AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

Chromosome-Level Genome Assembly of a Human Fungal Pathogen Reveals Synteny among Geographically Distinct Species

Mark Voorhies,[a] Shirli Cohen,[b] Terrance P. Shea,[c] Semar Petrus,[b*] José F. Muñoz,[c] Shane Poplawski,[b] William E. Goldman,[d] Todd P. Michael,[b*] Christina A. Cuomo,[c] Anita Sil,[a] Sinem Beyhan[b,e,f]

[a]Department of Microbiology and Immunology, University of California, San Francisco, San Francisco, California, USA
[b]Department of Infectious Diseases, J. Craig Venter Institute, La Jolla, California, USA
[c]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
[d]Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA
[e]Department of Medicine, University of California, San Diego, San Diego, California, USA
[f]Department of Biology, San Diego State University, San Diego, California, USA

Mark Voorhies and Shirli Cohen contributed equally to this work. Author order was determined by the duration worked on this project.

ABSTRACT *Histoplasma capsulatum*, a dimorphic fungal pathogen, is the most common cause of fungal respiratory infections in immunocompetent hosts. *Histoplasma* is endemic in the Ohio and Mississippi River Valleys in the United States and is also distributed worldwide. Previous studies have revealed at least eight clades, each specific to a geographic location: North American classes 1 and 2 (NAm 1 and NAm 2), Latin American groups A and B (LAm A and LAm B), Eurasian, Netherlands, Australian and African, and an additional distinct lineage (H81) comprised of Panamanian isolates. Previously assembled *Histoplasma* genomes are highly fragmented, with the highly repetitive G217B (NAm 2) strain, which has been used for most whole-genome-scale transcriptome studies, assembled into over 250 contigs. In this study, we set out to fully assemble the repeat regions and characterize the large-scale genome architecture of *Histoplasma* species. We resequenced five *Histoplasma* strains (WU24 [NAm 1], G217B [NAm 2], H88 [African], G186AR [Panama], and G184AR [Panama]) using Oxford Nanopore Technologies long-read sequencing technology. Here, we report chromosomal-level assemblies for all five strains, which exhibit extensive synteny among the geographically distant *Histoplasma* isolates. The new assemblies revealed that *RYP2*, a major regulator of morphology and virulence, is duplicated in G186AR. In addition, we mapped previously generated transcriptome data sets onto the newly assembled chromosomes. Our analyses revealed that the expression of transposons and transposon-embedded genes are upregulated in yeast phase compared to mycelial phase in the G217B and H88 strains. This study provides an important resource for fungal researchers and further highlights the importance of chromosomal-level assemblies in analyzing high-throughput data sets.

IMPORTANCE *Histoplasma* species are dimorphic fungi causing significant morbidity and mortality worldwide. These fungi grow as mold in the soil and as budding yeast within the human host. *Histoplasma* can be isolated from soil in diverse regions, including North America, South America, Africa, and Europe. Phylogenetically distinct species of *Histoplasma* have been isolated and sequenced. However, for the commonly used strains, genome assemblies have been fragmented, leading to underutilization of genome-scale data. This study provides chromosome-level assemblies of the commonly used *Histoplasma* strains using long-read sequencing technology. Comparative analysis of these genomes shows largely conserved gene order within the chromosomes. Mapping existing transcriptome data on these new assemblies reveals

clustering of transcriptionally coregulated genes. The results of this study highlight the importance of obtaining chromosome-level assemblies in understanding the biology of human fungal pathogens.

Histoplasma and the closely related pathogens *Blastomyces*, *Paracoccidioides*, and *Coccidioides* are thermally dimorphic fungi that cause fungal infections in immunocompetent and immunocompromised hosts (1). Histoplasmosis, a systemic infection caused by *Histoplasma*, is a significant cause of mortality in immunocompromised individuals (2). Although aggressive treatment with antifungals can be successful in clearing the infection, the mortality rate is estimated to be over 50% in some regions of the world (3, 4).

