saprobe
② Lineages
VNI
VNBI
VNBII
VNII
Cn
VNIV
Cd
VGIII
VGIV
VGI
VGIV
VGV
VGII
Cg
Concordance factors:
gCF|sCF (node ID)
Conflict-prone node
Example of well-supported node
29.9|33 (N43)
95.6|90.7
83.2|55.4
99.4|70.1
94.9|72.3
89|39.7
100|99.2
75.1|39.2
58.7|35.2
93.7|66.7
99.9|77.3
99.6|97.1
99.4|79.9
87.3|71.7 (N74)
0.1
All nodes:
SH-aLTR > 95%
UFboot > 95%
H99
125.91
Bt1
Bt206
Bt22
Ftc555-1
PMHc1023
T4
NIH12
NIH433
CA1280
CA1873
7685027
CBS11687
WM276
E566
IND107
CBS11718
MF34T
MF13
R265
CBS10090
CBS7841
CBS7855
CMW60451T
DSM27421
CBS7118
CBS6273
CBS6039
OR849T
DSM108352
DSM111204
DSM111205
DSM117830
DSM108351T
NCYC1536
DSM109063T
OR918T
CBS12478
C. neoformans
C. deneoformans
C. bacillisporus
C. decagattii
C. gattii
C. tetragattii
C. hyracis
C. deuterogattii
C. depauperatus
C. cederbergensis
C. floricola
C. wingfieldii
C. amylolentus
C. maquisensis
C. porticicola
C. fructiofici
C. jaspeensis
Kwoniella shandongensis
B
ASTRAL tree
Quartet support for main topology:
q1 (node ID)
35.64 (N32)
94.02 (N1)
All nodes: LPP ≥ 0.90
C
Strains = 39
Orthofinder
3,510 SC-OGs
Filtering
H99 proteins ≥ 100 aa
3,489
> 50% length of H99 protein
