

## Genetic and genomic analyses reveal boundaries between species closely related to *Cryptococcus* pathogens

Andrew Ryan Passer<sup>1¶</sup>, Marco A. Coelho<sup>1¶</sup>, Robert Blake Billmyre<sup>1#a</sup>, Minou Nowrousian<sup>2</sup>, Moritz Mittelbach<sup>3#b</sup>, Andrey M. Yurkov<sup>4</sup>, Anna Floyd Averette<sup>1</sup>, Christina A. Cuomo<sup>5</sup>, Sheng Sun<sup>1\*</sup>, and Joseph Heitman<sup>1\*</sup>

<sup>1</sup> Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, North Carolina, USA

<sup>2</sup> Lehrstuhl für Allgemeine und Molekulare Botanik, Ruhr-Universität Bochum, Bochum, Germany

<sup>3</sup> Geobotany, Faculty of Biology and Biotechnology, Ruhr-University Bochum, 44801 Bochum, Germany

<sup>4</sup> Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

<sup>5</sup> Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA

<sup>#a</sup>Current address: Stowers Institute for Medical Research, Kansas City, Missouri, USA

<sup>#b</sup>Current address: Plant Protection Office Berlin, Mohriner Allee 137, 12347 Berlin, Germany

\* Corresponding authors:

Email: sheng.sun@duke.edu (SS) and heitm001@duke.edu (JH)

¶These authors contributed equally to this work

## Abstract

Speciation is a central mechanism of biological diversification. While speciation is well studied in plants and animals, in comparison, relatively little is known about speciation in fungi. One fungal model is the *Cryptococcus* genus, which is best known for the pathogenic *Cryptococcus neoformans*/*Cryptococcus gattii* species complex that causes over 200,000 new infections in humans annually. The closest non-human pathogenic relatives are the sibling species, *Cryptococcus amyloletus* and *Tsuchiyaea wingfieldii*. However, because relatively few isolates of each species are available, it is unclear whether they represent divergent lineages of the same species or different biological species. The recent isolation of an additional strain, preliminarily identified as *T. wingfieldii*, prompted us to reexamine this group as it may inform about the evolutionary processes underlying the diversification of both non-pathogenic and pathogenic *Cryptococcus* lineages. Using genomic data, we reappraised the phylogenetic relationship of the four available strains and confirmed the genetic separation of *C. amyloletus* and *T. wingfieldii* (now *Cryptococcus wingfieldii*), and revealed an additional cryptic species, for which the name *Cryptococcus floricola* is proposed. Comparison of full-length chromosome assemblies revealed approximately 6% pairwise sequence divergence between the three species, and identified significant genomic changes, including inversions as well as a reciprocal translocation that involved inter-centromeric ectopic recombination, which together likely impose significant barriers to genetic exchange. Using genetic crosses, we show that while *C. wingfieldii* cannot interbreed with any of the other strains, *C. floricola* can undergo sexual reproduction with *C. amyloletus*. However, most of the spores resulting from this cross were inviable, and many were sterile, indicating that the two species are genetically isolated through intrinsic post-zygotic barriers and possibly due to niche differentiation. Genome sequencing and analysis of the progeny demonstrated decreased recombination frequency during meiosis in heterospecific crosses compared to *C. amyloletus* conspecific crosses. This study advances our

understanding of speciation in fungi and highlights the power of genomics in assisting our ability to correctly identify and discriminate fungal species.

## Author Summary

The idea of species as discrete natural units seems rather intuitive for most people, just as cells are the basic units of life. However, when observing variation across a species range, boundaries can become blurred making it less than obvious when different populations evolve into separate species. Additionally, separate species can still interbreed, such as lions breeding with tigers to produce a liger or a tigon (depending on the paternal and maternal species of origin), but the resulting offspring is usually inviable or sterile, which in turn is evidence that the parents involved are distinct species. Therefore, what species are and how they originate is still an open question in evolutionary biology. While recent advances have been made in the fields of animal and plant speciation, many other important components of biological diversity, such as fungi, are still understudied. Genome sequencing is now providing new tools to address the genetic mechanisms that drive divergence and reproductive isolation between populations, including genetic incompatibilities, sequence divergence, and chromosomal rearrangements. Here we focus on the *Cryptococcus amyloletus* species complex, a non-pathogenic fungal lineage closely related to the human pathogenic *Cryptococcus neoformans*/*Cryptococcus gattii* complex. Using genetic and genomic analysis we reexamined the species boundaries of four available isolates within the *C. amyloletus* complex and revealed three genetically isolated species. The genomes of these species are ~6% divergent and exhibit chromosome rearrangements, including translocations and small-scale inversions. Although two of the species (*C. amyloletus* and newly described *C. floricola*) are still able to interbreed, the resulting hybrid progeny were mostly inviable, and many were sterile, indicating that barriers to reproduction

have already been established. Our results will foster additional studies addressing the transitions between non-pathogenic and pathogenic *Cryptococcus* lineages.

## Introduction

Speciation is the process by which two species are formed from a common ancestor and is one of the main biological processes generating biodiversity. For organisms that reproduce sexually, including most animals and plants, as well as many fungi, this process implies the development of reproductive barriers inhibiting gene flow between diverging populations. These barriers can be pre- and post-zygotic depending on whether they operate before or after fertilization. Pre-zygotic barriers can be environmental, such as temporal and/or spatial separation of sexual reproduction among closely related species, or they can be biological, when mating is prevented through unsuccessful gamete recognition. Post-zygotic reproductive barriers are often associated with the generation of hybrid progeny that are either inviable or sterile and are expected to arise as a result of divergence between nascent species. In these cases, negative epistatic interactions between mutations fixed independently in the diverging lineages are expected to play a prominent role when brought together in the same individual (known as Dobzhansky–Müller incompatibility) [1].

For fungal species, pre-zygotic barriers may be achieved, for example, through loss of the pheromone and pheromone receptor interaction that usually initiates mating between cells with compatible mating types. As to post-zygotic barriers, these can be achieved through several different mechanisms. First, mating type-specific transcription factors that determine compatibility after cell fusion could be incompatible between mating partners of divergent lineages, thereby preventing the formation of an active heterodimer in the zygote that is critical for sexual reproduction to proceed (reviewed in [2, 3]). Second, the segregation of co-adapted

nuclear and cytoplasmic elements (e.g. mitochondria) could affect the fitness or even viability of the zygotes, and subsequent sexual development [4-7]. Third, the accumulation of genetic differences (including both sequence divergence and chromosomal rearrangements) between incipient species or diverging populations could compromise meiosis after hybridization, likely by disrupting homologous recombination and faithful chromosomal segregation, and thus, producing progeny with imbalanced genetic material that are either inviable or sterile [8-11]. Indeed, it has been recently shown that in interspecific crosses of the yeasts *Saccharomyces cerevisiae* and *Saccharomyces paradoxus*, the majority of the observed hybrid inviability could be attributed to meiosis I chromosomal nondisjunction, whose frequency could be significantly reduced by partially impairing the activity of genes that prevent homologous recombination between nonidentical sequences, suggesting that the chromosomal mis-segregation during meiosis is likely due to the presence of sequence divergence [12].

It has recently been estimated that there may be as many as 2.2 to 3.8 million fungal species in the world [13-15]. Fungi can influence the recycling of nutrients in diverse ecosystems as free-living organisms [16], or impact the health of many plants and animals positively as commensals or negatively as pathogens [17]. The definition of species concept affects how studies defining species are carried out and therefore has a great influence on diversity studies [18]. While a vast diversity in the fungal kingdom is appreciated, the definition of species and the identification of species boundaries in fungi are not always straightforward. Different approaches of defining species can result in detecting different entities. Species defined based on phenotypic or morphological variation may not necessarily be the same as species defined based on reproductive isolation, otherwise known as the biological species concept [19]. While many fungal species can be defined under the biological species concept, with robust sexual reproduction within but not between closely related species, frequently the biological species definition cannot be readily applied. Obviously, it cannot be applied to define species without a

known sexual cycle. Difficulties also arise because closely related fungal species may undergo hybridization and often produce viable hybrid progeny (albeit with highly reduced spore viability rates) that can propagate asexually through mitosis or, in rare cases, can even engage in sexual reproduction through backcrossing with either or both of the parental species [10, 11, 20, 21]. An alternative approach that avoids some of these pitfalls is to use DNA sequences to differentiate populations and define species. Thus, many fungal species are now defined through phylogenetic-based approaches [22], consisting on the analysis of divergence between lineages using selected DNA sequences that can distinguish a broad range of fungi (DNA-barcoding [23]), or by comparing whole genome sequences and presence of other genomic changes, including chromosomal rearrangements. One advantage of applying such criteria is that these genomic changes tend to occur and can be recognized before divergence has accumulated in other aspects of fungal biology, such as mating behavior or morphology. However, because slight differences can be found among virtually any group of fungi, the phylogenetic species concept alone can sometimes encourage extreme division of species into ever-smaller groups. Therefore, when possible, a combination of approaches allows the most accurate assessment of the dynamics underlying speciation.

The *Cryptococcus neoformans*/*Cryptococcus gattii* species complex is a group of closely related basidiomycete yeasts that has been used as model organisms to study fungal pathogenesis and for antifungal drug discovery [24, 25]. There are currently seven defined species within this species complex [26]. They mostly infect immunocompromised hosts, with infection initiating in the lungs. If not treated, it will disseminate to other organs, especially the brain, where it causes meningoencephalitis. Pathogenic *Cryptococcus* species collectively cause over 200,000 infections annually, making them one of the leading human fungal pathogens [27]. While many factors have been identified in the *C. neoformans*/*C. gattii* complex that contribute to virulence, it is still not fully understood how their pathogenesis evolved. Interestingly, many of the species

that are closely related to the pathogenic *Cryptococcus* species complex, such as *Cryptococcus amyloletus*, *Cryptococcus depauperatus*, and *Cryptococcus luteus* are not known to cause disease in plants or animals [28-31], and are instead regarded as saprobes or mycoparasites [32]. Of these sister species, *C. amyloletus* is most closely related to the pathogenic *C. neoformans/C. gattii* complex. Phylogenetic analyses as well as recent extensive genomic comparison studies have shown that substantial sequence divergence and chromosomal rearrangements have accumulated since the separation of the *C. neoformans/C. gattii* and *C. amyloletus* lineages [28-31]. One such rearrangement is associated with the transition from an ancestral tetrapolar breeding system (wherein two independent and unlinked mating-type loci determine pre- and post-fertilization compatibility (reviewed in [3]), to the extant bipolar breeding system of the pathogenic species that have linked mating-type (*MAT*) loci [33]. Such a transition determining reproductive compatibility is beneficial in inbreeding mating systems [34, 35] and was likely selected for in these fungi [36]. Genomic rearrangements can thus serve as direct targets for natural selection to act upon and are potentially of functional significance for virulence and the emergence of *Cryptococcus neoformans/C. gattii* species as major human pathogens.

A total of four isolates of the *C. amyloletus* species complex are available for study. These include the type strain of *Tsuchiyaea wingfieldii* CBS7118, and the recently isolated strain DSM27421 [37], which together with the two *C. amyloletus* strains form the sister clade to the pathogenic *Cryptococcus* lineage [38]. While isolates are routinely identified using partial ribosomal gene sequences, the type strains of *C. amyloletus* and *T. wingfieldii* were also studied using additional protein coding genes, but the two species were considered taxonomic synonyms because of the relatively low divergence of the analyzed sequences [38]. However, no whole genome sequencing and karyotypic data were available for either of the two isolates of *T. wingfieldii*, and it remained unclear whether they are mating compatible with other strains. Thus, the taxonomic status of strains CBS7118 and DSM27421 still needs to be fully established.

Given their close phylogenetic relationship with *C. amyloletus*, they could be closely related diverging lineages of the same species, or they could represent distinct species. Furthermore, the addition of two more isolates in the sister clade of the pathogenic *Cryptococcus* lineage would also allow comparative genomic analyses to identify with higher accuracy the genetic divergence and chromosomal rearrangements segregating within and between the pathogenic and non-pathogenic sister lineages, and thus provide further insights into the evolution of pathogenesis in the *C. neoformans/C. gattii* species complex.

In this study, we reexamine the species boundaries of the four available isolates within the *C. amyloletus* complex using phylogenetic, genetic, and genomic approaches, and revealed three genetically isolated species. We show that while strain CBS7118 does not mate with any of the strains tested, strain DSM27421 could undergo sexual reproduction with *C. amyloletus* to produce spores. However, by analyzing the F1 progeny of crosses between DSM27421 and *C. amyloletus* strains, we show definitively the presence of post-zygotic reproductive barriers between the two lineages, similar to those observed among the sister species in the *C. neoformans/C. gattii* species complex. These results confirm the genetic separation of *C. amyloletus* and strain DSM27421, for which the name *Cryptococcus floricola* is proposed. Additionally, we sequenced and assembled the complete genomes of the isolates CBS7118 and DSM27421. Genomic comparison of the two newly assembled genomes, together with those from *C. amyloletus*, identified both species- and clade-specific sequence polymorphisms and chromosomal rearrangements. We discuss our findings in the context of speciation among closely related fungal lineages, including the establishment of reproductive isolation through both sequence divergence and chromosomal rearrangements.



## Results

### Whole-genome phylogenetic analysis suggests three distinct species within the *C.*

#### *amylolentus* complex

Mittelbach *et al.* first isolated strain DSM27421 in 2012 from flower nectar collected in Tenerife, Canary Islands, Spain [37]. This isolate was originally identified as *T. wingfieldii* (a taxonomic synonym of *C. amylolentus*) based on sequence similarity of the D1/D2 domain of large subunit (26/28S) ribosomal RNA gene (LSU), which is a commonly used barcode in yeast identification. Because the LSU nucleotide sequences of the type strains of *C. amylolentus* and *T. wingfieldii* differ only in one nucleotide and one gap, additional genes were employed to try to resolve the phylogenetic relationship of the four strains in the *C. amylolentus* species complex. First, a maximum likelihood gene tree was inferred using a concatenated alignment of the internal transcribed spacer region (ITS1-5.8S-ITS2) and partial sequences of the *RPB1* and *TEF1* genes (Fig 1A). In this tree, the group comprised by strains *T. wingfieldii* DSM27421, *T. wingfieldii* CBS7118, *C. amylolentus* CBS6039, and *C. amylolentus* CBS6273 branched out separately from the other members of the *Cryptococcus* genus in a subtree that received 100% bootstrap support. However, the delimitation between *T. wingfieldii* and *C. amylolentus* remained unclear as the placement of DSM27421 and CBS7118 in a separate branch from the two *C. amylolentus* strains was weakly supported (Fig 1A). The maximum pairwise distance between these four strains was 0.012 substitutions per site, which is less than, for example, the distance between the pathogenic species *C. neoformans* and *C. deneoformans* (0.028 substitutions per site).

To obtain a finer resolution of the *C. amylolentus* species complex, we generated Illumina paired-end sequencing data and draft genome assemblies for *T. wingfieldii* strains CBS7118 and DSM27421. A maximum likelihood phylogeny inferred from the alignment of

4,896 single-copy genes shared among the studied taxa and an outgroup species, *C. depauperatus* (CBS7855) revealed two robustly supported branches: one containing *C. amylolentus* strains CBS6039 and CBS6273 and the other containing the two *T. wingfieldii* strains DSM27421 and CBS7118 (Fig 1B). Because bootstrap values on concatenated datasets can be misleading, we also bootstrapped well-supported single copy gene trees (i.e. with > 50% of 1000 bootstrap replicates at all nodes) and used them to infer both the gene support frequency (GSF) that indicates the percentage of individual gene trees that contain a given bipartition, and the internode certainty (IC) that quantifies the certainty of a bipartition [39, 40]. In this analysis, 92% of the well supported gene trees recovered the *C. amylolentus* clade, but the clade containing *T. wingfieldii* DSM27421 and CBS7118 strains was only recovered in 58% of the supported trees. Therefore, while gene trees for the two *C. amylolentus* isolates show strong support for a single clade, as expected for a single species, there is not as strong evidence supporting a single clade for the two *T. wingfieldii* strains, raising the possibility that they may represent separate species. In line with this, while the two *C. amylolentus* isolates shared 99.6% sequence identity at the whole genome level, the isolates CBS7118 and DSM27421 were clearly more divergent from each other, sharing only 94.4% gene identity between them, and each only sharing 93.5% identity with *C. amylolentus* (Fig 1C). Together, our data suggest that there are three phylogenetically distinct species present in this complex, with CBS6039 and CBS6273 representing *C. amylolentus*, CBS7118 representing the formerly described *T. wingfieldii* that we rename as the new taxonomic combination *Cryptococcus wingfieldii*, and DSM27421 representing a third, yet undescribed species that we named *Cryptococcus floricola*.