Histoplasma is endemic in the Ohio and Mississippi River Valleys in the United States and is also distributed worldwide, mostly in North America, South America, and Africa. Several *Histoplasma* genomes are publicly available and have been used for detailed analysis of fungal genetics. Seminal studies by Kasuga et al. classified *Histoplasma* isolates into at least eight geographically isolated clades: North American classes 1 and 2 (NAm 1 and NAm 2), Latin American groups A and B (LAm A and LAm B), Eurasian, Netherlands, Australian and African, as well as a distinct lineage (H81) comprised of three Panamanian isolates (5, 6). Later, six phylogenetic groups within the Latin American groups were described (7). Further phylogenetic analysis of 30 additional unassembled genomes (10 NAm 1, 11 NAm 2, 4 LAm A, 3 Panama, and 2 Africa) divided *Histoplasma* species into five genetically distinct lineages, and four of them have been renamed as follows: *Histoplasma capsulatum* (H81 lineage), *Histoplasma mississippiense* (NAm 1), *Histoplasma ohiense* (NAm 2), and *Histoplasma suramericanum* (LAm A) (8).

Histoplasma exists in a saprophytic hyphal form in the soil and produces asexual spores termed conidia. Once conidia are inhaled by the mammalian host, the fungus transitions into its pathogenic yeast form, which can proliferate within host macrophages and cause disease. We and others have been studying the temperature-regulated gene networks that control cell morphology and virulence in *Histoplasma*. We have shown that four transcriptional regulators, Ryp1, -2, -3, and -4, are major regulators of cell morphology and virulence gene expression and are required for yeast-phase growth (9–11), while the signaling mucin Msb2 is required for hyphal-phase growth (12). In these studies, we have used a wide variety of high-throughput molecular biology techniques, including transcriptional profiling, chromatin immunoprecipitation-on-chip analyses, forward genetic screens, and mapping genomic modifications, to study the biology of this important fungal pathogen. However, any chromosome-level patterns in these data were obscured due to the fragmented nature of the existing *Histoplasma* genome assemblies.

Emerging sequencing technology provides complete genome assemblies that can be leveraged in high-throughput analyses. Specifically, long-read sequencing tools such as Oxford Nanopore Technologies have demonstrated significant promise in the field of fungal genomics (13–16). In this study, we resequenced five strains of *Histoplasma*, G217B, H88, G184AR, G186AR, and WU24, which belong to four distinct populations (5). Using a combination of long-read (Oxford Nanopore [ONT]) and short-read (Illumina) sequencing, we *de novo* assembled all five genomes to the chromosomal level. Comparison of the assembled genomes unveiled largely syntenic regions of the chromosomes. Further examination of these genomes revealed that a region containing *RYP2*, a regulator of yeast-phase growth, is duplicated in the *Histoplasma* G186AR strain. Re-analyses of high-throughput data sets also revealed that the transposon genes, as well as genes embedded in transposon-rich regions, display more abundant transcription in the yeast phase in the G217B and H88 strains. This study highlights that complete genome assemblies allow new insights to be drawn from genomic data sets and will open opportunities for future studies of this organism.

RESULTS

Resequencing of *Histoplasma* strains reveals chromosomal-level genome assemblies. Previously assembled *Histoplasma* genomes are highly fragmented, with the highly repetitive G217B strain having 261 contigs (see Table S1A in the supplemental material). To achieve a full assembly of the repeat regions and characterize the large-scale genome architecture of *Histoplasma* species, we resequenced five *Histoplasma* strains (WU24, G217B, H88, G186AR, and G184AR) using Oxford Nanopore Technologies (ONT) to assemble complete chromosomes followed by polishing using Illumina short reads (Table S1A and Fig. S1). The five strains are derived from four distinct populations (5), three of which have recently been proposed as distinct species (7, 8). WU24 is from the North American 1 (NAm 1) clade and renamed as *Histoplasma mississippiense*; G217B is from North American 2 (NAm 2) and renamed as *Histoplasma ohiense*; G186AR and G184AR are Panamanian from the H81 lineage and are renamed as *Histoplasma capsulatum*; and H88 is from the African clade (5–8).