3,321
Align & Trim
MAFFT (L-INS-i)
TrimAl (gappyout)
Filtering
> 30% Parsimony-informative sites
PhyKIT (pis)
2,687 SC-OGs
SC-OGs
Species
2,687 SC-OGs
ML tree inference
Coalescence-based tree inference
IQ-TREE (-p)
IQ-TREE (-S)
Partitioned supermatrix
Per-gene ML tree inference
ML Species tree
Collapse branches < 70% UFboot
Newick utils
(nw_ed "i & b<=70" o)
ASTRAL-III
Genealogical concordance & support
ASTRAL Species tree
SH-aLRT|UFBoot
gCF|sCF
LPP|Quartet support
D
gCF/gDF
sCF/sDF
node N43
node N74
Percentage (%)
29.88% 803 trees
27.95% 751 trees
27.87% 749 trees
14.29% 384 trees
33.05% 609 sites
33.81% 623 sites
33.14% 610 sites
87.31% 2346 trees
7.03% 189 trees
5.43% 146 trees
0.22% 6 trees
71.73% 357 sites
15.41% 76 sites
12.87% 64 sites
gCF
gDF1
gDF2
gDFP
sCF
sDF1
sDF2
E
ASTRAL support
node N32
node N1
Normalized quartet support (%)
35.64% LPP:0.90
31.58% LPP:0.04
32.78% LPP:0.06
94.02% LPP:1.00
2.00% LPP:0.00
3.98% LPP:0.00
t1
t2
t3





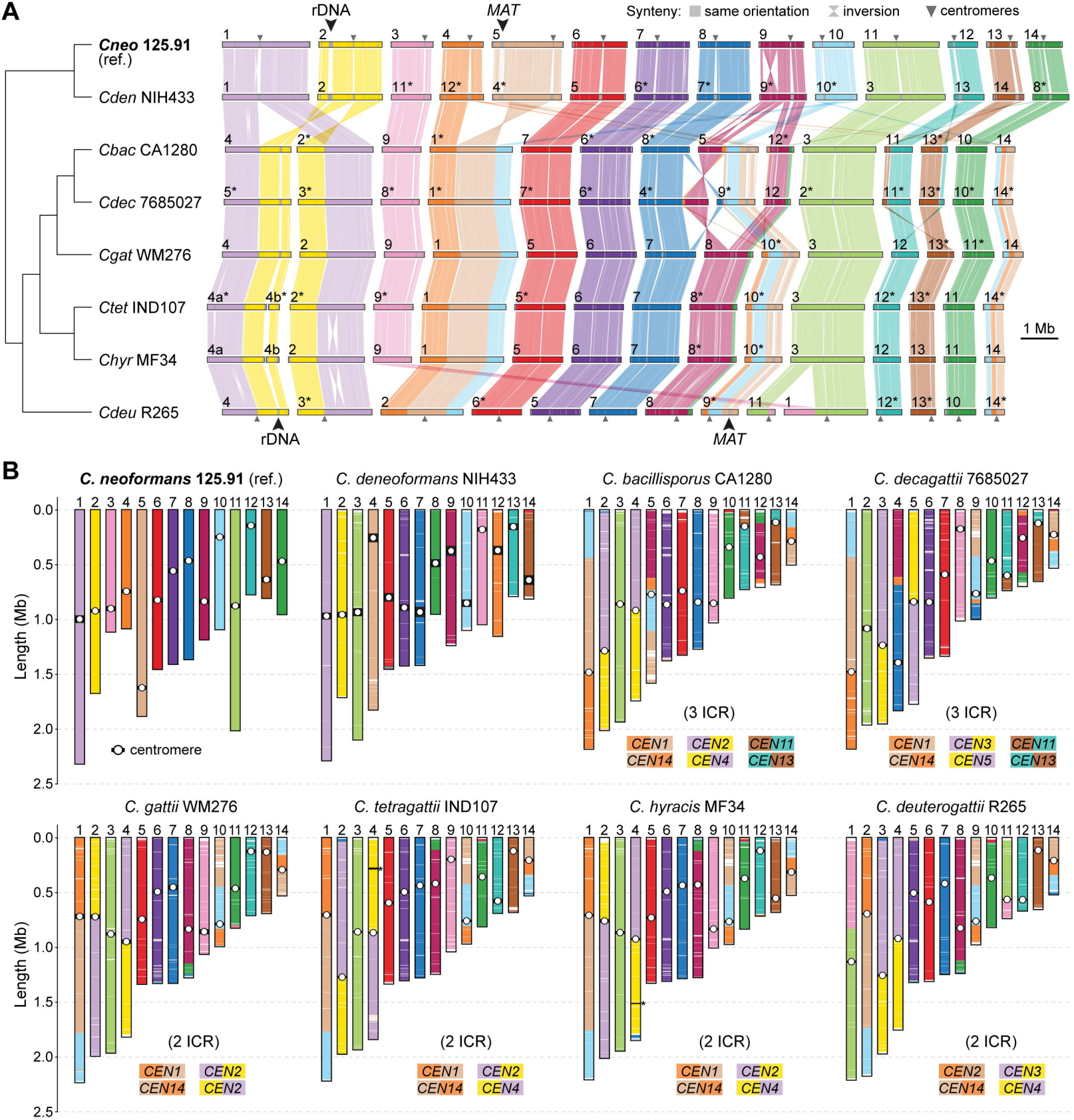


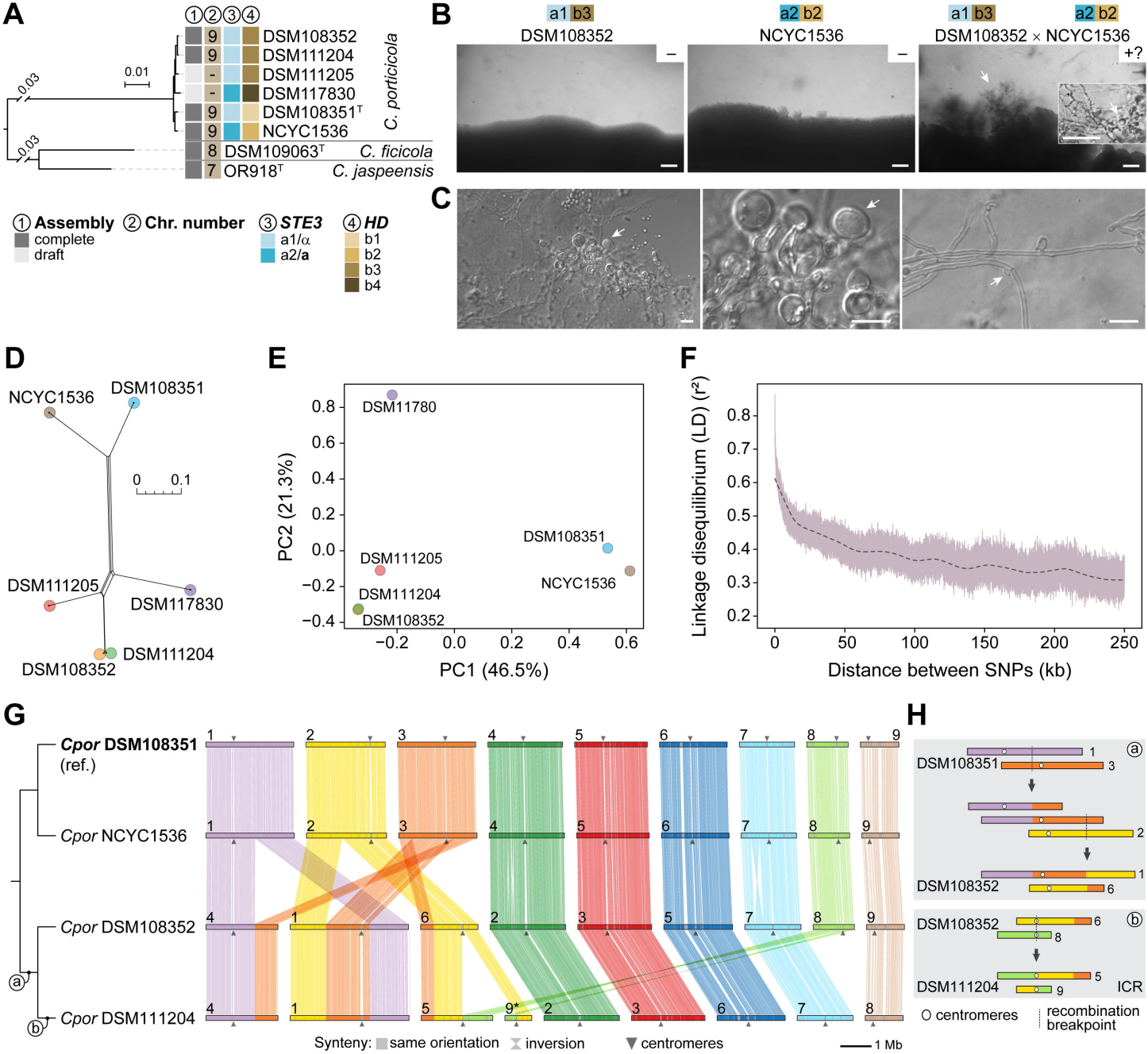


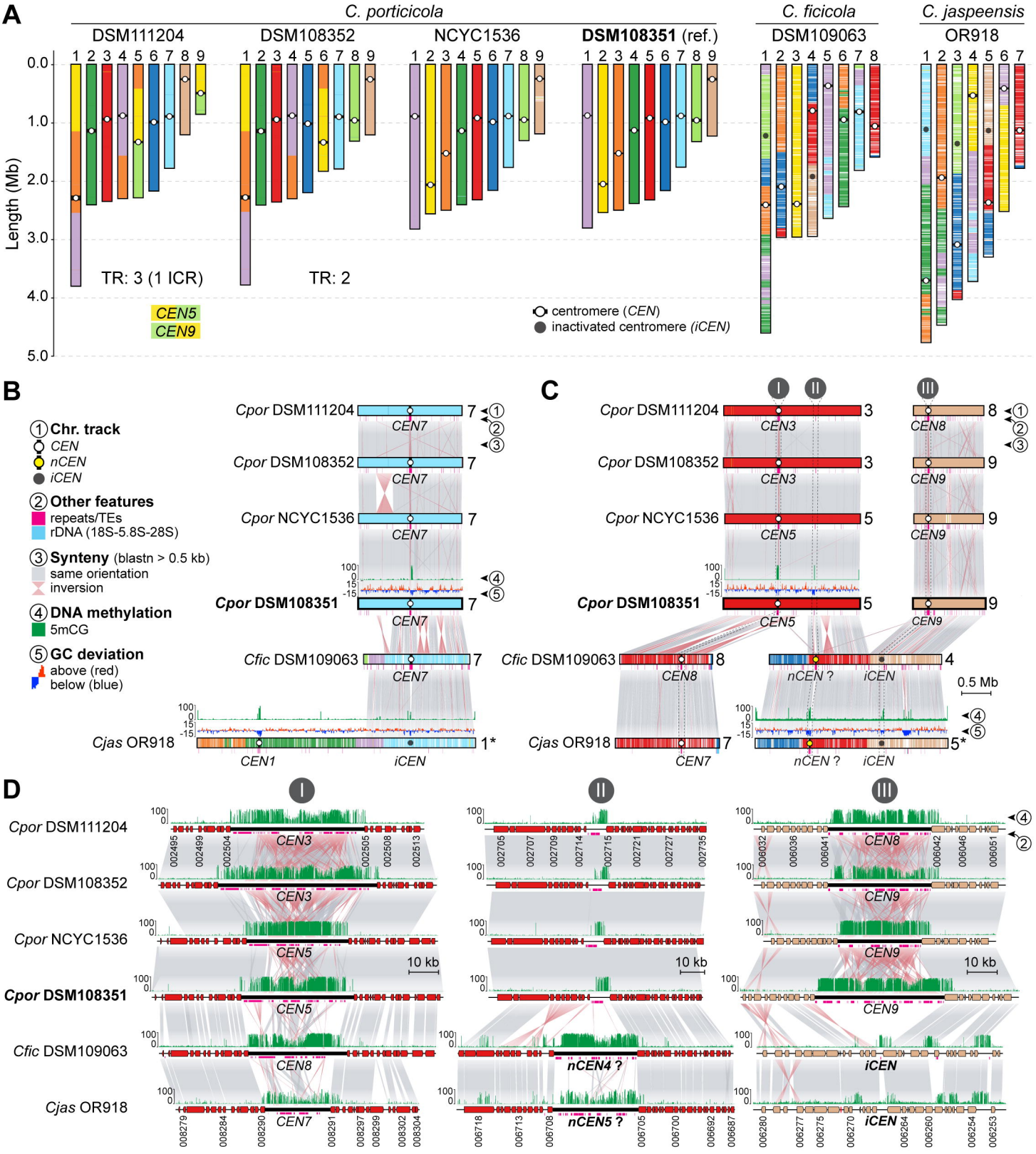
F

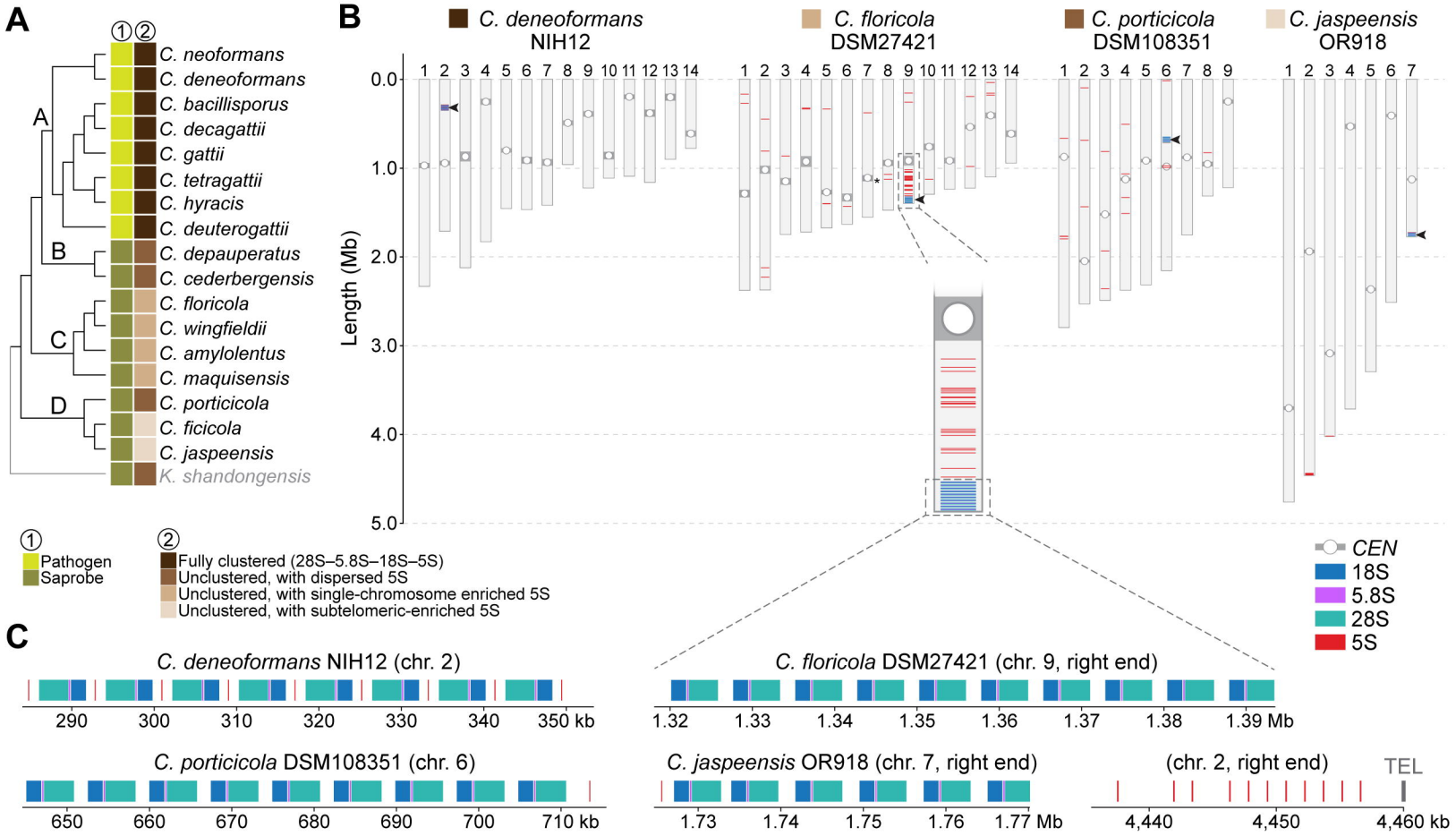