However, such sequence divergence does not necessarily ensure that these isolates represent fully established, reproductively isolated species. Thus, we further analyzed several aspects of these strains, including 1) their ability to undergo inter-species hybridization and when hybridization does occur, the consequences of meiosis on the viability and genetic composition

of any hybrid progeny; 2) the presence of chromosomal rearrangements, and together with sequence divergence, their collective effects on meiosis and chromosomal segregation; and 3) whether their phenotypic characteristics and physiological profiles are consistent with them belonging to distinct species.

### **Post-zygotic, but not pre-zygotic reproductive barriers, established between DSM27421 and *C. amylolentus***

In basidiomycete yeasts, mating compatibility is determined at two different levels. First, cells must undergo reciprocal exchange of mating pheromones that are recognized by specific receptors, both encoded by the pheromone/receptor (*P/R*) locus. After cell fusion, distinct homeodomain transcription factors, encoded by the *HD* locus of each mating partner, interact to initiate sexual development. For successful mating and completion of the sexual cycle, both compatibility factors must be heterozygous in the product of mating (e.g. zygote or dikaryon) [3]. Therefore, pre-zygotic barriers are absent when a cross between two strains results in normal development of sexual structures. For *Cryptococcus*, complex sexual development includes hyphae with complete clamp connections, basidia and basidiospores. To assess this, CBS7118 and DSM27421 were crossed with each other and with all of the available strains of *C. amylolentus* (i.e. the two parental isolates, CBS6039 and CBS6273, and both F1 and F2 progeny derived from a cross between these two strains; Table 1). All combinations of A1 or A2 (*P/R* locus) and B1 or B2 (*HD* locus) *MAT* alleles were accounted for in the *C. amylolentus* strains employed for crosses with CBS7118 and DSM27421. Following incubation for two weeks at room temperature in the dark, no mating was observed in any of the crosses that involved CBS7118, and hence this strain is either sterile or incompatible with all of the strains tested (Table 1). On the other hand, strain DSM27421 successfully mated with the *C. amylolentus* tester strains having A1B1 or A1B2 mating type, but not those with A2B1 or A2B2 mating type,

indicating that this strain has an A2B3 mating genotype. Sparse aerial hyphae were observed around the periphery of the mating colonies and projecting away from the agar surface. The aerial hyphae displayed prominent fused clamp connections indicative of dikaryon formation (Figs 2A and 2C), and the distal ends of the hyphae differentiated into round, unicellular basidia (Figs 2B and 2D). Four long chains of spores budded from the apical surface of the basidium (Fig 2D), with the four most distal spores remaining firmly attached to each other. The sexual structures observed in heterospecific crosses (i.e. crosses between strains of different species; e.g. DSM27421 x CBS6039) were morphologically indistinguishable from those previously reported for the *C. amyloletus* CBS6039 x CBS6273 conspecific cross [29], although visually much less abundant.

For full interfertility, spore formation and viability should occur at similar rates in both conspecific and heterospecific crosses. Diminished spore formation or viability in a heterospecific cross is indicative of a post-zygotic barrier. It has been shown that the conspecific cross between *C. amyloletus* CBS6039 and CBS6273 has an average spore germination rate of 50%, increasing up to 64% in spores dissected from F1 intercrosses [33]. In contrast, heterospecific crosses between DSM27421 and *C. amyloletus* had an overall germination rate of only 10% (n=846 spores from 28 basidia of 6 independent crosses). The average germination rate per basidium for each cross ranged from < 1% to 22% (Table 2). Interestingly, spores from the cross between DSM27421 and CBS6039 had the lowest germination rate, with only one spore germinating out of 153 spores dissected from 6 basidia. At the other end of the spectrum, the cross between DSM27421 and *C. amyloletus* progeny SSC125 produced 3 basidia with at least a 20% germination rate, although a fourth basidium from this cross yielded no viable spores (Table 2). Taken together, the lower number of spores and the high percentage of inviable progeny indicate that *C. amyloletus* and DSM27421 exhibit considerable intrinsic post-zygotic isolation, further supporting the assignment of DSM27421 to a different biological species.

## **Comparison of chromosome-level genome assemblies reveals significant differences between the three species**

Post-zygotic reproductive isolation between *C. amyloletus* and DSM27421 could be attributed to genetic incompatibilities. For example, large chromosomal rearrangements could impose a significant barrier through the generation of unbalanced progeny. To investigate this, chromosome-level assemblies for DSM27421 and CBS7118 were generated using Pacific Biosciences (PacBio) and Oxford Nanopore (ONT) long-read sequencing platforms, and subsequently compared to the available *C. amyloletus* CBS6039 reference assembly (GenBank: GCA\_001720205). For each strain, we obtained more than 150× read depth from various PacBio and ONT sequencing runs and used different assembly strategies with Canu to account for the variable read length and error rate of the generated data (S1 Table). Of the resulting assemblies, those with a lower number of contigs and a higher base level accuracy were selected for further analysis after multiple iterations of Illumina-read-based error correction using Pilon (see Materials and methods for details).

The final genome assemblies of DSM27421 and CBS7118 are approximately 21.7 and 20.8 Mb in size and consist of 15 and 14 nuclear contigs, respectively, plus the mitochondrial genome (S1 Table). Fluorescence-Activated Cell Sorting (FACS) analyses were consistent with the two strains being haploid (S1 File), and contour-clamped homogeneous electric field (CHEF) electrophoresis confirmed that DSM27421 and CBS7118 have 14 chromosomes each, as does *C. amyloletus* (S1 Fig) [33]. This indicates that all but one of the chromosomes in both strains were assembled into single contigs. Indeed, only chromosome 7 of DSM27421 was found to be fragmented into two contigs (7q and 7p; Fig 3A and S2C Fig), both representing arms of the same chromosome broken at the centromere. This was confirmed from chromoblot hybridization analyses using probes specific to the opposite ends of each contig (S1D and S1E Figs), and by

*in-silico* detection of regions highly enriched with Long Terminal Repeat (LTR) retrotransposons (Fig 3) at one end of each of the contigs. Such LTR-rich regions in *C. amyloletus* and *C. neoformans* were previously shown to be associated with centromeric regions [33, 41, 42]. Finally, we annotated protein coding genes, tRNA and rRNA genes, identified transposable elements and telomeric repeats, and predicted centromeric regions (see Material and Methods; Fig 3, S2 Fig, S1, S2 and S3 Tables).

Although sharing the same number of chromosomes, whole-genome comparisons revealed several structural differences between *C. amyloletus* and the two other strains, CBS7118 and DSM27421. These differences include: (i) several small-scale intrachromosomal inversions mostly occurring at subtelomeric regions (S3 Fig) and at the *P/R MAT* locus (see below); (ii) a higher number of transposable elements (TE) (S2 Table) mainly associated with longer centromeric regions in DSM27421 and CBS7118 compared to *C. amyloletus* (S4 Fig and S3 Table); and importantly (iii), a reciprocal translocation involving chromosomes 3 and 12 that distinguishes *C. amyloletus* from the other two strains. Interestingly, the break point of the translocation is located within the centromeres, suggesting it was likely result of inter-centromeric ectopic recombination mediated by common transposable elements present within the centromeres (Fig 3 and S1 Fig).

Of the four analyzed strains, only CBS7118 failed to demonstrate sexual reproduction. We therefore asked whether this strain still retains intact *MAT* loci or if it had lost some *MAT* genes as part of gross deletions or chromosomal rearrangements. It has been shown previously that the *P/R* and *HD MAT* loci in *C. amyloletus* are located on different chromosomes [33]. BLAST searches using the *C. amyloletus* pheromone receptor (*STE3*) and *HD1/HD2* genes as query confirmed that the same holds true in both DSM27421 and CBS7118 – the *P/R* locus is located on chromosomes 11 and 9, and the *HD* locus is found on chromosomes 10 and 11, respectively (Fig 3; S2 and S3 Figs). In addition to other genes, the *P/R* locus of each strain

contains a single mating pheromone receptor gene (*STE3*) and three-to-four identical putative pheromone precursors genes with a C-terminal CAAX domain that is characteristic of fungal mating pheromones [3, 43], all apparently intact (Fig 4A and S5 Fig). Further comparison of pheromone precursor proteins found that CBS7118 and *C. amyloletus* CBS6039 shared identical proteins as did DSM27421 and *C. amyloletus* CBS6273 (S5B Fig). Similarly, phylogenetic analyses revealed that the *STE3* alleles from *C. amyloletus* CBS6039 and CBS7118 group together with A2/a alleles from more distantly related species, whereas the *STE3* alleles of *C. amyloletus* CBS6273 and DSM27421 clustered with other A1/α alleles (S5A Fig). Such trans-specific polymorphism is typical of basidiomycete pheromone receptor genes and is expected for genes ancestrally recruited to the *MAT* locus and maintained across speciation by balancing selection [3, 36, 44-47].

In basidiomycete species, the *P/R* locus usually exhibits synteny breaks and sequence divergence between opposite mating types, while synteny is often more conserved across species when *P/R* regions of the same mating type are compared [48-51]. This is presumably associated with suppression of recombination between mating types at the *P/R* locus. Accordingly, while extensive gene shuffling has occurred between the *P/R* loci of *C. amyloletus* CBS6039 and CBS6273 (Fig 4A), the extant rearrangements between CBS6039 and CBS7118 can be simply explained by two successive inversions in CBS7118: the first, spanning the chromosomal segment between the *MYO2/05748* and *RPL39* genes, and the second, involving the region between the *MYO2/05748* and *LPD1* genes (S6A Fig). The rearrangement of the *P/R* locus between CBS6273 and DSM27421 is even simpler. There is a single large inversion that spans from the left pair of pheromone genes all the way to the right pair of pheromone genes (Fig 4A and S6B Fig). Eighteen genes are predicted within this inverted region in DSM27421, all of which have corresponding orthologs in CBS6273. In contrast, two genes of unknown function in CBS6273 that lie between *STE3* and *STE12* are absent in the other three strains. Interestingly, the

breakpoints of these inversions seem to be associated with the presence of identical and divergently oriented pheromone genes (or their remnants) (Fig 4A). It is therefore likely that these genes might have acted as inverted repeats mediating the formation of intrachromosomal inversion loops, thereby facilitating the close apposition of the respective regions for recombination to occur. We defined the boundaries of the *P/R* locus in DSM27421 and CBS7118 to the same regions previously assigned in *C. amyloletus* (highlighted with yellow background in Fig 4A) [33]. It should be noted, however, that in DSM27421 the *P/R* locus may extend further to the left flank given the presence of yet another inversion specific to this species. The subsequent characterization of these inversion breakpoints at the nucleotide level revealed two genes (*06914* and *06935*) that share 98% sequence identity and are inverted relative to one another (Fig 4A), possibly facilitating this rearrangement.

In contrast to the *P/R* locus, the *HD* locus is smaller, and its gene content is largely conserved among the four strains, with divergence primarily occurring at the nucleotide level (Fig 4B). An exception is the presence of one predicted gene that lies upstream of *CAP1* in the two *C. amyloletus* strains, which is absent in CBS7118. Although this additional gene is also missing from DSM27421, this strain also has extensive sequence added to the left of the genes *06787* and *06789* (CBS6039 nomenclature). While this could indicate an expansion of the *HD MAT* locus in DSM27421, determining its precise length will require the analysis of additional isolates as they become available. Therefore, in considering the present data, the *HD* locus most likely includes only the *SXII* (*HD1*) and *SXI2* (*HD2*) genes, or encompasses a few more genes on either side of the *SXII/SXI2* gene core, as previously proposed for *C. amyloletus* [33] (Fig 4B). Among these, the Sxi1 and Sxi2 protein sequences varied considerably between the four strains, consistent with them playing critical roles in mating type determination. For instance, the Sxi1 protein of *C. amyloletus* CBS6039 only shares 91.8%, 88.7%, and 86.5% sequence similarity with the Sxi1 proteins of CBS6273, CBS7118 and DSM27421, respectively.



Taken together, we obtained complete genome assemblies for DSM27421 and CBS7118 and found substantial genomic differences between these two strains and *C. amyloletus*, which is consistent with the hypothesis that DSM27421 and CBS7118 are distinct species. In addition, we found no clear evidence suggesting that CBS7118 was sterile as a result of mutations in the mating type-determining genes.

### **Genotypic and phenotypic analysis of the progeny from the crosses between *C. amyloletus* and *C. floricola* DSM27421**

A sample of progeny derived from two heterospecific crosses of DSM27421 and *C. amyloletus* were characterized further. While many crosses had low rate of spore production, we specifically collected progeny from basidia 2 and 3 of the DSM27421 x SSC120 cross that had relatively high germination rates and from basidium 1 of the DSM27421 x SSC129 cross (Fig. 5A; Tables 2 and 3). In all, 39 progeny were collected for analysis. The *TEF1* gene, encoding the translation elongation factor 1 $\alpha$ , was initially used as a marker in a PCR-RFLP assay to identify different genotypes among the progeny of the same basidium (Table 3). The progeny from basidium 2 of the DSM27421 x SSC120 cross had inherited *TEF1* from only DSM27421, and the progeny from basidium 1 of the DSM27421 x SSC129 cross had inherited *TEF1* from only SSC129. However, in the progeny from basidium 3 of the DSM27421 x SSC120 cross, a mix of parental *TEF1* alleles were found, indicating that at least two genotypes were represented in this set.

To further genotype and follow the mating type of each of the meiotic progeny, crosses of each were performed with the tester strains listed in Table 1 (results shown in Table 3). Sixteen of the progeny were sterile. All of the progeny from DSM27421 x SSC129 mated as mating type A1B1. The progeny from basidium 2 of DSM27421 x SSC120 either mated as A2B3 or were sterile, indicating that at least two genotypes were present in that progeny set. Similarly, the

progeny from basidium 3 of DSM27421 x SSC120 either mated as A1B3 or were sterile, except one isolate (ARP56) that mated as A2B2. Thus, at least three genotypes (out of four possible) were represented in this other progeny set. We identified the *MAT* alleles of the sterile progeny, using PCR-RFLP for the *STE3* and *HD* mating type-specific genes. The sterile progeny from basidium 3 of DSM27421 x SSC120 all had an A1B2 genotype, and in basidium 2 of the same cross the sterile progeny were all A2B2.

The presence of multiple genotypes in the haploid progeny from two different basidia of a heterospecific cross is suggestive of random assortment during meiosis. However, the nature and frequency of meiotic recombination remained elusive. It was recently shown that the frequency of meiotic recombination in conspecific crosses of *C. amyloletus* is on average 1.59 crossovers/Mb and all chromosomes have at least one crossover event [33]. Because sequence divergence between recombining chromosomes of different species may decrease the rate of meiotic recombination, for instance due to interference by the mismatch repair machinery that prevents recombination between diverged sequences [52-55], we sought to compare the findings in *C. amyloletus* with the meiotic recombination patterns in heterospecific crosses of *C. amyloletus* x DSM27421. To this end, paired-end Illumina sequencing was utilized to sequence the genomes of the *C. amyloletus* parents of the progeny collected, as well as a subset of the DSM27421 x *C. amyloletus* progeny. Specifically, we sequenced (i) two F<sub>1</sub> progeny, SSA790 and SSB821, that were derived from the CBS6039 x CBS6273 conspecific cross; (ii) the F<sub>2</sub> progeny recovered from cross SSA790 x SSB821 that could successfully mate with DSM27421 (i.e. SSC103, SSC120, SSC125, SSC129, SSC130; Fig 5A and Table 1); and (iii) 20 of the progeny derived from the heterospecific crosses of DSM27421 x SSC120 and DSM27421 x SSC129 (Tables 1 and 3).