Our results revealed that the genomes of G186AR (31 Mb), G184AR (31 Mb), and WU24 (32 Mb) are significantly smaller than H88 (38 Mb) and G217B (40 Mb), consistent with previously observed coverage differences (8). For two of the *Histoplasma* strains, WU24 and H88, we achieved complete telomere-to-telomere assemblies; H88 has 6 chromosomes while WU24 has 7 chromosomes (Fig. 1). The nearly complete assemblies have 7 large contigs each, with either 12 (G186AR) or 11 (G217B and G184AR) assembled telomeres. For the G217B and G184AR assemblies, we were able to locate the 45S rDNA repeat arrays near the ends of large contigs, consistent with their subtelomeric positioning in other strains (Fig. 1). These observations are consistent with either 6 or 7 chromosomes for the G217B, G186AR, and G184AR strains.

Next, we mapped existing gene annotations to the new assemblies using BLAT (17). For the G217B, H88, G186AR, and G184AR strains, 88 to 92% of the RNA-Seq based annotations (18)—using the G184AR annotations for both G184AR and G186AR—transferred exactly, and 96% of them transferred with no changes to the coding sequence (Table S1B). For WU24, 77% of the previously available gene predictions transferred exactly and 81% transferred with no changes to the coding sequence. Additionally, the mapping of the mating type locus (MTL) based on the previous MTL characterization (19–22) revealed that G217B and WU24 carry *MAT1-1* allele, and thus are "+" mating type isolates, whereas G186AR, G184AR, and H88 carry *MAT1-2* allele, and thus are "−" mating type isolates.

***Histoplasma* genomes have highly variable repetitive regions.** The genomes of *Histoplasma* and *Blastomyces* have been observed to contain a substantial, but widely variable, amount of repetitive long terminal repeat (LTR) transposon content (8, 23, 24). We annotated transposons using two complementary approaches: searching for broadly conserved features using LTRHarvest (25) and searching for protein coding regions homologous to the *gag* and *pol* open reading frames of a previously annotated *Histoplasma* LTR transposon using TBLASTN (26). The regions identified by the two methods had extensive overlap and were merged for further analysis. Results of our analysis revealed that the genomes of G217B and H88 are composed of 20% and 16% LTR transposon sequence, respectively, compared to 4 to 6% for the other three genomes (Fig. 1 and Table S2). This difference in repeat content is sufficient to explain nearly all of the size differences among the genomes; after removing transposon sequence, the genomes differ by at most 3 Mb (~10% of the genome size). Visually, the transposon distribution appeared to be punctate rather than diffuse throughout the genome, which is reminiscent of the transposon distribution in *Blastomyces* (24). To examine this pattern further in *Histoplasma*, we joined transposons within 50 kb of each other. This analysis gave only 66 to 128 transposon-rich regions per genome, with the transposon-rich G217B having only 107 regions compared to 128 for the relatively transposon-poor WU24, confirming the punctate distribution of transposons within *Histoplasma* genomes.