dDDH (interspecies)

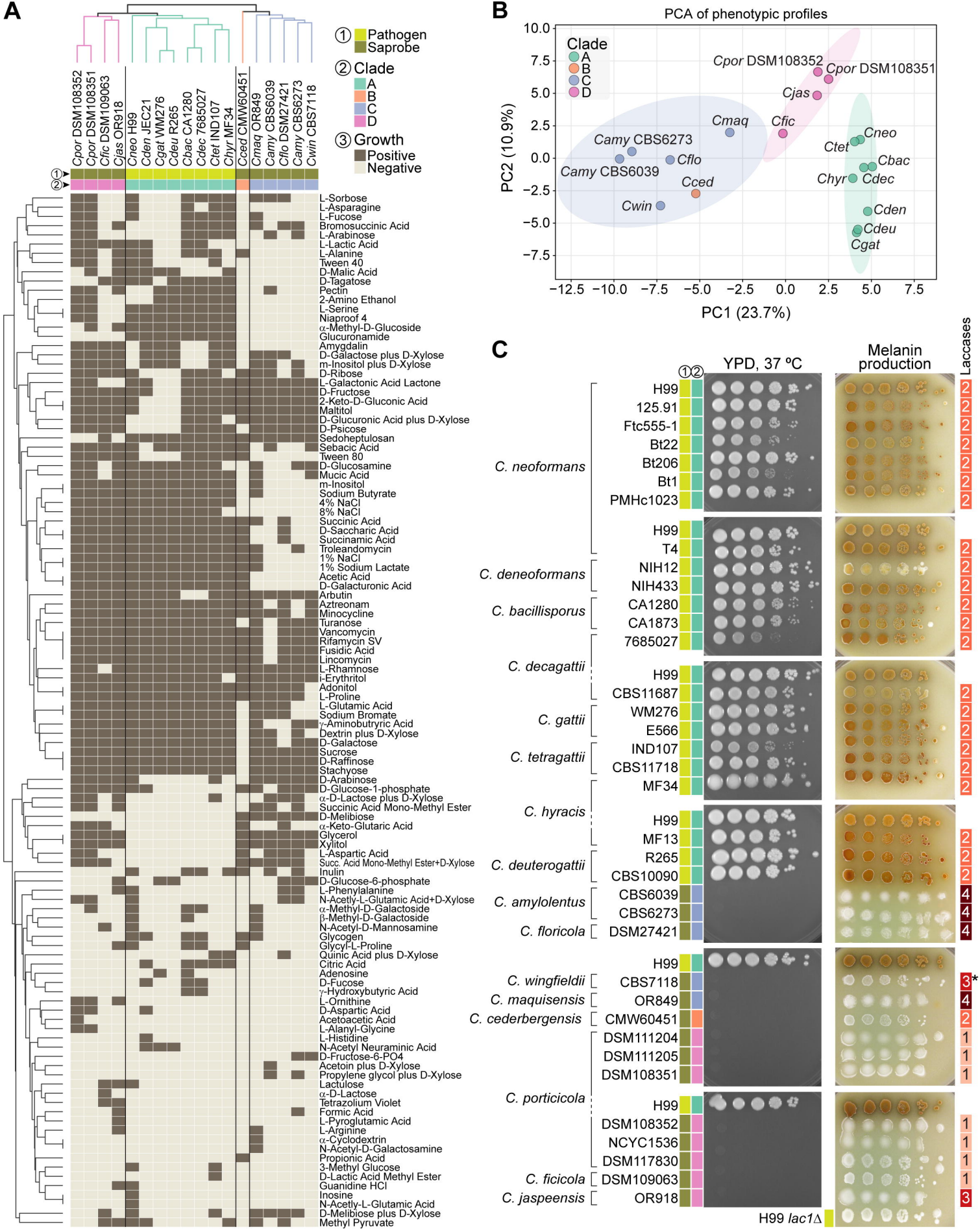
Pair	Strain 1	Strain 2	DDH%	Prob. ≥ 70%
1	*Cbac* CA1873	*Cdec* CBS11687	69.3	77.08
2	*Cbac* CA1280	*Cdec* 7685027	68.6	75.66
3	*Ctet* IND107	*Chyr* MF13	62.8	61.25
4	*Ctet* CBS11718	*Chyr* MF34	62.1	59.06


A
rDNA
MAT
Synteny: same orientation inversion centromeres
Cneo 125.91 (ref.)
Cden NIH433
Cbac CA1280
Cdec 7685027
Cgat WM276
Ctet IND107
Chyr MF34
Cdeu R265
1 Mb
B
C. neoformans 125.91 (ref.)
C. deneoformans NIH433
C. bacillisporus CA1280
C. decagattii 7685027
C. gattii WM276
C. tetragattii IND107
C. hyracis MF34
C. deuterogattii R265
Length (Mb)
centromere
(3 ICR)
CEN1 CEN2 CEN11 CEN14 CEN4 CEN13
(3 ICR)
CEN1 CEN3 CEN11 CEN14 CEN5 CEN13
(2 ICR)
CEN1 CEN2 CEN14 CEN2
(2 ICR)
CEN1 CEN2 CEN14 CEN4
(2 ICR)
CEN1 CEN2 CEN14 CEN4
(2 ICR)
CEN2 CEN3 CEN14 CEN4


A
① ② ③ ④
DSM108352
DSM111204
DSM111205
DSM117830
DSM108351T
NCYC1536
DSM109063T
OR918T
C. porticicola
C. ficicola
C. jaspeensis
0.01
0.03
0.03
① Assembly
complete
draft
② Chr. number
③ STE3
a1/α
a2/a
④ HD
b1
b2
b3
b4
B
a1 b3
DSM108352
a2 b2
NCYC1536
a1 b3
a2 b2
DSM108352 × NCYC1536
–
–
+?