To examine recombination genome-wide and at high spatial resolution, we mapped the Illumina reads to either *C. amyloletus* CBS6039 or DSM27421 reference assemblies and

generated plots for the distribution of single-nucleotide polymorphisms (SNPs) along the length of each chromosome for each of the datasets. In this approach, crossovers along the chromosomes can be scored as transitions between haplotype segments of the two parental strains (Fig 5B). As expected, at least one crossover was generally present on each chromosome of the *C. amyloletus* F<sub>1</sub> progeny (SSA790 and SSB821), with a range of 1 to 6. Additional crossovers were observed in the F<sub>2</sub> progeny, and in six instances no measurable recombination was detected (tracks enclosed by red lines in Fig 5B). It should be noted, however, that crossovers between homozygous regions of the two interbred F<sub>1</sub> strains (SSA790 and SSB821) are undetectable with this approach, resulting in lower estimates. Indeed, the crossover frequency estimated for the *C. amyloletus* F<sub>2</sub> progeny was 1.32 crossovers/Mb on average, in comparison with 1.67 crossovers/Mb of the F<sub>1</sub> progeny. Hence, when considering both the *C. amyloletus* F<sub>1</sub> and F<sub>2</sub> progeny sets, the average of 1.42 crossovers/Mb across all chromosomes should be regarded as the very lowest estimate.

For the *C. amyloletus* x SSC120 heterospecific cross, SNP distribution revealed only five different genotypes (out of 8 possible) among the 17-sequenced progeny derived from two different basidia (i.e. two independent meiosis). Three genotypes were recovered from basidium 3 and only two from basidium 2 (Fig 5C). The missing genotypes in each case most likely represent inviable meiotic products rather than biased sampling, as all spores that germinated were analyzed (Tables 2 and 3). Furthermore, spores with similar genotypes in each basidium are mitotic clones that resulted from multiple rounds of mitosis to generate chains of spores following a single meiotic event [33, 56]. Among the progeny representing unique genotypes, no recombination was detected in eight of the chromosomes (in a total of 14 cases; tracks enclosed by red lines in Fig 5C). This is observed, for example, in five out of 14 nuclear chromosomes of strain ARP56 (numbered 3, 8, 11, 12 and 13) and in three chromosomes of strain ARP61 (3, 10 and 14) (Fig 5C). Nonetheless, at least one crossover per homeologous chromosome pair was

detected in each of the two meiotic events analyzed. The range of the observed crossovers for the recombinant chromosomes was 1 to 5, and the crossover frequency along each chromosome ranged between 0.46 (chromosome 3) and 1.68 (chromosome 7), with an average of 0.97 crossovers/Mb when all chromosomes are considered. Hence, the crossover frequency in the heterospecific crosses is on average lower than that estimated from *C. amyloletus* conspecific crosses (this study and [33]).

Next, to test whether aneuploidy was present in the progeny set of both conspecific and heterospecific crosses, the relative chromosome copy numbers were determined from read counts. The nuclear chromosomes of all progeny had read depths that, relative to the average total read depth, were about one (or equal to zero after log<sub>2</sub> transformation) (S7 Fig). However, a large segment of doubled read count for chromosome 10 was observed for ARP45 and ARP60, indicating that the cell population of these strains included a significant number of  $n + 1$  aneuploids. A third isolate (ARP44) appears to be otherwise genotypically identical to ARP45 and ARP60 but is euploid. No other signs of aneuploidy were present in the sequenced progeny set (S7 Fig and S1 File).

In addition to visualizing inheritance of nuclear DNA, SNP mapping was also used to examine the inheritance of mitochondrial DNA (mtDNA). In crosses between individuals of different species, or between extensively divergent populations of the same species, uniparental inheritance (UPI) of mitochondria may be altered due to a failure of the mechanisms governing UPI. This may lead to the generation of progeny with mtDNA from the parent that normally does not contribute (known as mtDNA leakage) and facilitate recombination of mitochondrial genomes [57-59]. Of the *C. amyloletus* F<sub>1</sub> progeny, SSA790 (mating type A1B2) inherited its mtDNA from CBS6273 (A2B2), but SSB821 (A2B1) inherited its mtDNA from CBS6039 (A1B1) (Fig 5B). On the other hand, all the *C. amyloletus* F<sub>2</sub> progeny inherited their mtDNA from SSB821 (mating type A2B1), which is a pattern consistent with UPI and in line with

previous studies in *C. amyloletus* [29]. Likewise, the progeny derived from the DSM27421 x *C. amyloletus* heterospecific crosses all inherited mitochondrial DNA from DSM27421 (mating type A2B3) as assessed either by SNP mapping or using a PCR-RFLP assay that is specific for the mitochondrial small subunit (mitSSU) ribosomal RNA gene and can distinguish between the two parental mitochondrial alleles (Table 3). The inheritance of CBS6039 mitochondria by SSB821 presumably represents an error in the execution of uniparental inheritance, which is occasionally observed among a small proportion of the progeny in species with UPI of mitochondria [29, 60, 61]. Therefore, it seems that UPI of mitochondria was not affected in the heterospecific crosses and mitochondria were inherited from the A2 parent during sexual reproduction.

Interestingly, the genomes of the three viable progeny from the DSM27421 x SSC129 cross (ARP68, ARP70, and ARP71) had a different result (Fig 5C). No crossovers were seen at any point in their genomes, and all nuclear DNA was inherited from SSC129. Conversely, all mitochondrial DNA was inherited from DSM27421. This finding indicates that, for these three progeny, karyogamy had not occurred, but the nuclear genome of one parent was combined with the mitochondrial genome of the other, a phenomena known as cytoduction [62].

Species that are products of recent divergence may still display signatures of continued exchange of genes [63-66]. Given that DSM27421 can still produce viable progeny and undergo meiotic recombination with at least one strain of *C. amyloletus*, we searched for genomic signatures that could be suggestive of recent introgression between *C. amyloletus*, DSM27421 and CBS7118. Pairwise divergence was used as a proxy to search for evidence of the incorporation of DNA segments (regions >5 kb) of a given strain into the genome of another strain. In this approach, putative introgressed regions would be detected as genomic tracts with near zero sequence divergence. Each genome was independently used as reference to account for regions that could be missing in some of the strains. In all comparisons, the overall uniformity of

sequence divergence across the genome (S8 Fig) suggests that there has been little, if any, recent introgression between the three lineages, which is consistent with the low spore viability of the tested heterospecific crosses.

**Taxonomic description of *Cryptococcus wingfieldii* and *Cryptococcus floricola* and phenotypic comparison with *C. amylolentus***

Description of *Cryptococcus wingfieldii* (van der Walt, Y. Yamada & N.P. Ferreira) Yurkov, A.R. Passer, M.A. Coelho, R.B. Billmyre, M. Nowrousian, M. Mittelbach, C.A. Cuomo, A.F. Averette, S. Sun & Heitman, comb. nov. (MB 829726).

Basionym: *Sterigmatomyces wingfieldii* Van der Walt, Y. Yamada & N.P. Ferreira, *Antonie van Leeuwenhoek* 53, 138, 1987 (MB 133483).

Holotype: PREM 48490 in the Herbarium for Fungi of the Research Institute for Plant Protection, Pretoria, South Africa

Ex-type cultures: CBS 7118, PYCC 5373, JCM 7368, NRRL Y-17143, DSM 107903.

Physiological characteristics of the species are provided in [67].

The species is known from a single culture isolated from insect frass in South Africa. This yeast was originally described in the genus *Sterigmatomyces* and later transferred to the monotypic genus *Tsuchiyaea* as *Tsuchiyaea wingfieldii* (discussed in [68] and [38]). Phylogenetic analyses suggested its close relationship with *Cryptococcus amylolentus*, so *Tsuchiyaea wingfieldii* was considered to be a synonym of *C. amylolentus* [38, 68]. Results obtained in the present study suggest that *Cryptococcus amylolentus*, *Tsuchiyaea wingfieldii*, and the newly described *Cryptococcus floricola* (see below) represent genetically isolated species.

Description of *Cryptococcus floricola* Yurkov, A.R. Passer, M.A. Coelho, R.B. Billmyre, M. Nowrousian, Mittelbach, C.A. Cuomo, A.F. Averette, S. Sun & Heitman, sp. nov. (MB 829727)

Etymology: The species epithet “floricola” refers to its origin of isolation, flower nectar.

After growth on 5% malt agar plates for 2 weeks at 25°C, the streak culture was mucoid, smooth and cream in color, partially transparent with a glistening surface. Upon aging, the colony turned dull and tan and appears wrinkled. After growth on YM agar plates for 7 d at 25°C, cells were ellipsoidal, fusoidal, elongate to cylindrical (4–10 × 3–5 μm), occurred singly and proliferated by polar budding (Fig 6). Pseudohyphae and true hyphae were not observed after 1 mo on Dalmau plate culture on PDA and CMA at 16–22°C. Chlamyospore-like cells were observed in older (3 to 4 weeks) cultures on YM agar and PDA incubated at 16°C. Ballistospores were not observed.

Sexual reproduction with compatible strains of *Cryptococcus amylolentus* observed on V8 agar (pH 5) after growth for 1 week in darkness at room temperature (22°C to 24°C). Fused clamp cells are present. Aerial hyphae are produced. The tips of the aerial hyphae form 3–6 × 4–6 μm basidia with four parallel spore chains of budding globose (1.6–2 × 2–2.5 μm) basidiospores arising from the apical surface of basidia.

Assimilation of carbon compounds: growth on D-glucose, D-galactose, L-sorbose, D-glucosamine, D-ribose, D-xylose, L-arabinose, D-arabinose, L-rhamnose, sucrose, maltose, trehalose, α-Methyl-D-glucoside, cellobiose, salicin, melibiose, lactose, raffinose, melezitose, soluble starch, glycerol, erythritol, ribitol, xylitol, L-arabinitol, D-glucitol, D-mannitol, myo-inositol, 2-Keto-D-gluconate, D-gluconate, D-gluconate, D-galacturonate, succinate, citrate, ethanol, palatinose, L-malic acid and gentiobiose. No growth occurs on inulin, galactitol, DL-lactate, methanol, quinic acid, D-glucarate, galactaric acid, Tween 40 and Tween 80, and nitrate and nitrite.

Urea hydrolysis and DBB reaction were positive. Growth in 50% and 60% D-glucose was positive. Growth in the presence of 1%, 4%, 5%, 8% and 10% NaCl was positive. Maximum growth temperature: 30°C.

Molecular characteristics (type strain): nucleotide sequences of ITS-LSU (D1/D2 domains) rRNA are deposited in NCBI/EMBL (GenBank) under the accession number HG421442.

Deposits: holotype MoM 837<sup>T</sup> isolated from nectar of *Echium leucophaeum* in Tenerife, Canary Islands, Spain (28° 33.80' N, 16° 17.41' W), preserved in a metabolically inactive state at the German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany. Ex-type cultures are deposited in the German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany (DSM 27421), the CBS yeast collection of the Westerdijk Fungal Biodiversity Institute, Utrecht, The Netherlands (CBS 15421<sup>T</sup>) and the Portuguese Yeast Culture Collection (PYCC), Caparica, Portugal (PYCC 8315<sup>T</sup>).

Notes: The species differs from closely related species *Cryptococcus amylolentus* and *Cryptococcus wingfieldii* in the ability to grow on D-glucosamine,  $\alpha$ -Methyl-D-glucoside,  $\beta$ -Methyl-D-glucoside, citric acid, and some aldaric acids such as D-malic, D-saccharic, D-tartaric, L-malic, L-tartaric, and mucic acids (Fig 6). Additional isolates of the three species can be identified using partial nucleotide sequences of *TEF1* and *RPB1* genes, which showed 97-98% similarity in pairwise comparisons.



## Discussion

Speciation requires the establishment of reproductive isolation, and thus, the cessation of gene flow, between genetically diverged lineages. With the advancement of genome sequencing techniques, it has become less challenging to obtain data on whole genome sequences and compare genetic and genomic divergence between closely related species. This is particularly true for eukaryotic microbes like fungi with relatively small and simple genomes. However, it remains less straightforward to obtain the complete view of the reproductive compatibility between diverging lineages that could represent cryptic or nascent species. This could be due to several factors: 1) environmental cues for mating can vary among closely related fungal species and thus the conditions that induce sexual reproduction between different species can be difficult to reconstitute in the laboratory; and 2) even in cases where mating structures are observed between potential different species, it is still important to assess the viability of the mating progeny, which are sometimes difficult to recover, but nevertheless are the true hallmarks of whether reproductive barriers are present.

In our study, we observed 93.5% - 94.4% pairwise genome level sequence similarity among the genomes of *C. amyloletus* CBS6039, *C. wingfieldii* CBS7118, and *C. floricola* DSM27421. These values are slightly higher, but generally comparable to those observed among the species in the *C. neoformans/C. gattii* species complex [69, 70]. This finding is also consistent with the observed chromosomal rearrangements among these species, where although inversions have been identified, they are generally small, simple, and concentrated in the subtelomeric regions, and we identified only one balanced translocation that is shared by isolates CBS7118 and DSM27421. In contrast, multiple chromosomal translocations, as well as large and complex inversion/transposition regions have been identified among species in the *C. neoformans/C. gattii* complex [10, 41]. Interestingly, the breakpoint of the translocation identified in CBS7118 and DSM27421 is located within the centromeric region, suggesting it

might be the result of inter-centromeric ectopic recombination mediated by shared transposable element present in the centromeres, which have been shown in previous studies to play important roles in the genome evolution of *C. neoformans*/*C. gattii* complex as well as their closely related species [33].

Comparison of the *MAT* loci showed that the *P/R* locus region in *C. wingfieldii* CBS7118 and *C. floricola* DSM27421 has each undergone further rearrangements compared to *C. amyloletus* CBS6039 and CBS6273, in regions both within as well as flanking the *MAT* regions. However, the *P/R* alleles of the same mating type in different species are still more similar to each other than those of the opposite mating types in the same species, which is consistent with observations in other basidiomycetes species [48-51, 71].

Our analyses also suggest that while the same transposable elements are enriched in the centromeric regions of all four isolates analyzed, *C. wingfieldii* CBS7118 and *C. floricola* DSM27421 have additional copies of these transposable elements, leading to predicted centromeres that are larger than those of *C. amyloletus* isolates (S4 Fig). The primary function of the centromere is to generate a functional kinetochore for faithful chromosome segregation and as a few studies suggest [72, 73] differences in centromere structure may mediate or reinforce hybrid incompatibility, thus facilitating speciation. In *C. neoformans*, centromeres can undergo expansion/contraction likely through recombination mechanisms involving transposable elements [42]. Thus, it is possible that *C. wingfieldii* and *C. floricola* have evolved to sustain larger centromeres. Whether or not the relatively larger centromeres of *C. floricola* have expanded centromeric heterochromatin relative to the smaller *C. amyloletus* centromeres that could possibly lead to unequal tension on centromeres remains unknown but will be an interesting topic to explore in future studies.

Taken together, our genomic comparison analyses collectively suggest that isolates CBS7118 and DSM27421 represent distinct species, *C. wingfieldii* and *C. floricola*, respectively,

that along with *C. amyloletus* form the sister clade to the pathogenic *C. neoformans/C. gattii* species complex.

In our study, *C. wingfieldii* (CBS7118) did not mate with *C. floricola*, or with any of the *C. amyloletus* strains. As mentioned earlier, this could be due to our inability to recapitulate the ideal conditions for *C. wingfieldii* to undergo sexual reproduction. Alternatively, it could be that *C. wingfieldii* has already undergone enough divergence that pre-zygotic reproductive barriers have already been established between *C. wingfieldii* and the other two species, *C. amyloletus* and *C. floricola*. If that is the case, this is not yet reflected at the level of the pheromone genes, which are identical between *C. wingfieldii* and *C. amyloletus* CBS6039 (S5 Fig). Another possibility could be that cell-cell fusion still occurs between *C. wingfieldii* and *C. amyloletus*, but the down stream sexual development governed by the *HD* genes is compromised, and thus, no clear signs of mating (e.g. filamentation) could be detected. On the other hand, we observed successful sexual reproduction between *C. floricola* DSM27421 and the *C. amyloletus* tester strains with varying genetic backgrounds, with mating structures highly similar to those observed in crosses between *C. amyloletus* isolates [29]. Importantly, we were able to successfully recover mating progeny from these crosses. However, when the spores were dissected and analyzed, the progeny from crosses between *C. floricola* and *C. amyloletus* showed a highly reduced germination rate, indicating that most of the sexual progeny are inviable. A similar the low spore viability is seen in crosses between sister-species of the *C. neoformans/C. gattii* complex [10, 11]. This suggests that post-zygotic instead of pre-zygotic reproductive isolation has been established between *C. floricola* and *C. amyloletus*.