***Histoplasma* strains have highly syntenic chromosomes.** Earlier studies showed that there is variability in chromosome size among *Histoplasma* strains (27, 28). Our
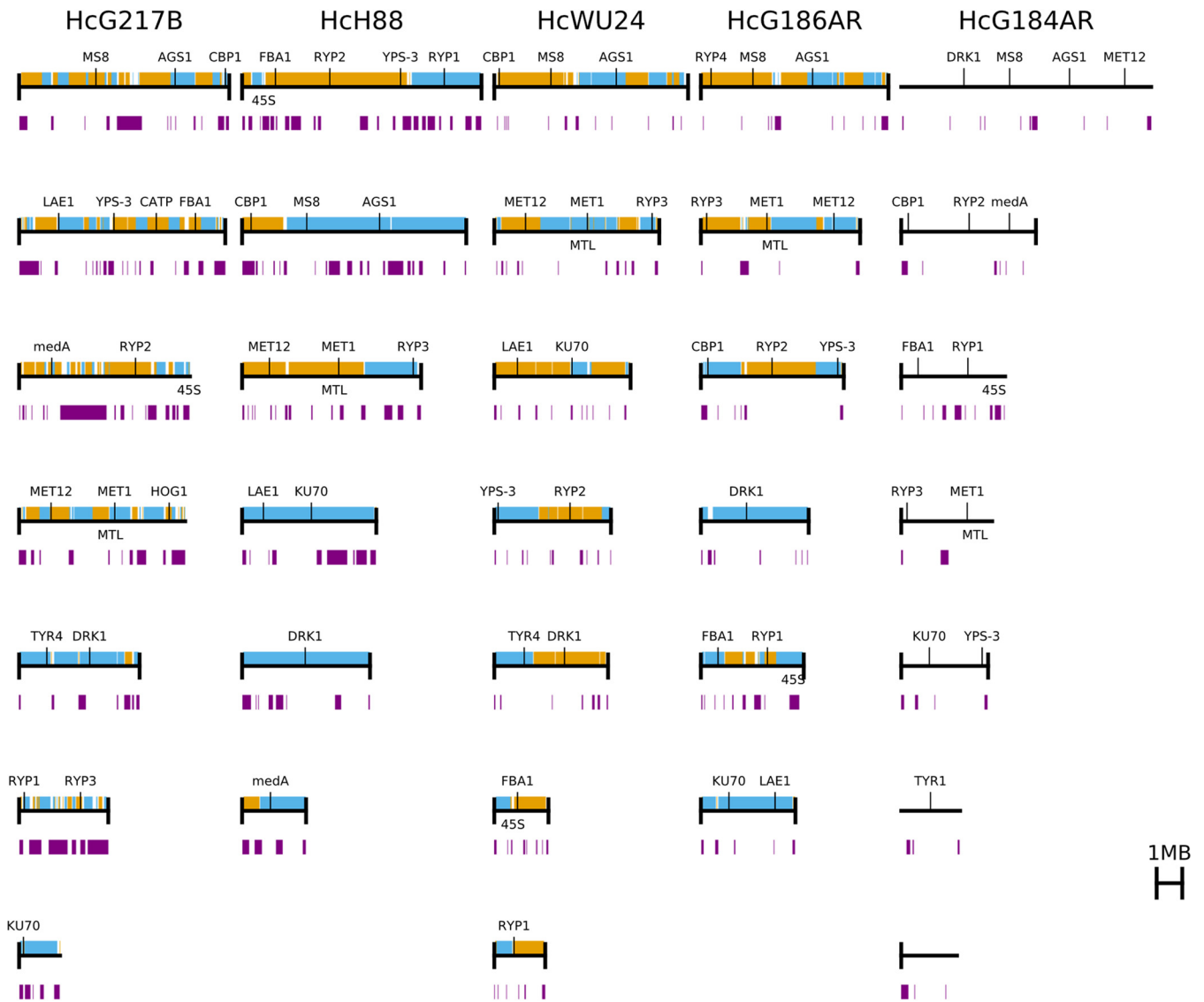
FIG 1 *Histoplasma* genomes are assembled at the chromosomal level. The newly assembled chromosomes of *Histoplasma* genomes are shown as black horizontal lines. Chromosomes or contigs are plotted largest to smallest. Telomeres are shown with black vertical lines at the end of each chromosome when present. Contigs from the previous assemblies are overlaid using alternating colors. Repeat regions (LTR transposons) are indicated with purple blocks below each chromosome. Genes of interest are displayed above each chromosome.