C
D
NCYC1536
DSM108351
0
0.1
DSM111205
DSM117830
DSM111204
DSM108352
E
DSM11780
DSM108351
DSM111205
NCYC1536
DSM111204
DSM108352
PC2 (21.3%)
PC1 (46.5%)
F
Linkage disequilibrium (LD) (r²)
Distance between SNPs (kb)
G
Cpor DSM108351 (ref.)
Cpor NCYC1536
Cpor DSM108352
Cpor DSM111204
Synteny:
same orientation
inversion
centromeres
1 Mb
H
DSM108351
DSM108352
DSM108352
DSM111204
ICR
centromeres
recombination breakpoint


A
C. porticicola
C. ficicola
C. jaspeensis
DSM111204
DSM108352
NCYC1536
DSM108351 (ref.)
DSM109063
OR918
Length (Mb)
TR: 3 (1 ICR)
TR: 2
CEN5
CEN9
centromere (CEN)
inactivated centromere (iCEN)
B
① Chr. track
CEN
nCEN
iCEN
② Other features
repeats/TEs
rDNA (18S-5.8S-28S)
③ Synteny (blastn > 0.5 kb)
same orientation
inversion
④ DNA methylation
5mCG
⑤ GC deviation
above (red)
below (blue)
Cpor DSM111204
Cpor DSM108352
Cpor NCYC1536
Cpor DSM108351
Cfic DSM109063
Cjas OR918
CEN7
CEN1
iCEN
C
Cpor DSM111204
Cpor DSM108352
Cpor NCYC1536
Cpor DSM108351
Cfic DSM109063
Cjas OR918
CEN3
CEN5
CEN8
CEN9
nCEN ?
iCEN
CEN7
0.5 Mb
D
Cpor DSM111204
Cpor DSM108352
Cpor NCYC1536
Cpor DSM108351
Cfic DSM109063
Cjas OR918
CEN3
CEN5
CEN8
CEN7
CEN9
nCEN4 ?
nCEN5 ?
iCEN
10 kb


A
①
②
C. neoformans
C. deneoformans
C. bacillisporus
C. decagattii
C. gattii
C. tetragattii
C. hyracis
C. deuterogattii
C. depauperatus
C. cederbergensis
C. floricola
C. wingfieldii
C. amylolentus
C. maquisensis
C. porticicola
C. ficicola
C. jaspeensis
K. shandongensis
A
B
C
D
① Pathogen
Saprobe
② Fully clustered (28S–5.8S–18S–5S)
Unclustered, with dispersed 5S
Unclustered, with single-chromosome enriched 5S
Unclustered, with subtelomeric-enriched 5S
B
C. deneoformans NIH12
C. floricola DSM27421
C. porticicola DSM108351
C. jaspeensis OR918
Length (Mb)
CEN
18S
5.8S
28S
5S
C
C. deneoformans NIH12 (chr. 2)
290 300 310 320 330 340 350 kb
C. floricola DSM27421 (chr. 9, right end)
1.32 1.33 1.34 1.35 1.36 1.37 1.38 1.39 Mb
C. porticicola DSM108351 (chr. 6)
650 660 670 680 690 700 710 kb
C. jaspeensis OR918 (chr. 7, right end)
1.73 1.74 1.75 1.76 1.77 Mb
(chr. 2, right end)
TEL
4,440 4,450 4,460 kb


A
① Pathogen
Saprobe
② Clade
A
B
C
D
③ Growth
Positive
Negative
Cpor DSM108352
Cpor DSM108351
Cfic DSM109063
Cjas OR918
Cneo H99
Cden JEC21
Cgat WM276
Cdeu R265
Cbac CA1280
Cdec 7685027
Ctet IND107
Chyr MF34
Cced CMW60451
Cmaq OR849
Camy CBS6039
Cflo DSM27421
Camy CBS6273
Cwin CBS7118
L-Sorbose
L-Asparagine
L-Fucose
Bromosuccinic Acid
L-Arabinose
L-Lactic Acid
L-Alanine
Tween 40