Post-zygotic isolation in this case could have resulted from compromised meiosis due to the elevated levels of genetic divergence, as well as the presence of chromosomal rearrangements such as translocations and inversions. Together, these differences could have reduced the frequency of crossing-over and resulted in chromosome mis-segregation, thus

leading to unbalanced meiotic products. Consistent with this hypothesis, when progeny from crosses between *C. amyloletus* and *C. floricola* were analyzed, we found that crossovers occurred at a lower frequency than in crosses between *C. amyloletus* isolates [33]. Specifically, we observed no signs of crossover in several chromosomes, which likely resulted from chromatids that did not participate in crossover during meiosis I, similar to the observations reported in crosses within the *C. neoformans/C. gattii* species complex [74, 75]. Alternatively, the reduced spore viability could be due to the presence of mechanisms such as Dobzhansky-Muller incompatibility that have evolved between the diverging lineages. It should be noted that the reduction of spore viability from crosses between *C. amyloletus* and *C. floricola* appears to be less severe than in the case of interspecific crosses in the *C. neoformans/C. gattii* complex, suggesting that speciation among *C. amyloletus* and *C. floricola* likely occurred more recently.

The isolates representing *C. amyloletus*, *C. floricola*, and *C. wingfieldii* have been isolated from different geographic areas. It is possible that geographic isolation further enhances genetic isolation among these species [19]. A combination of genetic incompatibility, and differences in distribution range and dispersal vectors could explain the lack of mating between these species. It is important to mention that all isolates of the *C. amyloletus* species complex were found in habitats associated with insects, including insect frass and floral nectar. Although *C. amyloletus* and *C. wingfieldii* have been found so far only in South Africa, their dispersal may depend on different vectors. In contrast to the other two species, *C. floricola* was isolated much farther away, on the Canary Islands (Macronesia), and may thus have a different range. The three species show distinctive phenotypic characteristics, and principle component analyses of the physiological profiles of the three species classified them into three distinct groups (Fig 6A). For example, it appears that *C. floricola* may have a unique ability to grow under high glucose conditions, which could have been selected for as *C. floricola* was originally found to be associated with flower nectar. It would be difficult to dissect whether the observed divergent

physiological features in these species are the causes or the consequences of their ecological separation leading to or facilitating genetic isolation and eventually the establishment of reproductive isolation between closely related nascent species.

The species *C. amyloletus*, *C. floricola*, and *C. wingfieldii* form a tight clade that is closely related to the major human fungal pathogen *C. neoformans/C. gattii* complex, which collectively cause over 200,000 cases annually [27]. Understanding and dissecting the genetic basis and evolutionary events that led the species in this complex to become successful human pathogens requires an analysis and comparison of the genomes of not only these pathogens, but also their non-pathogenic close relatives. It has been shown that *C. amyloletus* is not virulent in a murine model [28], although it might be virulent in insect models [76]. Both *C. floricola* and *C. wingfieldii* show no growth at elevated temperatures mimicking the body temperatures of mammals, suggesting they are also likely avirulent in humans. Thus, these three species provide a unique opportunity to gain further insights into the evolution of the species in the pathogenic *C. neoformans/C. gattii* species complex.

## **Materials and methods**

### **Mating crosses, spore dissection, and genotyping**

Plates of V8 (pH = 5) mating medium [77] were inoculated with suspensions of yeast cells in water. For each pair-wise cross, 10  $\mu$ L of each strain was pipetted onto the same spot of medium. The plates were then incubated in the dark at 24°C for at least one week and checked periodically thereafter. Mating responses were observed with a compound light microscope and scored as positive if aerial hyphae, basidia, and spores were observed. Spore germination was quantified by using a micromanipulator to transfer individual basidia from mating plates to YPD plates and then to separate individual spores into a defined grid. The plates were incubated at

24°C to allow colonies to form. The proportion of spores that germinated was calculated for each basidium by dividing the number of spores that formed a colony by the number of spores that were plated. The mating type of the viable progeny was determined by crossing them with tester strains (indicated in Table 1). For the sterile progeny, the mating genotype was determined via a PCR-RFLP assay using the primers listed in Table S4 targeting a small region of the *P/R* or *HD* locus, and then digested using a restriction enzyme (BsaHI and Tsp45I, for *P/R* and *HD* loci respectively) that yielded a restriction pattern unique to each *MAT* alleles. A similar approach was used to identify different alleles of the mitochondrial small-subunit rRNA using MseI.

## Microscopy

Inoculation and incubation conditions for microscopy specimens were the same as for the other mating crosses. For the light micrographs, a cross of DSM27421 x CBS6039 on V8 (pH = 5) mating medium was incubated for 39 days. The crosses were viewed with a Zeiss Scope.A1 microscope and photographed with a Zeiss Axiocam 105 Color camera. For the scanning electron micrographs (SEM), a cross of DSM27421 x CBS6039 on V8 (pH = 5) mating medium was incubated for 11 days. SEM was performed at the North Carolina State University Center for Electron Microscopy, Raleigh, NC, USA. Agar blocks approximately 0.5 cm<sup>3</sup> containing hyphae on the edges of mating patches were excised and fixed in 0.1 M sodium cacodylate buffer (pH = 6.8) containing 3% glutaraldehyde at 4°C for several weeks. Before viewing, the agar blocks were rinsed with cold 0.1 M sodium cacodylate buffer (pH = 6.8) three times and then dehydrated in a graded series of ethanol to 100% ethanol. The blocks were critical-point dried with liquid CO<sub>2</sub> (Tousimis Research Corp.) and sputter coated with 50 Å of gold/palladium using a Hummer 6.2 sputter coater (Anatech USA). The samples were viewed at 15 kV with a JSM 5900LV scanning electron microscope (JEOL) and captured with a Digital Scan Generator (JEOL) image acquisition system.

## **DNA extraction and genome sequencing of DSM27421 and CBS7118 strains**

Genomic DNA was extracted using a modified CTAB protocol as previously reported [42]. High molecular weight DNA samples were obtained by spooling out the precipitated DNA using a glass rod instead of using centrifugation. Genomic DNA size and integrity were confirmed by CHEF as previously described [33, 42]. Sequencing of the DSM27421 and CBS7118 genomes was carried out using Illumina, Pacific Biosciences (PacBio), and Oxford Nanopore (ONT) technologies. For Illumina sequencing of the CBS7118 genome, two libraries were constructed with average insert sizes of 187 bases and 1.9 kilobases (jumping library). For the fragment library, 100 ng of genomic DNA was sheared to ~250 bp using a Covaris LE instrument and prepared for sequencing as previously described [78]. The ~2 kb jumping library was prepared using the 2-to-5-kb insert Illumina Mate-pair library prep kit (V2; Illumina) as previously described [79]. These libraries were sequenced by the Broad Institute Genomics Platform on the Illumina HiSeq2000 to generate paired 101 base reads. The genome of DSM27421 was sequenced from a small insert-size library (~350 bp) on a HiSeq 2500 to generate 151 base reads. For PacBio sequencing, large-insert size libraries (15-20 kb) were generated and run on a PacBio RS II or Sequel (2.0 chemistry) systems (S1 Table). PacBio sequencing and Illumina sequencing of DSM27421 were performed at the Sequencing and Genomic Technologies Core Facility of the Duke Center for Genomic and Computational Biology. Nanopore sequencing was performed as per manufacturer's guidelines. Libraries were prepared with the SQK-LSK108 1D ligation Sequencing kit and run for 48 hours in a R9 flow cells (FLO-MN106) using the MinION system. MinION sequencing and live base-calling was controlled using Oxford Nanopore Technologies MinKNOW v.1.10.16 software.

## **Genome assembly and gene prediction of strains DSM27421 and CBS7118**

The initial assembly of CBS7118 was generated from approximately 100X of reads from the fragment library and 50X of reads from the jumping library using ALLPATHS-LG [80] version R47093. The resulting assembly consisted of 83 scaffolds and 196 contigs. For DSM27421, the initial assembly was generated using SPAdes v3.10 [81], resulting in 573 scaffolds and 636 contigs. Improved assemblies using both PacBio and Nanopore long-read read data were generated with Canu v1.7 [82] using the default parameters and an estimated genome size of 20 Mb. Because the read length profile of each sequencing run differed considerably (S1 Table), different read length combinations were tested as input for Canu. For DSM27421, the final draft assembly was generated by combining the read data from two PacBio runs (reads above 10 kb from #run1 and all reads from #run2) and the ONT reads above 10 kb (#run1) (see S1 Table). For the CBS7118 strain, combining the ONT and PacBio reads resulted in more fragmented assemblies than using the PacBio data alone, most likely due to a smaller read length and higher noise level of the ONT reads, and thus, only PacBio reads were used (S1 Table). Genome assembly statistics and additional genomic features of each strain are reported in S1 Table. The accuracy of the resulting assemblies was improved by correcting errors using five rounds of Pilon (v1.22) polishing ('--fix all' setting) [83] using the Illumina reads mapped to the respective assembly using BWA-MEM (v0.7.17-r1188) [84]. Gene models were predicted *ab initio* using MAKER v2.31.18 [85] with predicted proteins from *Cryptococcus neoformans* H99 [41] and *Cryptococcus amyloletus* CBS6039 [33] as input.

### **Whole genome pairwise identity and phylogenetic analyses**

To calculate average pairwise identities (Fig 1C), genomes were aligned using NUCmer (v3.22) from the MUMmer package [86] with parameter '-mum'. Alignments were filtered with delta-filter using parameters '-1' to select 1-to-1 alignments allowing for rearrangements and '-l 100' to select a minimal alignment length of 100 bases. A maximum likelihood (ML) phylogram



to confirm the phylogenetic placement *C. amyloletus* species complex within the *Cryptococcus* lineage was inferred from a concatenated alignment of the internal transcribed spacer region (ITS1, 5.8S and ITS2), *RPB1* and *TEF1* genes. Available sequences were obtained from GenBank and additional sequences were determined by Sanger sequencing and deposited in GenBank (S2 File). The individual genes sequences were aligned by MAFFT (v7.245) [87] using the E-INS-i algorithm, and subsequently concatenated. A maximum likelihood phylogram was generated for this dataset in MEGA v6.06 [88] using the following parameters: uniform rates, complete gap deletion, subtree pruning and regrafting level 5, very weak branch swap filter, BioNJ initial tree, and using the Kimura 2-parameter substitution model [89]. Phylogram stability was measured with 1,000 bootstrap replicates.

For finer resolution of the *C. amyloletus* complex, orthologs were identified between *C. wingfieldii* (CBS7118), *C. floricola* (DSM27421), and *C. amyloletus* (CBS6039 and CBS6273), and an outgroup (*C. depauperatus* CBS7855) based on BLASTP pairwise matches with expect < 1e-5 using ORTHOMCL (v1.4). A phylogeny was inferred from 4,896 single copy genes as follows. Individual proteins were aligned using MUSCLE [90], the individual alignments were concatenated, and poorly aligning regions removed with trimAl [91]. This sequence was input to RAxML (v8.2.4; raxmlHPC-PTHREADS-SSE3) [92] and a phylogeny estimated in rapid bootstrapping mode with model PROTCATWAG and 1,000 bootstrap replicates. To estimate the gene level support for the best tree obtained from this analysis, the individual gene trees were inferred from protein alignments using RAxML with the same settings. The subset of gene trees with at least 50% bootstrap support at all nodes were input to RAxML with the best tree to estimate gene support frequency (GSF) and internode certainty (IC) at each node [39, 40] using settings -f b or -f i respectively.

### **Analysis of genomic features and synteny comparison**

Repetitive DNA content, including transposable elements, was analyzed with RepeatMasker (Smit AFA, Hubley R, Green P. RepeatMasker Open-4.0. 2013-2015, <http://www.repeatmasker.org>), using REPBASE v23.09 [93], and with TransposonPSI (<http://transposonpsi.sourceforge.net/>). Centromeres were predicted upon detection of centromere-associated LTR elements previously reported in *C. amyloletus* (Tcen1 to Tcen6) [33] and *C. neoformans* (Tcn1 to Tcn6) [41, 94]. These elements were identified from the RepeatMasker output and confirmed by BLASTN using the *C. neoformans* sequences as query. Most of these elements mapped to the largest ORF-free region in each contig, including one end of contigs 7a and 7b in the DSM27421 assembly that represent arms of the same chromosome. Centromere locations were further refined by mapping onto the DSM27421 and CBS7118 assemblies the position of the centromere flanking genes previously identified in *C. amyloletus* [33], using BLAST analyses. The final centromere length was measured as the intergenic region between the centromere flanking genes (S3 Table) and subsequently compared to the previously reported centromere lengths of *C. neoformans* H99, *C. deuterogattii* R265 and *C. deneoformans* JEC21 [42] (S4 Fig). Statistical tests (Tukey-Kramer HSD test) were performed using JMP Pro 13 (SAS Institute). The GC content was calculated in non-overlapping 5-kb windows using a modified perl script (gcSkew.pl, <https://github.com/Genomics/scripts/blob/master/AssemblyTools/gcSkew.pl>) and plotted as the deviation from the genome average for each contig. Ribosomal RNA genes (18S, 5.8S, 25S, and 5S) and tRNA genes were inferred and annotated using RNAmmer (v1.2) [95] and tRNAscan-SE (v2.0) [96], respectively. Telomeric repeats were identified using the EMBOSS fuzznuc function [97] based on known telomere repeat sequences of *C. amyloletus*. A search pattern of  $2 \times C(3,4)GCTAAC$  was used, allowing for minor variation between the repeats.

Synteny comparison across the genomes of all strains was conducted using megablast (word size: 28) and plotted, together with the other genomic features, using Circos (v0.69-6) [98]

(as shown in Fig 3 and S2A and S2B Figs). Additional whole-genome alignments were conducted with Satsuma (<https://github.com/bioinfologics/satsuma2>) [99] with the default parameters, and the output was sequentially passed to the visualization tools ‘BlockDisplaySatsuma’ and ‘ChromosomePaint’ included in the same package to generate a postscript file. Centromere and other genomic features were superimposed at scale in the final figure (shown in S2C-E Figs) based on their respective genome coordinates. Linear synteny comparisons shown in S3 and S4A Figs were generated with the Python application Easyfig [100] with following settings: -svg -f gene frame -legend both -leg\_name locus\_tag -blast\_col 230 230 230 230 230 230 -blast\_col\_inv 241 209 212 241 209 212 -bo F -width 5000 -ann\_height 450 -blast\_height 300 -f1 T -f2 10000 -min\_length 1000.

### **CHEF analysis and chromoblots**

CHEF gel electrophoresis and chromoblot analyses were carried out as described in a previous study [28]. Probes that hybridized to each chromosome arm were generated by PCR using the primers listed in Table S4.

### **Analysis of mating type regions**

*MAT* regions were identified by BLAST searches against the well-annotated *MAT*-derived proteins from *C. neoformans*, and manually reannotated if necessary. Synteny between *MAT* regions of different strains was based on bidirectional BLAST analyses of the corresponding predicted proteins. The short pheromone precursor genes (or their remnants) were not always found among the predicted genes and were thus identified manually.

### **Read mapping, variant calling and filtering, aneuploidy, and genome-wide recombination**

The genomes of *C. amyloletus* F1 and F2 progeny, and the progeny derived from DSM27421 x *C. amyloletus* heterospecific crosses, were sequenced with Illumina paired-end sequencing on a HiSeq 4000. Read lengths were 100 or 150 bases, depending on the run. Genomic signatures consistent with meiotic recombination and aneuploidy were inferred, respectively, from SNP distribution and read depth obtained for each chromosome. Paired-end reads of *C. amyloletus* F1 and F2 progeny were mapped to the *C. amyloletus* CBS6039 reference genome whereas those resulting from DSM27421 x *C. amyloletus* crosses were mapped to the newly generated DSM27421 assembly. In both cases, reads were mapped using BWA-MEM short read aligner (v0.7.17-r1188) with default settings. SNP discovery, variant evaluation and further refinements were carried out with the Genome Analysis Toolkit (GATK) best practices pipeline [101] (v4.0.1.2), including Picard tools to convert SAM to sorted BAM files, fix read groups (module: ‘AddOrReplaceReadGroups’; SORT\_ORDER=coordinate), and mark duplicates. Variant sites were identified with the HaplotypeCaller from GATK using the haploid mode setting and only high-confident variants that passed a filtration step were retained (‘VariantFiltration’ module used the following criteria:  $DP < 20 \parallel QD < 15.0 \parallel FS > 60.0 \parallel MQ < 55.0 \parallel SOR > 4.0$ ). Finally, filtered variants found in each contig/chromosome were binned into 5-kb windows and parsed into a tab-delimited format to allow visualization in Circos.