results confirm these early findings and reveal that the repeat content is highly variable among the *Histoplasma* strains, which contributes to size differences. To investigate whether the gene order is conserved despite the chromosome size differences, we mapped the locations of orthologous genes among the *Histoplasma* genomes using our previous ortholog mappings for G217B, G186AR, and H88 (18) supplemented with INPARANOID (29) mappings to WU24, and assuming equivalent gene sets for the highly similar G186AR and G184AR strains. This ortholog-based synteny analysis gave results consistent with whole-genome nucleotide alignments using NUCmer from MUMmer 3 (30). Our analyses showed that the genomes of H88, WU24, G217B, and G186AR are highly syntenic, with chromosomes 2, 3, and 5 of H88 essentially being conserved, with a few rearrangements, in the other strains (Fig. 2 and Fig. S4). Additionally, chromosome 4 of H88 is conserved in G186AR and is conserved but fused to different regions in G217B (with content orthologous to H88 chromosome 1) and WU24 (with content orthologous to H88 chromosome 6). Interestingly, G184AR is more dramatically rearranged relative to G186AR, which was isolated from the same Panama population and is highly identical at the nucleotide level (Fig. 2 and Fig. S4).
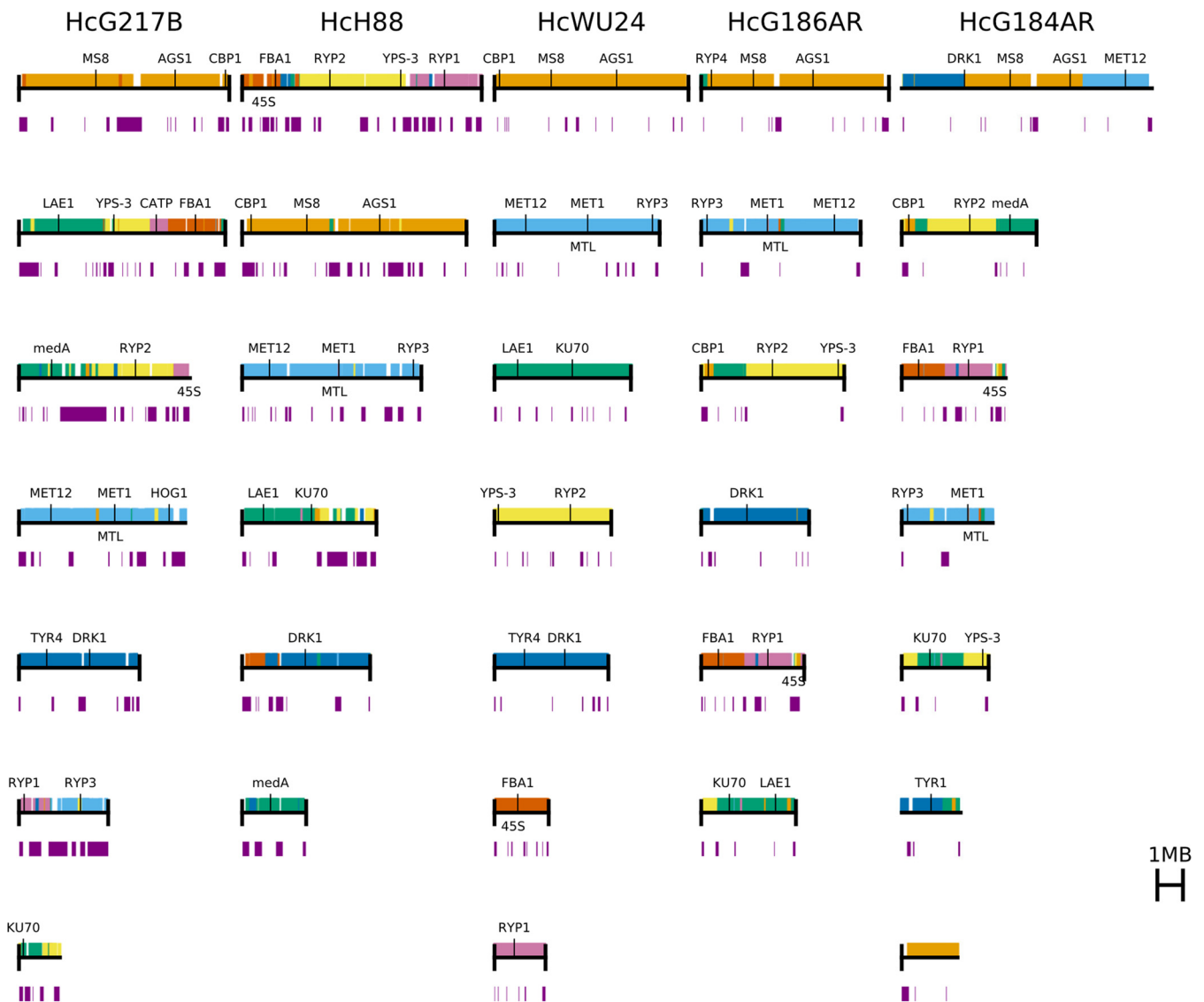
FIG 2 The *Histoplasma* genomes are highly syntenic. Chromosomes are sorted largest to smallest, and regions of synteny are colored according to the WU24 chromosomes in other strains. Purple bars below the chromosomes indicate the repetitive (transposon) regions.