D-Malic Acid
D-Tagatose
Pectin
2-Amino Ethanol
L-Serine
Niaproof 4
α-Methyl-D-Glucoside
Glucuronamide
Amygdalin
D-Galactose plus D-Xylose
m-Inositol plus D-Xylose
D-Ribose
L-Galactonic Acid Lactone
D-Fructose
2-Keto-D-Gluconic Acid
Maltitol
D-Glucuronic Acid plus D-Xylose
D-Psicose
Sedoheptulosan
Sebacic Acid
Tween 80
D-Glucosamine
Mucic Acid
m-Inositol
Sodium Butyrate
4% NaCl
8% NaCl
Succinic Acid
D-Saccharic Acid
Succinamic Acid
Troleandomycin
1% NaCl
1% Sodium Lactate
Acetic Acid
D-Galacturonic Acid
Arbutin
Aztreonam
Minocycline
Turanose
Vancomycin
Rifamycin SV
Fusidic Acid
Lincomycin
L-Rhamnose
i-Erythritol
Adonitol
L-Proline
L-Glutamic Acid
Sodium Bromate
γ-Aminobutyric Acid
Dextrin plus D-Xylose
D-Galactose
Sucrose
D-Raffinose
Stachyose
D-Arabinose
D-Glucose-1-phosphate
α-D-Lactose plus D-Xylose
Succinic Acid Mono-Methyl Ester
D-Melibiose
α-Keto-Glutaric Acid
Glycerol
Xylitol
L-Aspartic Acid
Succ. Acid Mono-Methyl Ester+D-Xylose
Inulin
D-Glucose-6-phosphate
L-Phenylalanine
N-Acetyl-L-Glutamic Acid+D-Xylose
α-Methyl-D-Galactoside
β-Methyl-D-Galactoside
N-Acetyl-D-Mannosamine
Glycogen
Glycyl-L-Proline
Quinic Acid plus D-Xylose
Citric Acid
Adenosine
D-Fucose
γ-Hydroxybutyric Acid
L-Ornithine
D-Aspartic Acid
Acetoacetic Acid
L-Alanyl-Glycine
L-Histidine
N-Acetyl Neuraminic Acid
D-Fructose-6-PO4
Acetoin plus D-Xylose
Propylene glycol plus D-Xylose
Lactulose
α-D-Lactose
Tetrazolium Violet
Formic Acid
L-Pyroglutamic Acid
L-Arginine
α-Cyclodextrin
N-Acetyl-D-Galactosamine
Propionic Acid
3-Methyl Glucose
D-Lactic Acid Methyl Ester
Guanidine HCl
Inosine
N-Acetly-L-Glutamic Acid
D-Melibiose plus D-Xylose
Methyl Pyruvate
B
PCA of phenotypic profiles
Clade
A
B
C
D
Cpor DSM108352
Cpor DSM108351
Cjas
Cmaq
Cfic
Camy CBS6273
Cneo
Ctet
Cbac
Camy CBS6039
Cflo
Chyr
Cdec
Cced
Cwin
Cden
Cdeu
Cgat
PC2 (10.9%)
PC1 (23.7%)
10.0
7.5
5.0
2.5
0.0
−2.5
−5.0
−7.5
−12.5
−10.0
−7.5
−5.0
−2.5
0.0
2.5
5.0
7.5
C
YPD, 37 ºC
Melanin production
Laccases
C. neoformans
H99
125.91
Ftc555-1
Bt22
Bt206
Bt1
PMHc1023
H99
T4
C. deneoformans
NIH12
NIH433
C. bacillisporus
CA1280
CA1873
C. decagattii
7685027
H99
CBS11687
C. gattii
WM276
E566
C. tetragattii
IND107
CBS11718
C. hyracis
MF34
H99
MF13
C. deuterogattii
R265
CBS10090
C. amylolentus
CBS6039
CBS6273
C. floricola
DSM27421
H99
C. wingfieldii
CBS7118
C. maquisensis
OR849
C. cederbergensis
CMW60451
DSM111204
DSM111205
DSM108351
C. porticicola
H99
DSM108352
NCYC1536
DSM117830
C. ficicola
DSM109063
C. jaspeensis
OR918
H99 lac1Δ
3*