Recombination tracts are observed as transitions between haplotype segments from the two parental strains along the chromosomes. Read count data was used to screen for gross aneuploidy of chromosomes. First, reads were counted in 5-kb non-overlapping windows using the module ‘count\_dna’ from the Alfred package (v0.1.7) (<https://github.com/tobiasrausch/alfred>) and using the BAM file obtained after read mapping as input. Then, the resulting read counts were median normalized and log<sub>2</sub>-transformed, and finally parsed into a tab-delimited format and plotted as a heatmap in Circos.

## Divergence plots

To detect regions of introgression (> 5kb), Illumina reads generated for each of the strains were treated with the methods described above for the alignment to each of the reference genome assemblies (*C. amyloletus* CBS6039, DSM27421 or CBS7118). The resulting consensus genotype in the Variant Call Format was converted to FASTQ format by limiting maximum depth to 200 to avoid overrepresented regions. A FASTA file was then generated, in which bases with quality lower than 20 (equivalent to 99% of accuracy) were masked to lowercase and ambiguous bases were subsequently converted to an “N.” Divergence per site ( $k$ , with Jukes–Cantor correction) between pair of strains was estimated in VariScan v.2.0.3 [102] using a non-overlapping sliding-window of 5,000 sites. For easier visualization and interpretation, each datapoint in S8 Fig is the average of itself and two filtered windows on either side. Because highly divergent regions are challenging to align to a reference genome, all divergence estimates should be regarded as minimum estimates.

## FACS analysis

To determine the ploidy of the full set of strains used in this study (S1 Data), isolates were cultured on YPD medium for 2 to 4 days at 25°C and processed for flow cytometry as previously described [103], but without sonication. For each sample, approximately 10,000 cells were analyzed on the FL1 channel on a Becton-Dickinson FACScan at Duke Cancer Institute Flow Cytometry Shared Resource. *C. deneoformans* JEC21 (D $\alpha$ ) and *C. deneoformans* XL143 ( $\alpha$ DD $\alpha$ ) were used as haploid and diploid controls, respectively.

## Physiological tests

Phenotype microarray testing of carbon sources was examined using the BIOLOG MicroStation and YT, FF, and GEN III MicroPlates following the manufacturer's instructions

(BIOLOG Inc., Hayward, CA, USA). Yeasts were incubated on potato dextrose agar (PDA, Difco) at room temperature. Yeast biomass was harvested from PDA, suspended in the inoculation solution IF-B (BIOLOG Inc., Hayward, CA, USA), and the turbidity was adjusted to the transmittance value provided by the manufacturer. MicroPlates were sealed to prevent desiccation and incubated at room temperature and measured after 1, 2, 3, and 4 weeks, with the optical density recorded at 590 and 750 nm. The ability to utilize particular substrates by individual strains was recorded as positive, weak or negative.

Utilization of D-glucosamine, maltose, methyl- $\alpha$ -D-glucoside, melezitose, soluble starch, D-glucitol, galactitol, DL-lactic, succinic, citric, and aldaric acids were performed in 3.5 ml liquid media according to commonly used protocols [104, 105]. Growth tests in the presence of 50% glucose, 60% glucose, 5% NaCl, 8% NaCl, and 10% NaCl was were performed in 3.5 ml liquid media according to commonly used protocols [105].

A total of 156 growth responses were recorded with YT, FF, and GEN III MicroPlates, invariable results were discarded and results from 31 variable tests were visualized with a heatmap and principle component analysis (PCA) using ClustVis web tool [106]. Growth response results were transformed into “1” for positive growth, “0.5” for weak growth, and “0” for negative growth. The following settings were used for the analysis: SVD method with imputation; no transformation; no row scaling; no row centering; constant columns were removed. A heatmap was produced using the following settings: no scaling was applied to rows; both rows and columns are clustered using correlation distance and average linkage. PCA was performed with the following settings: no scaling was applied to rows; SVD with imputation is used to calculate principal components.

### **Data availability statement**

Sequencing reads for *C. floricola* DSM27421 (BioProject PRJNA496466), *C. wingfieldii* CBS7118 (BioProject PRJNA496468), and the progeny of crosses (BioProject PRJNA496469) are available in the NCBI SRA database. The DSM27421 and CBS7118 genome assemblies have been deposited at DDBJ/ENA/GenBank under the accession numbers RRZH00000000 and CP034261 to CP034275, respectively. Other sequence accession numbers are listed in File S2.

**Table 1. Strains used in this study.**

Strain	Organism/species	Mating type	Source of strain
CBS7118	<i>Cryptococcus wingfieldii</i>	-	frass from scolytid beetles in South Africa
<b>DSM27421</b>	<i>Cryptococcus floricola</i>	A2B3	nectar from the flower <i>Echium leucophaeum</i> in Tenerife, Spain
CBS6039	<i>Cryptococcus amyloletus</i>	A1B1	frass of the beetle <i>Enneadesmus forficulus</i> in South Africa
CBS6273	<i>Cryptococcus amyloletus</i>	A2B2	frass of the beetle <i>Sinoxylon ruficorne</i> in South Africa
<b>SSA790</b>	<i>Cryptococcus amyloletus</i>	A1B2	progeny of CBS6039 x CBS6273
<b>SSB821</b>	<i>Cryptococcus amyloletus</i>	A2B1	progeny of CBS6039 x CBS6273
<b>SSC103</b>	<i>Cryptococcus amyloletus</i>	A1B1	progeny of SSA790 x SSB821
SSC104	<i>Cryptococcus amyloletus</i>	A2B2	progeny of SSA790 x SSB821
SSC118	<i>Cryptococcus amyloletus</i>	A2B2	progeny of SSA790 x SSB821
SSC119	<i>Cryptococcus amyloletus</i>	A2B2	progeny of SSA790 x SSB821
<b>SSC120</b>	<i>Cryptococcus amyloletus</i>	A1B2	progeny of SSA790 x SSB821
SSC123	<i>Cryptococcus amyloletus</i>	A2B1	progeny of SSA790 x SSB821
SSC124	<i>Cryptococcus amyloletus</i>	A2B2	progeny of SSA790 x SSB821
<b>SSC125</b>	<i>Cryptococcus amyloletus</i>	A1B1	progeny of SSA790 x SSB821
SSC128	<i>Cryptococcus amyloletus</i>	A2B1	progeny of SSA790 x SSB821
<b>SSC129</b>	<i>Cryptococcus amyloletus</i>	A1B1	progeny of SSA790 x SSB821
<b>SSC130</b>	<i>Cryptococcus amyloletus</i>	A1B2	progeny of SSA790 x SSB821

Strains selected for Illumina whole-genome sequencing are indicated in boldface

**Table 2. Germination rates of spores dissected from crosses between DSM27421 x *C. amyloletus*.**

cross	basidium #	spores plated	spores germinated	Average germination rate/basidium
DSM27421 (A2B3) × CBS6039 (A1B1)	1	11	0	0.57%
	2	16	0	
	3	42	0	
	4	29	1 (3%)	
	5	28	0	
	6	27	0	
DSM27421 (A2B3) × SSC103 (A1B1)	1	10	0	5.79%
	2	48	0	
	3	17	1 (6%)	
	4	46	0	
	5	39	9 (23%)	
DSM27421 (A2B3) × SSC120 (A1B2)	1	14	0	16.50%
	2	34	7* (21%)	
	3	63	29* (46%)	
	4	36	0	
	5	34	11 (32%)	
	6	22	0	
DSM27421 (A2B3) × SSC125 (A1B1)	1	9	0	22.17%
	2	56	20 (36%)	
	3	56	11 (20%)	
	4	15	5 (33%)	
DSM27421 (A2B3) × SSC129 (A1B1)	1	19	3* (16%)	3.95%
	2	56	0	
	3	29	0	
	4	53	0	
DSM27421 (A2B3) × SSC130 (A1B2)	1	10	0	12.82%
	2	14	0	
	3	13	5 (38%)	

Progeny selected for Illumina whole-genome sequencing are indicated with asterisks.



**Table 3. Mating ability and genotypic characterization of progeny from crosses between *C. amyloletus* and DSM27421.**

Origin	Strain	Mating phenotype	Mating genotype	mitSSU genotype
-	SSC120	A1B2	A1B2	0
-	SSC129	A1B1	A1B1	0
-	DSM27421	A2B3	A2B3	1
DSM27421 (A2B3) × SSC120 (A1B2), basidium 3	<b>ARP32</b>	-	A1B2	1
	ARP33	A1B3	A1B3	1
	<b>ARP34</b>	-	A1B2	1
	<b>ARP35</b>	-	A1B2	1
	<b>ARP36</b>	-	A1B2	1
	ARP37	A1B3	A1B3	1
	<b>ARP38</b>	-	A1B2	1
	ARP39	-	A1B2	1
	ARP40	A1B3	A1B3	1
	ARP41	-	A1B2	1
	ARP42	-	A1B2	1
	ARP43	-	A1B2	1
	<b>ARP44</b>	A1B3	A1B3	1
	<b>ARP45</b>	A1B3	A1B3	1
	ARP46	-	A1B2	1
	ARP47	A1B3	A1B3	1
	ARP48	A1B3	A1B3	1
	ARP49	A1B3	A1B3	1
	ARP50	-	A1B2	1
	ARP51	A1B3	A1B3	1
	ARP52	A1B3	A1B3	1
	ARP53	-	A1B2	1
	ARP54	A1B3	A1B3	1
	ARP55	A1B3	A1B3	1
	<b>ARP56</b>	A2B2	A2B2	1
	ARP57	A1B3	A1B3	1
	ARP58	A1B3	A1B3	1
	<b>ARP59</b>	-	A1B2	1
<b>ARP60</b>	A1B3	A1B3	1	
DSM27421 (A2B3) × SSC120 (A1B2), basidium 2	<b>ARP61</b>	A2B3	A2B3	1
	<b>ARP62</b>	-	A2B2	1
	<b>ARP63</b>	-	A2B2	1
	<b>ARP64</b>	A2B3	A2B3	1
	<b>ARP65</b>	A2B3	A2B3	1
	<b>ARP66</b>	-	A2B2	1
	<b>ARP67</b>	A2B3	A2B3	1
DSM27421 (A2B3) × SSC129 (A1B1), basidium 1	<b>ARP68</b>	A1B1	A1B1	1
	<b>ARP70</b>	A1B1	A1B1	1
	<b>ARP71</b>	A1B1	A1B1	1

Progeny selected for Illumina whole-genome sequencing are indicated in boldface.

## Acknowledgments

We thank the Centraalbureau voor Schimmelcultures (CBS) (The Netherlands) for providing yeast strains, Marco Guerreiro (Ruhr University, Bochum) for providing assistance in preliminary identification of the DSM27421 isolate, and Valerie Lapham of the Center for Electron Microscopy at NC State University for assistance with SEM. MN thank Profs. Ulrich Kück and Christopher Grefen for support at the Botany Department of the Ruhr-University Bochum.

## References

1. Orr HA, Turelli M. The evolution of postzygotic isolation: accumulating Dobzhansky-Muller incompatibilities. *Evolution*. 2001;55(6):1085-94. pmid: 11475044.
2. Bennett RJ, Turgeon BG. Fungal Sex: The Ascomycota. *Microbiol Spectr*. 2016;4(5). pmid: 27763253.
3. Coelho MA, Bakkeren G, Sun S, Hood ME, Giraud T. Fungal Sex: The Basidiomycota. *Microbiol Spectr*. 2017;5(3). pmid: 28597825.
4. Burton RS, Pereira RJ, Barreto FS. Cytonuclear genomic interactions and hybrid breakdown. *Annu Rev Ecol Evol Syst*. 2013;44(1):281-302.
5. Chou JY, Leu JY. Speciation through cytonuclear incompatibility: insights from yeast and implications for higher eukaryotes. *Bioessays*. 2010;32(5):401-11. pmid: 20414898.
6. Chou J-Y, Hung Y-S, Lin K-H, Lee H-Y, Leu J-Y. Multiple molecular mechanisms cause reproductive isolation between three yeast species. *PLoS Biology*. 2010;8(7):e1000432. pmid: 20652018.
7. Lee HY, Chou JY, Cheong L, Chang NH, Yang SY, Leu JY. Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell*. 2008;135(6):1065-73. pmid: 19070577.

8. Chambers SR, Hunter N, Louis EJ, Borts RH. The mismatch repair system reduces meiotic homeologous recombination and stimulates recombination-dependent chromosome loss. *Mol Cell Biol.* 1996;16(11):6110-20. pmid: 8887641.
9. 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. pmid: 14635842.
10. Sun S, Xu J. Genetic analyses of a hybrid cross between serotypes A and D strains of the human pathogenic fungus *Cryptococcus neoformans*. *Genetics.* 2007;177(3):1475-86.
11. Sun S, Xu J. Chromosomal rearrangements between serotype A and D strains in *Cryptococcus neoformans*. *PLoS One.* 2009;4(5):e5524. pmid: 19436753.
12. Rogers DW, McConnell E, Ono J, Greig D. Spore-autonomous fluorescent protein expression identifies meiotic chromosome mis-segregation as the principal cause of hybrid sterility in yeast. *PLoS Biol.* 2018;16(11):e2005066. pmid: 30419022.
13. Blackwell M. The fungi: 1, 2, 3 ... 5.1 million species? *Am J Bot.* 2011;98(3):426-38. pmid: 21613136.
14. Hawksworth DL, Lucking R. Fungal diversity revisited: 2.2 to 3.8 million species. *Microbiol Spectr.* 2017;5(4). pmid: 28752818.
15. O'Brien HE, Parrent JL, Jackson JA, Moncalvo JM, Vilgalys R. Fungal community analysis by large-scale sequencing of environmental samples. *Appl Environ Microbiol.* 2005;71(9):5544-50. pmid: 16151147.
16. Floudas D, Binder M, Riley R, Barry K, Blanchette RA, Henrissat B, et al. The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. *Science.* 2012;336(6089):1715-9. pmid: 22745431.
17. Heitman J, Howlett BJ, Crous PW, Stukenbrock EH, James TY, Gow NAR. *The Fungal Kingdom: American Society of Microbiology; 2017.*
18. Agapow PM, Bininda-Emonds OR, Crandall KA, Gittleman JL, Mace GM, Marshall JC, et al. The impact of species concept on biodiversity studies. *Q Rev Biol.* 2004;79(2):161-79. pmid: 15232950.
19. Taylor JW, Turner E, Townsend JP, Dettman JR, Jacobson D. Eukaryotic microbes, species recognition and the geographic limits of species: examples from the kingdom Fungi. *Philos Trans R Soc Lond B Biol Sci.* 2006;361(1475):1947-63. pmid: 17062413.
20. Liti G, Barton DB, Louis EJ. Sequence diversity, reproductive isolation and species concepts in *Saccharomyces*. *Genetics.* 2006;174(2):839-50. pmid: 16951060.