Moreover, the locations of the transposon-rich regions are not conserved. Instead, the transposons appear to be inserted at idiosyncratic locations in each genome without large disruption of the conserved gene order (Fig. 2, Fig. S4 and S5). Nevertheless, there does appear to be conservation of transposon-adjacent context for some genes. For example, the *CBP1* gene, encoding the most studied virulence factor of *Histoplasma* (31) has a conserved subtelomeric location that is flanked by transposons on the telomere side in WU24, G186AR, and G184AR, flanked on both sides in G217B, and embedded in a transposon-rich region in H88 (Fig. 2 and Fig. S5B). The *CBP1* chromosome in WU24, G217B, and H88 also contains *RYP4*, a regulator of morphology and dimorphism in *Histoplasma* (9).

An extensive synteny analysis between *Histoplasma* strains reveals conservation of gene order around many annotated genes (Fig. S5). One of the most striking degrees of synteny can be observed in chromosome 3 of H88, which contains the mating type locus (19, 22) as well as most of the genes for sulfur assimilation; *SRE1*, a regulator of iron acquisition (32); and *RYP3*, a regulator of morphology and dimorphism in *Histoplasma* (11). This chromosome is syntenic along almost its entire length within the *Histoplasma* strains (Fig. S4B and S5C). In addition, high conservation of the region
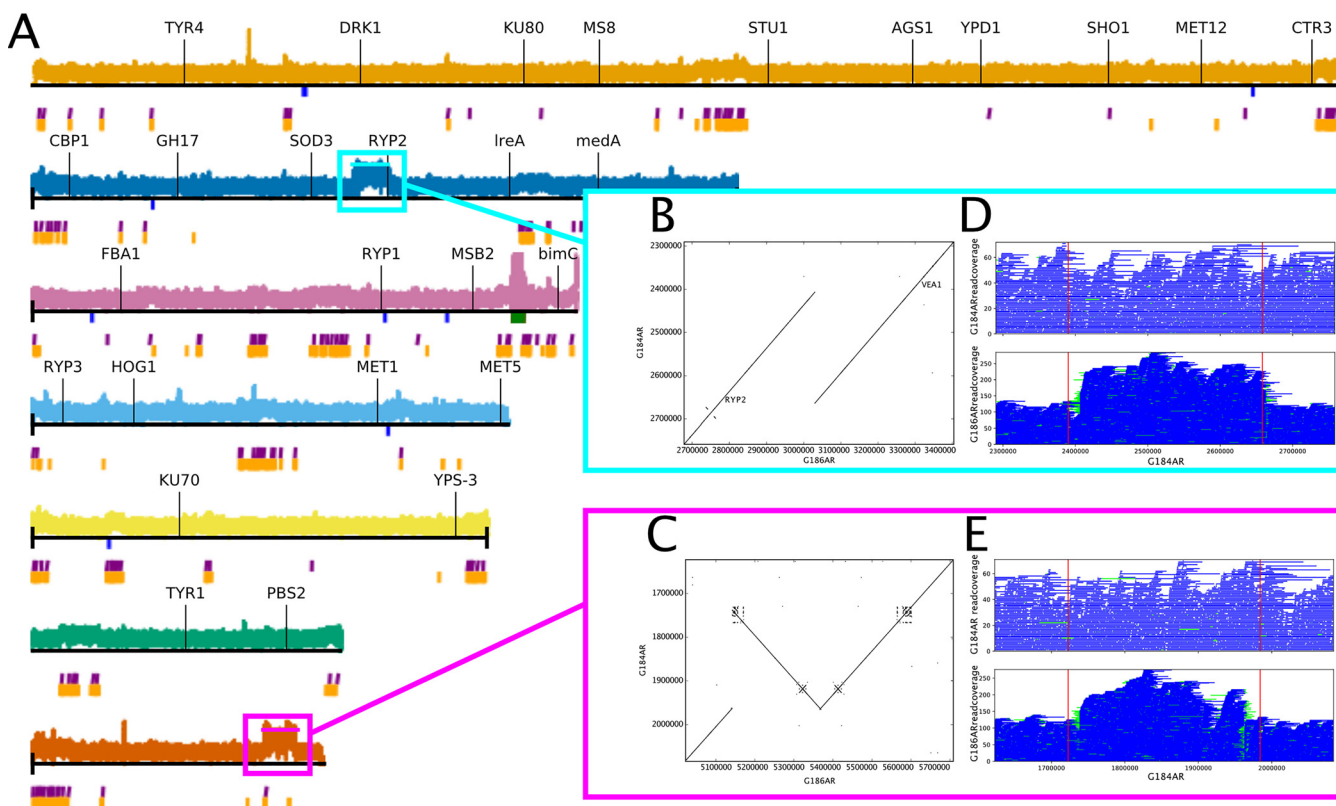
FIG 3 G186AR strain contains two duplicated regions. (A) Coverage for mapping G186AR Illumina reads onto the current G184AR assembly. (B, C) BLASTN-based dotplots for the duplicated regions. (D, E) Coverage for mapping G184AR and G186AR ONT reads onto the current G184AR assembly.