21. Skosireva I, James TY, Sun S, Xu J. Mitochondrial inheritance in haploid x non-haploid crosses in *Cryptococcus neoformans*. *Curr Genet*. 2010;56(2):163-76. pmid: 20127336.
22. Taylor JW, Jacobson DJ, Kroken S, Kasuga T, Geiser DM, Hibbett DS, et al. Phylogenetic species recognition and species concepts in fungi. *Fungal Genet Biol*. 2000;31(1):21-32. pmid: 11118132.
23. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, Levesque CA, et al. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *Proc Natl Acad Sci U S A*. 2012;109(16):6241-6. pmid: 22454494.
24. Bouklas T, Fries BC. *Cryptococcus neoformans* constitutes an ideal model organism to unravel the contribution of cellular aging to the virulence of chronic infections. *Curr Opin Microbiol*. 2013;16(4):391-7. pmid: 23631868.
25. Loftus BJ, Fung E, Roncaglia P, Rowley D, Amedeo P, Bruno D, et al. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science*. 2005;307(5713):1321-4. pmid: 15653466.
26. 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 Genet Biol*. 2015;78:16-48. pmid: 25721988.
27. Rajasingham R, Smith RM, Park BJ, Jarvis JN, Govender NP, Chiller TM, et al. Global burden of disease of HIV-associated cryptococcal meningitis: an updated analysis. *Lancet Infect Dis*. 2017;17(8):873-81. pmid: 28483415.
28. Findley K, Rodriguez-Carres M, Metin B, Kroiss J, Fonseca A, Vilgalys R, et al. Phylogeny and phenotypic characterization of pathogenic *Cryptococcus* species and closely related saprobic taxa in the Tremellales. *Eukaryot Cell*. 2009;8(3):353-61. pmid: 19151324.
29. Findley K, Sun S, Fraser JA, Hsueh Y-P, Averette AF, Li W, et al. Discovery of a modified tetrapolar sexual cycle in *Cryptococcus amyloletus* and the evolution of *MAT* in the *Cryptococcus* species complex. *PLoS Genet*. 2012;8(2):e1002528.
30. 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. pmid: 20502678.
31. 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. pmid: 20224779.
32. 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;10.1007/978-3-319-61575-2\_7. Cham: Springer International Publishing; 2017. p. 179-210.
  33. 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. pmid: 28800596.
  34. James TY. Why mushrooms have evolved to be so promiscuous: Insights from evolutionary and ecological patterns. Fungal Biol Rev. 2015;29(3):167-78.
  35. Nieuwenhuis BP, Billiard S, Vuilleumier S, Petit E, Hood ME, Giraud T. Evolution of uni- and bifactorial sexual compatibility systems in fungi. Heredity. 2013;111(6):445-55. pmid: 23838688.
  36. Fraser JA, Diezmann S, Subaran RL, Allen A, Lengeler KB, Dietrich FS, et al. Convergent evolution of chromosomal sex-determining regions in the animal and fungal kingdoms. PLoS Biol. 2004;2(12):e384. pmid: 15538538.
  37. Mittelbach M, Yurkov AM, Nocentini D, Nepi M, Weigend M, Begerow D. Nectar sugars and bird visitation define a floral niche for basidiomycetous yeast on the Canary Islands. BMC Ecol. 2015;15:2. pmid: 25638173.
  38. 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. pmid: 26955199.
  39. Salichos L, Rokas A. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature. 2013;497(7449):327-31. pmid: 23657258.
  40. Salichos L, Stamatakis A, Rokas A. Novel information theory-based measures for quantifying incongruence among phylogenetic trees. Mol Biol Evol. 2014;31(5):1261-71. pmid: 24509691.
  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. pmid: 24743168.

42. 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;10.1073/pnas.1713725115. pmid: 29507212.
43. Kües U, James TY, Heitman J. Mating type in basidiomycetes: unipolar, bipolar, and tetrapolar patterns of sexuality. In: Pöggeler S, Wöstemeyer J, editors. *Evolution of fungi and Fungal-like organisms. The Mycota. 6*. Heidelberg, Germany: Springer Berlin Heidelberg; 2011. p. 97-160.
44. Branco S, Badouin H, Rodriguez de la Vega RC, Gouzy J, Carpentier F, Aguileta G, et al. Evolutionary strata on young mating-type chromosomes despite the lack of sexual antagonism. *Proc Natl Acad Sci U S A*. 2017;114(27):7067-72. pmid: 28630332.
45. Devier B, Aguileta G, Hood ME, Giraud T. Ancient trans-specific polymorphism at pheromone receptor genes in basidiomycetes. *Genetics*. 2009;181(1):209-23. pmid: 19001292.
46. May G, Shaw F, Badrane H, Vekemans X. The signature of balancing selection: fungal mating compatibility gene evolution. *Proc Natl Acad Sci U S A*. 1999;96(16):9172-7. pmid: 10430915.
47. Richman A. Evolution of balanced genetic polymorphism. *Mol Ecol*. 2000;9(12):1953-63. pmid: 11123608.
48. Branco S, Carpentier F, Rodriguez de la Vega RC, Badouin H, Snirc A, Le Prieur S, et al. Multiple convergent supergene evolution events in mating-type chromosomes. *Nat Commun*. 2018;9(1):2000. pmid: 29784936.
49. Coelho MA, Gonçalves P, Sampaio JP. Evidence for maintenance of sex determinants but not of sexual stages in red yeasts, a group of early diverged basidiomycetes. *BMC Evol Biol*. 2011;11(1):249.
50. 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. pmid: 23524993.
51. Maia TM, Lopes ST, Almeida JM, Rosa LH, Sampaio JP, Goncalves P, et al. Evolution of mating systems in basidiomycetes and the genetic architecture underlying mating-type determination in the yeast *Leucosporidium scottii*. *Genetics*. 2015;201(1):75-89. pmid: 26178967.

52. Borts RH, Haber JE. Meiotic recombination in yeast: alteration by multiple heterozygosities. *Science*. 1987;237(4821):1459-65. pmid: 2820060.
53. Datta A, Adjiri A, New L, Crouse GF, Jinks Robertson S. Mitotic crossovers between diverged sequences are regulated by mismatch repair proteins in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 1996;16(3):1085-93. pmid: 8622653.
54. Datta A, Hendrix M, Lipsitch M, Jinks-Robertson S. Dual roles for DNA sequence identity and the mismatch repair system in the regulation of mitotic crossing-over in yeast. *Proc Natl Acad Sci U S A*. 1997;94(18):9757-62. pmid: 9275197.
55. Emmanuel E, Yehuda E, Melamed-Bessudo C, Avivi-Ragolsky N, Levy AA. The role of *AtMSH2* in homologous recombination in *Arabidopsis thaliana*. *EMBO Rep*. 2006;7(1):100-5. pmid: 16311517.
56. Idnurm A. A tetrad analysis of the basidiomycete fungus *Cryptococcus neoformans*. *Genetics*. 2010;185(1):153-63. pmid: 20157004.
57. Gyllensten U, Wharton D, Josefsson A, Wilson AC. Paternal inheritance of mitochondrial DNA in mice. *Nature*. 1991;352(6332):255-7. pmid: 1857422.
58. Voelz K, Ma H, Phadke S, Byrnes EJ, Zhu P, Mueller O, et al. Transmission of hypervirulence traits via sexual reproduction within and between lineages of the human fungal pathogen *Cryptococcus gattii*. *PLoS Genet*. 2013;9(9):e1003771. pmid: 24039607.
59. Xu J, Yan Z, Guo H. Divergence, hybridization, and recombination in the mitochondrial genome of the human pathogenic yeast *Cryptococcus gattii*. *Mol Ecol*. 2009;18(12):2628-42. pmid: 19457185.
60. Yan Z, Hull CM, Sun S, Heitman J, Xu J. The mating type-specific homeodomain genes *SXII $\alpha$*  and *SXII2a* coordinately control uniparental mitochondrial inheritance in *Cryptococcus neoformans*. *Curr Genet*. 2007;51(3):187-95. pmid: 17186242.
61. Yan Z, Sun S, Shahid M, Xu J. Environment factors can influence mitochondrial inheritance in the fungus *Cryptococcus neoformans*. *Fungal Genet Biol*. 2007;44(5):315-22. pmid: 17092744.
62. Zakharov IA, Yurchenko LV, Yarovoy BP. Cytoaduction: Autonomous transfer of cytoplasmic hereditary factors during the mating of yeast cells. *Genetika*. 1969;9(5):136-41.
63. Corcoran P, Anderson JL, Jacobson DJ, Sun Y, Ni P, Lascoux M, et al. Introgression maintains the genetic integrity of the mating-type determining chromosome of the fungus *Neurospora tetrasperma*. *Genome Res*. 2016;26(4):486-98. pmid: 26893460.

64. Kavanaugh LA, Fraser JA, Dietrich FS. Recent evolution of the human pathogen *Cryptococcus neoformans* by intervarietal transfer of a 14-gene fragment. *Mol Biol Evol.* 2006;23(10):1879-90. pmid: 16870684.
65. Mallet J. Hybridization as an invasion of the genome. *Trends Ecol Evol.* 2005;20(5):229-37. pmid: 16701374.
66. Neafsey DE, Barker BM, Sharpton TJ, Stajich JE, Park DJ, Whiston E, et al. Population genomic sequencing of *Coccidioides* fungi reveals recent hybridization and transposon control. *Genome Res.* 2010;20(7):938-46. pmid: 20516208.
67. Van Der Walt JP, Yamada Y, Ferreira NP, Richards PDG. New basidiomycetous yeasts from Southern Africa. II. *Sterigmatomyces wingfieldii* sp. n. *Antonie van Leeuwenhoek.* 1987;53(3):137-42.
68. Fonseca Á, Boekhout T, Fell JW. *Cryptococcus* Vuillemin (1901). In: Kurtzman CP, Fell JW, Boekhout T, editors. *The Yeasts: A Taxonomic Study*; <https://doi.org/10.1016/B978-0-444-52149-1.00138-5>. 5th ed. London: Elsevier; 2011. p. 1661-737.
69. 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. pmid: 28611159.
70. 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. pmid: 26330512.
71. Yurkov A, Guerreiro MA, Sharma L, Carvalho C, Fonseca A. Multigene assessment of the species boundaries and sexual status of the basidiomycetous yeasts *Cryptococcus flavescens* and *C. terrestris* (Tremellales). *PLoS One.* 2015;10(3):e0120400. pmid: 25811603.
72. Guo X, Su H, Shi Q, Fu S, Wang J, Zhang X, et al. *De novo* centromere formation and centromeric sequence expansion in wheat and its wide hybrids. *PLoS Genet.* 2016;12(4):e1005997. pmid: 27110907.
73. Thomae AW, Schade GO, Padeken J, Borath M, Vetter I, Kremmer E, et al. A pair of centromeric proteins mediates reproductive isolation in *Drosophila* species. *Dev Cell.* 2013;27(4):412-24. pmid: 24239514.
74. Roth C, Sun S, Billmyre RB, Heitman J, Magwene PM. A high-resolution map of meiotic recombination in *Cryptococcus deneoformans* demonstrates decreased recombination in unisexual reproduction. *Genetics.* 2018;209(2):567-78. pmid: 29625994.



75. Sun S, Billmyre RB, Mieczkowski PA, Heitman J. Unisexual reproduction drives meiotic recombination and phenotypic and karyotypic plasticity in *Cryptococcus neoformans*. *PLoS Genet*. 2014;10(12):e1004849. pmid: 25503976.
76. Garcia-Solache MA, Izquierdo-Garcia D, Smith C, Bergman A, Casadevall A. Fungal virulence in a lepidopteran model is an emergent property with deterministic features. *mBio*. 2013;4(3):e00100-13. pmid: 23631914.
77. Sun S, Priest SJ, Heitman J. *Cryptococcus neoformans* mating and genetic crosses. *Curr Protoc Microbiol*. 2019;10.1002/cpmc.75:e75. pmid: 30661293.
78. Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey TM, et al. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol*. 2011;12(1):R1. pmid: 21205303.
79. Grad YH, Lipsitch M, Feldgarden M, Arachchi HM, Cerqueira GC, Fitzgerald M, et al. Genomic epidemiology of the *Escherichia coli* O104:H4 outbreaks in Europe, 2011. *Proc Natl Acad Sci U S A*. 2012;109(8):3065-70. pmid: 22315421.
80. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A*. 2011;108(4):1513-8. pmid: 21187386.
81. 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. pmid: 22506599.
82. 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. pmid: 28298431.
83. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014;9(11):e112963. pmid: 25409509.
84. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754-60. pmid: 19451168.
85. Cantarel BL, Korf I, Robb SM, Parra G, Ross E, Moore B, et al. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res*. 2008;18(1):188-96. pmid: 18025269.

86. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, et al. Versatile and open software for comparing large genomes. *Genome Biol.* 2004;5(2):R12. pmid: 14759262.
87. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772-80. pmid: 23329690.
88. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol Biol Evol.* 2013;30(12):2725-9. pmid: 24132122.
89. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol.* 1980;16(2):111-20. pmid: 7463489.
90. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792-7. pmid: 15034147.
91. 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. pmid: 19505945.
92. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics.* 2014;30(9):1312-3. pmid: 24451623.
93. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA.* 2015;6:11. pmid: 26045719.
94. Goodwin TJ, Poulter RT. The diversity of retrotransposons in the yeast *Cryptococcus neoformans*. *Yeast.* 2001;18(9):865-80. pmid: 11427969.
95. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 2007;35(9):3100-8. pmid: 17452365.
96. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res.* 2016;44(W1):W54-7. pmid: 27174935.
97. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. *Trends Genet.* 2000;16(6):276-7. pmid: 10827456.
98. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 2009;19(9):1639-45. pmid: 19541911.

99. Grabherr MG, Russell P, Meyer M, Mauceli E, Alfoldi J, Di Palma F, et al. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics*. 2010;26(9):1145-51. pmid: 20208069.
100. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. *Bioinformatics*. 2011;27(7):1009-10. pmid: 21278367.
101. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491-8. pmid: 21478889.
102. Vilella AJ, Blanco-Garcia A, Hutter S, Rozas J. VariScan: Analysis of evolutionary patterns from large-scale DNA sequence polymorphism data. *Bioinformatics*. 2005;21(11):2791-3. pmid: 15814564.
103. Fu C, Heitman J. *PRMI* and *KAR5* function in cell-cell fusion and karyogamy to drive distinct bisexual and unisexual cycles in the *Cryptococcus pathogenic* species complex. *PLoS Genet*. 2017;13(11):e1007113. pmid: 29176784.
104. Fonseca A. Utilization of tartaric acid and related compounds by yeasts: taxonomic implications. *Can J Microbiol*. 1992;38(12):1242-51. pmid: 1288842.
105. 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.
106. Metsalu T, Vilo J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. *Nucleic Acids Res*. 2015;43(W1):W566-70. pmid: 25969447.

## Figure legends

### **Fig 1. The DSM24721 strain is closely related to *Cryptococcus amylolentus* and *C. wingfieldii*.**

(A) Maximum likelihood (ML) phylogram inferred from a concatenated alignment of the internal transcribed spacer region (ITS1, 5.8S and ITS2), *RPB1* and *TEF1* genes, showing a close phylogenetic relationship between *C. amylolentus* complex and the pathogenic *Cryptococcus* species. All positions containing gaps and missing data were excluded. The tree was rooted with sequences of *Vanrija humicola*. There were a total of 1164 positions in the final dataset. (B) ML phylogeny reconstructed from on the concatenated protein alignments of 4,896 single-copy genes shared across the studied taxa and the outgroup *C. depauperatus*. Measures of gene support frequency (GSF) and internode certainty (IC) are shown at the nodes, and the tree certainty (TC) is given at the bottom. Branch lengths in both trees are given in number of substitutions per site. Former species names assigned to CBS7118 and DSM27421 are used in (A), whereas the new names or new combinations proposed in this study are used in (B). In both trees, bootstrap percentage values from 1,000 replicates are shown at the tree nodes and the type strain of each species is indicated by a superscript capital T. (C) The percentage of identical DNA base pairs is shown for each pairwise combination of the four strains.

### **Fig 2. Sexual structures formed in a cross between *C. amylolentus* CBS6039 and DSM27421.**

(A) Light micrograph of the hyphae and basidia with spore chains (arrows) that formed during this cross, on V8 (pH = 5) mating medium. Note the presence of clamp connections (arrowheads) at the junction of the hyphal cells. (B) Light micrograph of a basidium and spore chains at a higher magnification. (C) and (D) Scanning electron micrographs at a magnification of 10,000 × showing, respectively, details of a fused clamp cell (white arrow), and four chains of spores

arising from the apical surface of a basidium (ba). Scale bars = 50  $\mu\text{m}$  in (A), 20  $\mu\text{m}$  in (B) and 1  $\mu\text{m}$  in (C) and (D).