containing *DRK1*, a key regulatory gene of dimorphism in *Blastomyces* and *Histoplasma* (33), is also remarkable among the *Histoplasma* strains and the closely related fungal pathogens *Blastomyces*, *Paracoccidioides*, and *Coccidioides* (Fig. S5D).

Synteny analysis reveals duplication of a major regulator of fungal dimorphism. Our analysis also reveals the duplication of two ~250-kb regions in G186AR relative to G184AR. G186AR and G184AR are the two Panamanian isolates and have a median identity of 99.95% for all large (≥100 kb) alignable regions. The alignments were explored in more detail with BLASTN (Fig. 3B and C). Mapping of our G186AR Illumina reads to the G184AR assembly confirms ~2× coverage in these regions (Fig. 3A), as does mapping of previously published G186AR Illumina reads (SRR6243650 [8]). Mapping of our G186AR ONT reads to the G184AR assembly likewise confirms the duplication; the boundaries of the duplicated regions have soft-clipping for about half of the mapped reads, as would be expected if these reads originate from the internal junction of the duplicated region (Fig. 3D and E).

The first duplication (Fig. 3B) is a direct repeat from ~16 kb to the left (3') of *RYP2* to ~8 kb to the left (5') of *VEA1* (such that *RYP2* is duplicated but *VEA1* is not). *RYP2* and *VEA1* are both velvet transcription factors with roles in morphology in *Histoplasma* (11, 34). The duplicated region is not itself repetitive; it contains 110 gene annotations in G184AR (1 gene per 2.4 kb) and no LTR transposons (Fig. 3A).