**Fig 3. Genome-wide comparison between *C. amyloletus*, DSM24721 and CBS7118.**

Circos plot comparing CBS6039 and DSM27421 genome assemblies. Chromosome 7 of DSM27421 is broken into two contigs (7p and 7q), with centromere-specific transposable elements identified at one end of each contig. Of note is a high transposable element load in the DSM27421 genome compared to CBS6039 (track D) and a reciprocal translocation between chromosomes 3 and 12 with the breakpoint mapping within the centromere. (B) This translocation is also shared between CBS6039 and CBS7118, whereas (C) the DSM27421 and CBS7118 genomes are overall syntenic as shown by the links representing collinearity of genomic regions (track E). Small inversions are not represented (see S3 Fig). The chromosomal location of the *P/R* and *HD* mating-type loci is highlighted in green and orange, respectively. Other genomic features, such as tRNA and rRNA gene content, GC content and transposable elements, are depicted in different tracks as given in the key. Expanded views of panels B and C are illustrated in S2 Fig.

**Fig 4. Comparison of the *MAT* loci among the studied *Cryptococcus* strains.**

Synteny maps of the *P/R* (A) and *HD* (B) loci in *C. amyloletus* (CBS6039 and CBS6273), CBS7118, and DSM27421. In both panels, genomic tracks are ordered as shown in (B). Mating pheromone/receptor (*MFA/STE3*) genes and *HD* genes are depicted by dark green and dark orange arrows, respectively, showing the direction of transcription. Additional genes that are present in the fused *MAT* locus in the pathogenic *Cryptococcus* species are shown in light orange or green, while others are shown in white. Vertical blue bars connect orthologs with the same orientation, while pink bars indicate inversions. The regions spanning of the proposed *HD* and

*P/R* loci are highlighted in yellow. The structure of the *HD* locus is largely conserved between the four species, but the *P/R* locus underwent several gene rearrangements, even between strains of the same *P/R* mating-type (see S7 Fig for details). Pheromone gene remnants (*MFa*) in CBS7118 and CBS6039 strains are indicated by asterisks.

**Fig 5. Hybrid progeny show lower levels of recombination.**

Ancestry of the progeny set examined in this study is depicted. The *C. amylolentus* F<sub>2</sub> progeny analyzed are derived from four independent basidia, and four different mating types were scored among this progeny set. Of these, SSC120 and SSC129 were crossed with DSM27421. Basidium 3 from the SSC120 x DSM27421 cross produced progeny ARP32 to ARP60; basidium 2 from that same cross produced progeny ARP61 to ARP67. Basidium 1 from the SSC129 x DSM27421 cross produced progeny ARP68, ARP70, and ARP71. Illumina data was obtained for all strains except those that could not cross with DSM27421 (displayed with faded colors). (B) and (C) Recombination plots of the meiotic progeny derived from *C. amylolentus* conspecific crosses and from heterospecific crosses of DSM27421 with *C. amylolentus*, respectively. SNP distributions along the length of each chromosome for each strain is colored as shown in the key. Regions of frequent haplotype change are indicative of recombination. No crossovers were detected for several chromosomes (tracks enclosed by red lines), particularly in the heterospecific cross. For simplicity, additional basidiospores with identical genotypes (see Table 3) were not plotted in the figure.

**Fig 6. Phenotypic and morphological characteristics distinguishing *C. amylolentus*, *C. floricola*, and *C. wingfieldii*.** (A) The ability of strains of *C. amylolentus*, *C. floricola*, and *C. wingfieldii* to oxidize and assimilate different carbon sources assayed using Biolog YT, FF, and GEN III MicroPlates, is displayed as a heatmap. No scaling is applied to rows. Both rows and

columns are clustered using correlation distance and average linkage. (B) Principle component analysis showing the similarity of growth responses of strains of *C. amyloletus*, *C. floricola*, and *C. wingfieldii* assayed using Biolog YT, FF, and GEN III MicroPlates. Percent of the total variance described by the first two extracted factors are given on the axes. Macroscopic and microscopic morphology is shown for (C) *C. amyloletus* CBS6273, (D) *C. amyloletus* CBS6039, (E) *C. wingfieldii* CBS7118, and (E) *C. floricola* DSM27421. Colonies were grown for one week on YPD or V8 media (inset images) and imaged by light microscopy (left two segments of each panel; scale bars = 50  $\mu\text{m}$ ) and by SEM (right segment of each panel; scale bars = 10  $\mu\text{m}$ ). V8 medium promoted hyphal growth, except for CBS7118, which only grew as a yeast.

## Supporting Information

### S1 Fig. Validation of the DSM27421 genome assembly by chromoblot analysis.

(A) Electrophoretic karyotypes (CHEF) of *C. amyloletus* (CBS6039 and CBS6273), DSM27421 and CBS7118 and selected progeny derived from a conspecific (SS120) and a heterospecific cross (ARP60). Chromosomes of *Saccharomyces cerevisiae* served as size markers. (B) and (C) The size and number of the bands on the CHEF gel for DSM27421 and CBS7118, respectively, is in overall agreement with the contig size obtained from genome sequencing, except for the chromosomes/contigs containing the rDNA array and chromosome 7 of DSM27421, which is broken into two contigs (7q and 7p; asterisk indicates that the size is underestimated). (D) Circos plot showing the genome of DSM27421 assembled into 15 contigs (track A) and the position of Southern hybridization probes used for the chromoblot analysis (shown as bars in track B, color coded as their respective contigs). Two probes targeting different arms of the same chromosome were used for each contig. The black vertical bars at contig ends indicate the presence of telomeric repeats and grey bars indicate the positions of the

predicted centromeres. (E) Results of Southern hybridization using probes described in panel D. In each row, the image on the far left is the EtBr image of the CHEF gel used for the chromoblot analysis, with DSM27421 on the right and the *S. cerevisiae* on the left serving as a size ladder. Probes that hybridized to the same chromosome are grouped together. The labels at the bottom of the hybridization images indicate the names of the probes as described in panel D. It should be noted that probes targeting chromosome 9 (9a and 9b) did not work in our chromoblot analyses. This could be due to the fact that chromosome 9 contains the rDNA array, and because of array expansion and contraction, the population used to make CHEF plugs is comprised of cells with varying numbers of repeats in the array, and consequently, have variant of chromosome 9 with varying sizes that do not form a sharp band in the CHEF analysis.

**S2 Fig. Genomic features of *C. amyloletus*, DSM2741, and CBS7118 assemblies and overall synteny comparison.** (A) and (B) are extended views of Fig 3B and 3C, showing genomic features and overall genome synteny of CBS6039 vs. CBS7118 and DSM27421 vs. CBS7118, respectively. (C) and (D) Linear plots showing the chromosomal position of centromeres, the rDNA array, and *MAT* loci in DSM27421 (*C. floricola* sp. nov.) and CBS7118 (*C. wingfieldii* comb. nov.), respectively. Chromosomes are color coded based on their synteny with *C. amyloletus* CBS6039 chromosomes, except in (E) where DSM27421 was used as reference. A reciprocal translocation between chromosomes 3 and 12 seems to have been driven by intercentromeric recombination and differentiates *C. amyloletus* from the other two sibling species. For simplicity, small inversions are not represented (see S3 Fig).

**S3 Fig. Synteny analysis highlighting diverged and inverted genomic regions between *C. amyloletus*, DSM2741, and CBS7118.** Collinear (light blue) and inverted (red) genomic regions are shown along the length of each chromosome for each of the strains (same order in all



chromosomes). Chromosome/contig numbers in each species are indicated on the top, with an added 'r' indicating contigs with reversed orientation, and are color coded as in S2C Fig. Note a higher number of inversions and divergent regions (white spaces interleaved within blocks of collinearity) at subtelomeric regions, predicted centromeres, and the *P/R MAT* locus. Other features are given in the key.

**S4 Fig. Centromere length and synteny comparison among species.** (A) Synteny comparison across the centromeric region of *C. amyloletus* chromosome 1 is given as an example to illustrate the distinctive length-difference of *C. amyloletus* centromeres compared to DSM27421 and CBS7118. The centromeric regions are depicted as yellow boxes showing their respective length defined as the distance between conserved *CEN*-flanking ORFs (left, dark green; right, dark red) (see S3 Table for additional details). Vertical bars connect regions of similarity (BLASTN hits >1 kb) with the same (blue) or inverted (red) orientation. (B) Comparison of the predicted centromere length among the indicated species. Each dot represents one centromere (those depicted in panel A are shown as yellow dots), and the horizontal red lines depict mean centromere length of the corresponding species. Only 13 predicted centromeres are plotted for *C. floricola* DSM27421 because *CEN7* is incomplete. Other *Cryptococcus* species used in this comparison are *Cryptococcus deuterogattii* (*C. deu.*), *C. neoformans* (*C. neo.*) and *C. deneoformans* (*C. den.*) with *CEN* length obtained from [42]. Different capital letters at the top indicate significantly different means (Tukey-Kramer HSD test,  $p < 0.05$ ).

**S5 Fig. Phylogeny of the pheromone receptor (*STE3*) and predicted pheromone precursor genes.** (A) Phylogeny based on the nucleotide sequences of *STE3 $\alpha$*  (A1) and *STE3 $\alpha$*  (A2) genes from several Tremellales representatives. The *STE3* alleles of *C. amyloletus*, *C. wingfieldii* (CBS7118) and *C. floricola* (DSM27421) are associated with strains of the opposite mating types

in other species as a result of the original arbitrary assignment of mating types in *C. amyloletus* based on mating tests (indicated by asterisks; e.g. the *STE3α/A1* is found in the A2 strains CBS6273 and DSM27421). (B) Sequence alignment and comparison of MF $\alpha$  and MF $\alpha$  pheromone precursors from *C. neoformans*, *C. amyloletus*, DSM27421, and CBS7118. A consensus sequence is given at the top and the black arrow denotes the predicted cleavage site giving rise to the peptide moiety of the mature pheromone.

**S6 Fig. Putative inversion events leading to the extant *P/R* locus configuration in *C.***

***wingfieldii* CBS7118 and *C. floricola* DSM27421.** (A) Compared to *C. amyloletus* CBS6039,

the extant *P/R* locus configuration in *C. wingfieldii* CBS7118 appears to be the result of two successive inversions spanning the chromosomal segments highlighted in orange. (B) The *P/R* locus of *C. floricola* DSM27421 differs from *C. amyloletus* CBS6273 by a single large inversion spanning the chromosomal region highlighted in orange. In both cases, the breakpoints of these inversions are associated with the presence of identical pheromone genes that may constitute large inverted repeats that facilitated inversion.

**S7 Fig. The progeny of *C. amyloletus* x DSM27421 are predominantly euploid.**

Read depth (binned in 5-kb non-overlapping windows) was used to screen for chromosome aneuploidy. For each sequenced strain, read depth was normalized to the median read depth for that strain, log<sub>2</sub>-transformed and plotted as a heatmap with Circos. (A) All of the *C. amyloletus* parents used for crosses were euploid. (B) Two examples (ARP45 and ARP60) of  $n + 1$  aneuploidy in the hybrid progeny are shown.

**S8 Fig. Divergence plots showing no evidence of introgression between *C. amylolentus*, DSM27421 and CBS7118.** Each plot represents the divergence  $k$  (with Jukes-Cantor correction) relative to (A) *C. amylolentus* CBS6039, (B) *C. wingfieldii* CBS7118 and (C) *C. floricola* DSM27421 (y-axis values are in percentage; x-axis values are in Kb). Centromeres, *HD* and *P/R MAT* loci are highlighted as indicated in the key. Each data point represents the average divergence of itself plus two windows on each side.

**S1 Table. Genome sequencing data generated and final genome assembly statistics.**

**S2 Table. Transposable element content of *C. amylolentus*, *C. floricola* and *C. wingfieldii* genomes.** Estimates obtained from the RepeatMasker output using the combined databases: Dfam\_Consensus-20170127, RepBase-20170127.

**S3 Table. List of ORFs flanking the candidate centromere regions in *C. amylolentus*, *C. wingfieldii* and *C. floricola* and predict centromere length.**

**S4 Table. List of primers used in this study.**

**S1 File. Ploidy determination by FACS for strains used in this study.** *C. deneoformans* JEC21 (D $\alpha$ ) and *C. deneoformans* XL143 ( $\alpha$ DD $\alpha$ ) were used as haploid and diploid controls, respectively. Approximately 10,000 cells were analyzed. Propidium iodide area (PI-A) is shown in the x-axis.

**S2 File. Sequence accession numbers.**

Fig 1

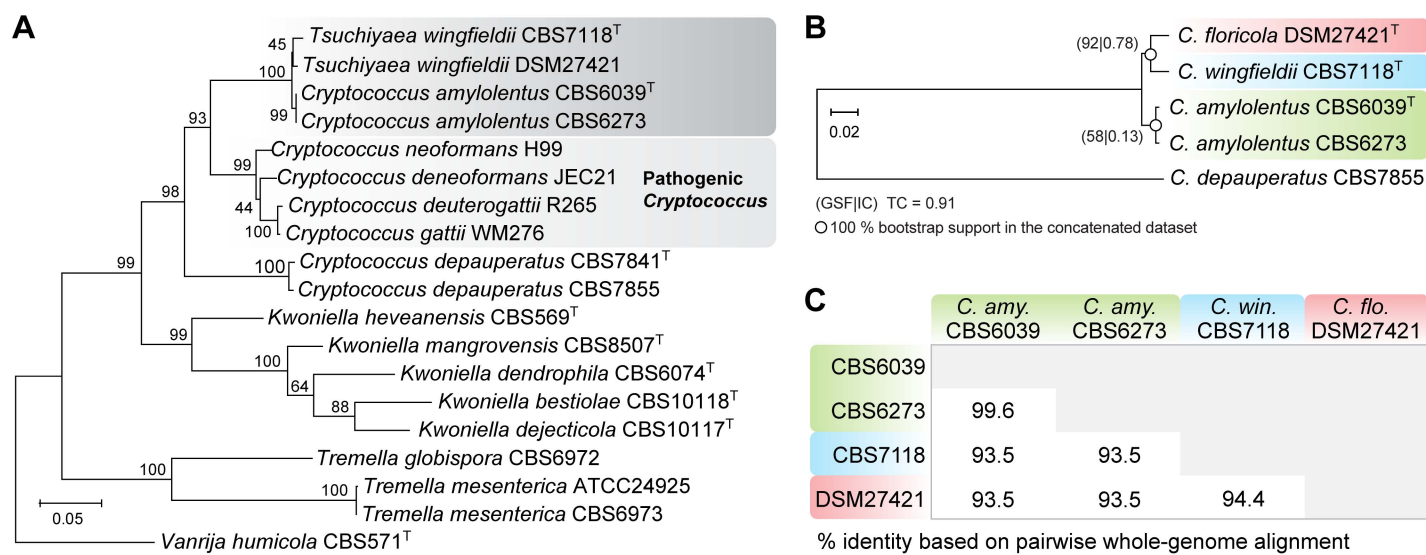


Fig 2

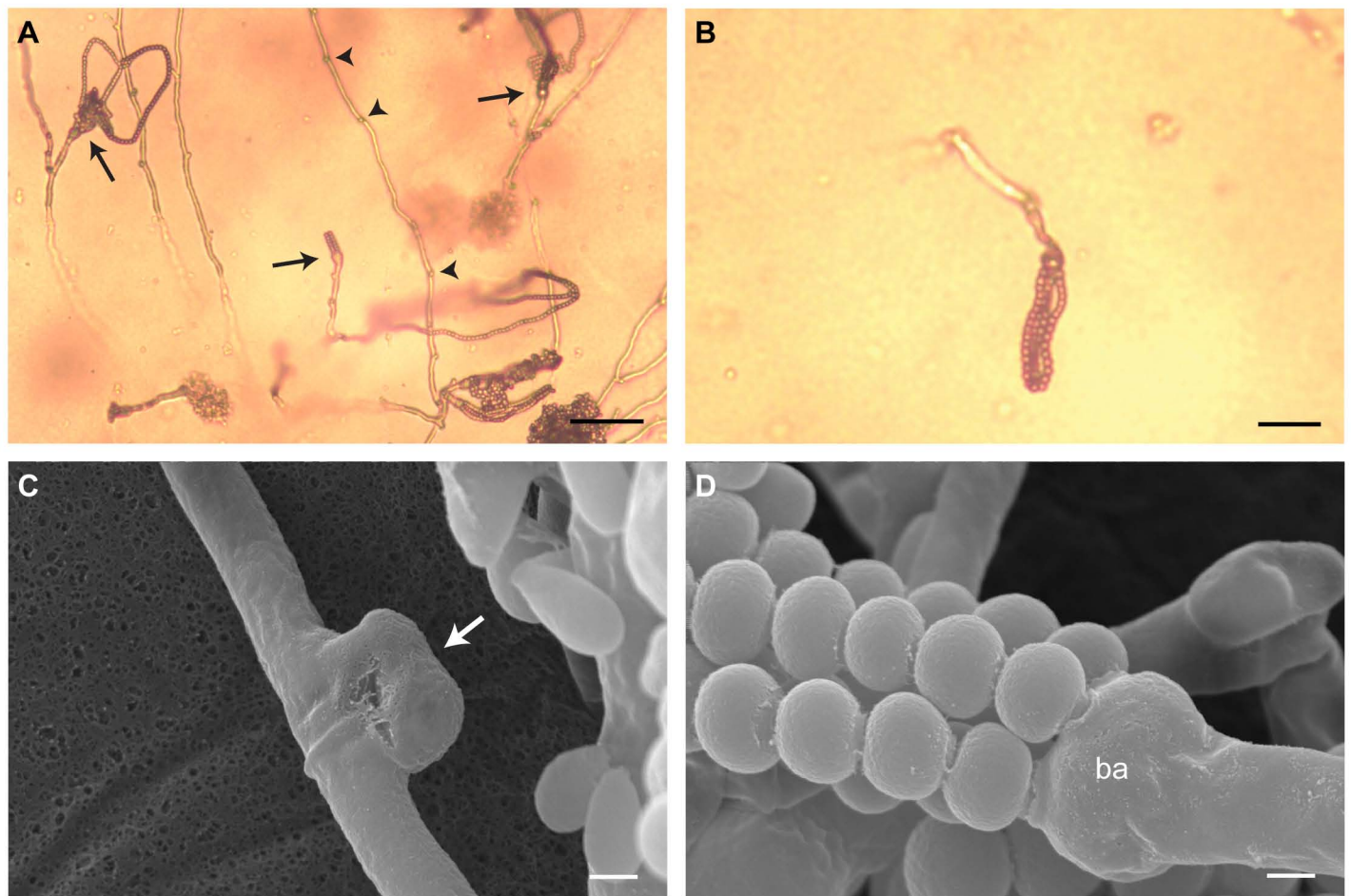


Fig 3

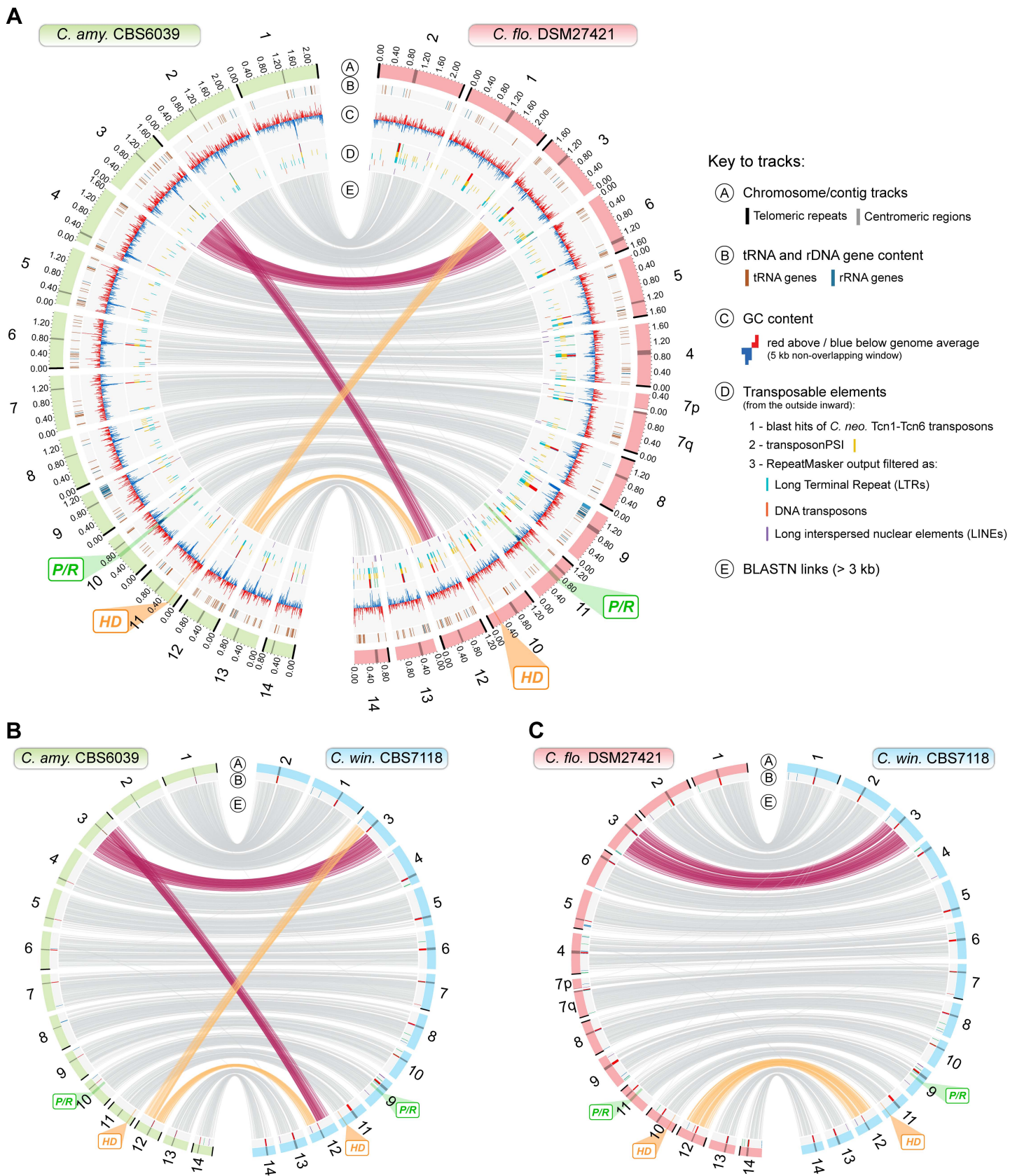


Fig 4

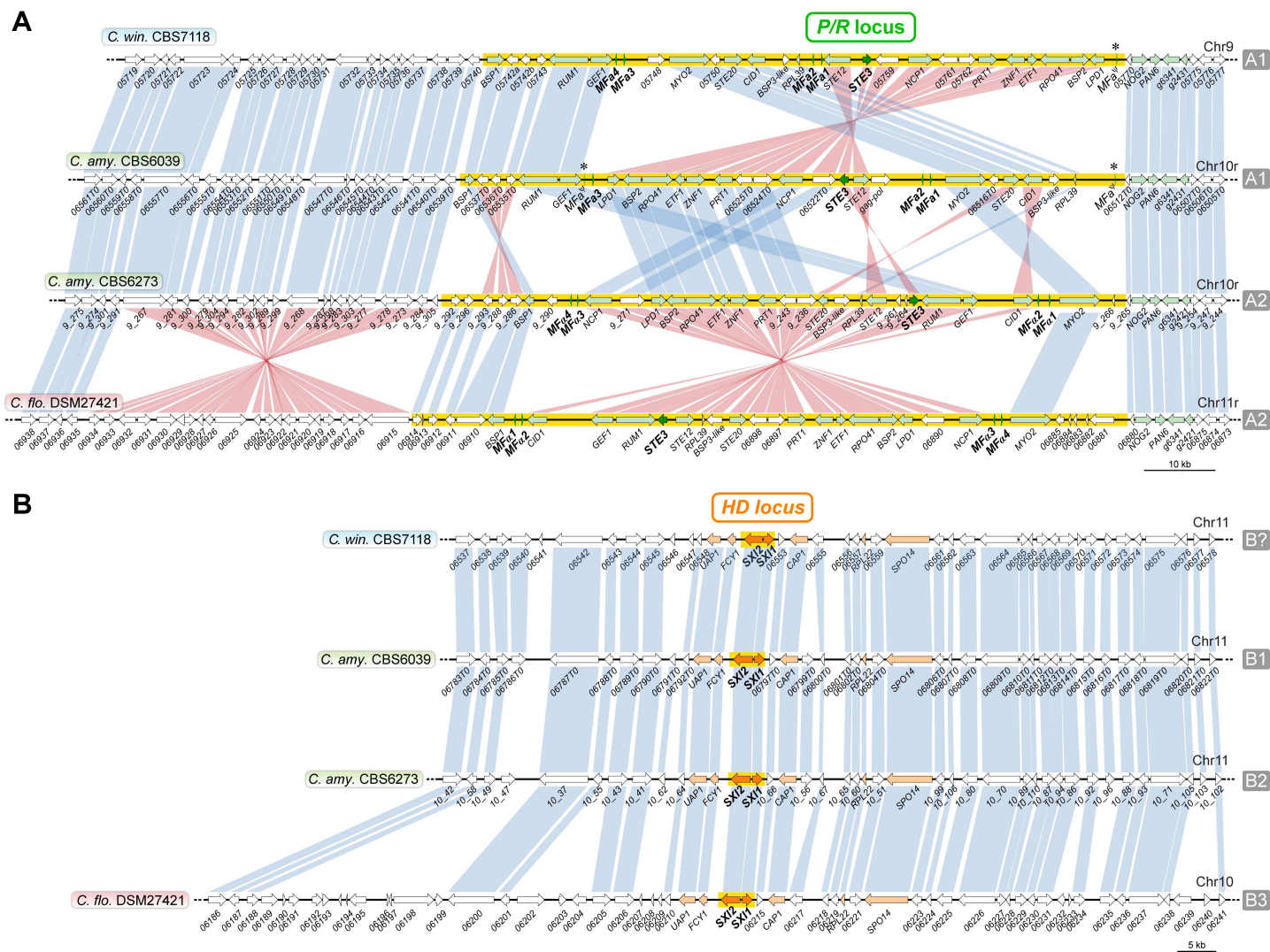


Fig 5

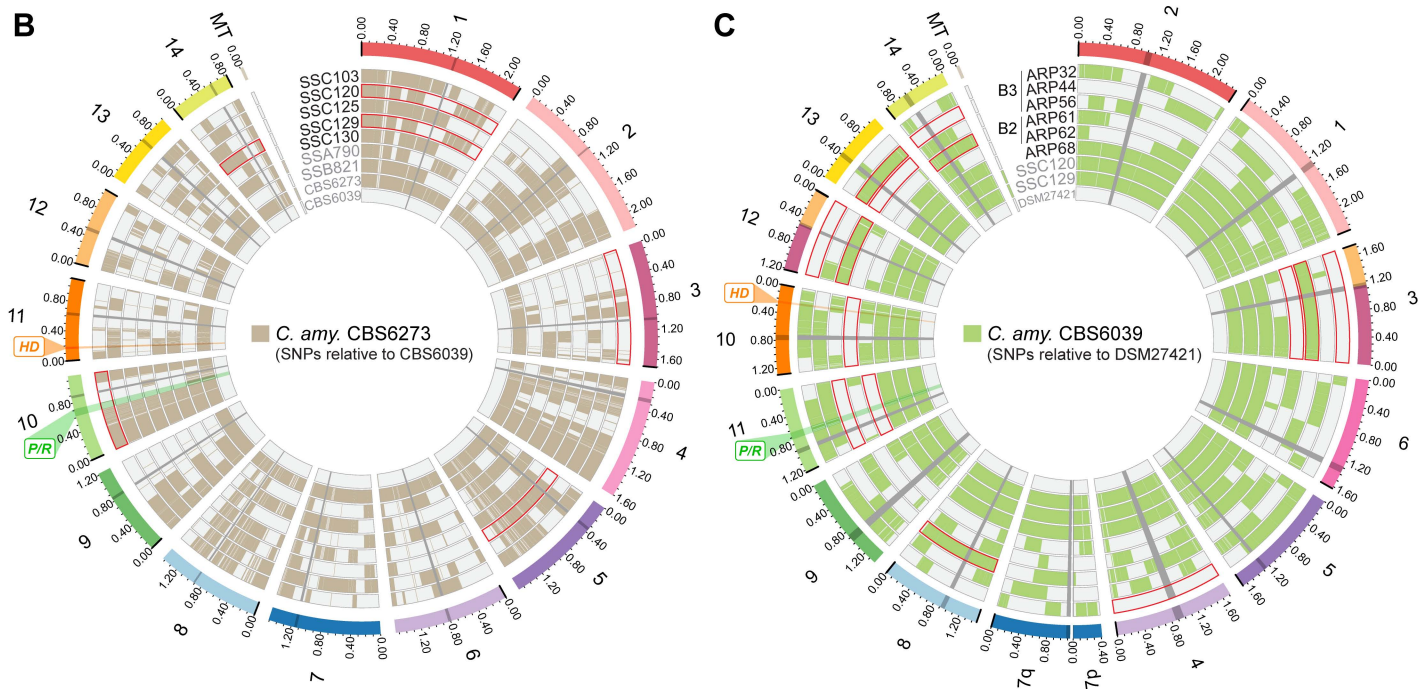
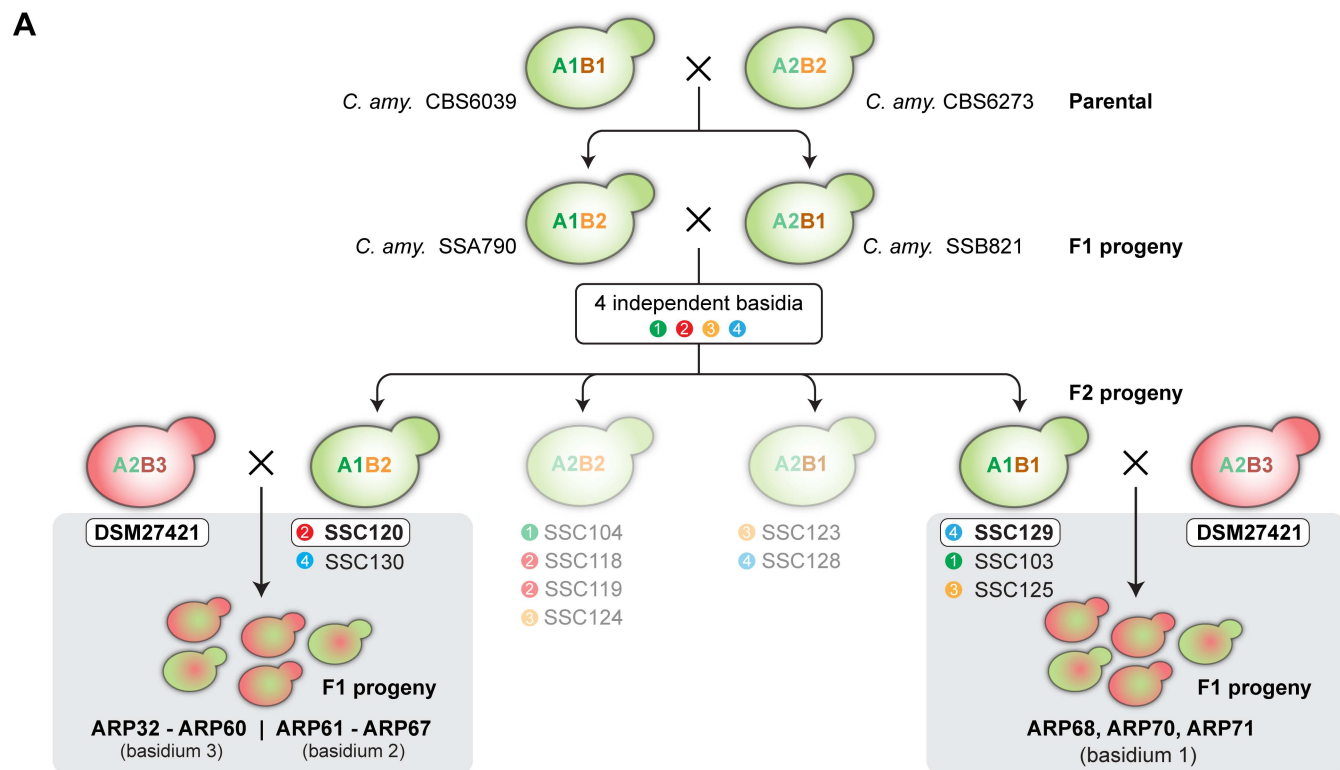




Fig 6