The second duplication (Fig. 3C) is an inverted repeat. While mostly genic with ~1 gene per 2.4 kb as for the first region, this second region does contain two sets of LTR transposons in G184AR. This duplicated region contains phospholipase B (*PLB1*) which was observed to have increased expression in G186AR yeast relative to G217B yeast (35). However, the same study also noted similar increased expression of phospholipase D and *OLE1*, which are not located in either duplicated region, so it is not clear if the copy number variation is an underlying cause of the expression difference. The yeast-enriched gene *TSA1* (GenBank AAK54753), a putative cysteine peroxidase, is also in the duplicated region.
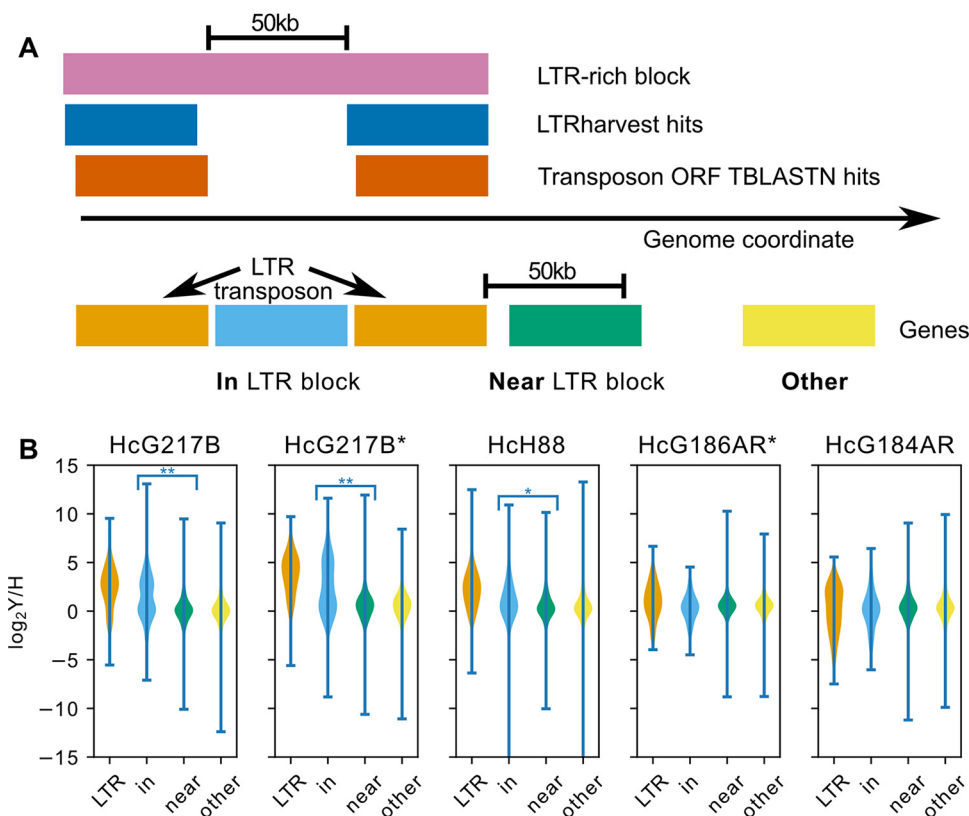
FIG 4 Transposon-embedded genes show yeast-phase enriched expression patterns. (A) Schematic of transcript classification by proximity to LTR transposon; *viz.*, "LTR" (orange), transcript overlaps transposon annotation; "in" (blue), transcript is between two transposon annotations within 50 kb of each other; "near" (green), transcript is within 50 kb of a transposon annotation; "other" (yellow), transcript does not fall into any of the previous categories. (B) Violin plots of yeast/hyphae expression ratios from (18) or (35); *, taken from the Kallisto analysis of (12), colored according to the same scheme as panel A. Significant differences by Wilcoxon rank sum test are indicated: **, $P < 2.2\text{e-}16$, *, $P = 4.7\text{e-}15$.

Transcripts embedded in transposon-rich regions are enriched in yeast phase. Fungal pathogens of plants contain transposon-rich genomic regions that are enriched for virulence factors with increased transcription in the host (36). In contrast, transposon-rich regions were not enriched for differential expression in *Blastomyces*, a close relative of *Histoplasma* (24). We performed similar analysis in *Histoplasma* using the previously published RNA-Seq data sets (18, 35) for four genomes, G217B, H88, G186AR, and G184AR. We classified genes either as "LTR" (transposon), "in" (inside the transposon-rich regions), "near" (within 50 kb of a transposon-rich region), or "other" (the remaining genes) based on their proximity to transposon-rich regions (Fig. 4, Fig. S6 and Data Set S1). In all cases, the expression of transposon-adjacent genes is indistinguishable from transposon-distant genes (Fig. 4). In G217B and H88, genes annotated as transposons have increased expression in yeast compared to hyphae (Fig. 4). Due to the high sequence identity among transposons, we cannot distinguish whether this differential transcription occurs at some or all of the transposon loci. In G217B, the transposon-embedded genes likewise have increased expression in yeast compared to hyphae, significantly distinguishable from the transposon-adjacent and transposon-distant genes ($P < 2.2\text{e-}16$, Wilcoxon test). On closer inspection, this expression is distributed bimodally, with the transposon-embedded genes appearing to have expression ratios drawn either from the transposon or general distributions (Fig. 4). In H88, there is a less pronounced but still significant ($P = 4.7\text{e-}15$, Wilcoxon test) increase in yeast/hyphae expression for transposon-embedded genes. For the Panama strains, G186AR and G184AR, neither the transposons nor the transposon-adjacent genes have expression distributions distinguishable from the remaining genes (Fig. 4).

DISCUSSION

Advancement of sequencing technology, in particular the long reads produced by PacBio or ONT, has greatly improved our understanding of fungal genomes (13, 14, 37). Among over 1,000 publicly available fungal genomes assemblies, there are currently 137 assemblies that have employed ONT sequencing. The most significant characteristic of these assemblies is that the N_{50} of Nanopore assemblies is about 1.4 Mb, whereas the N_{50} of all fungal assemblies is about 140 kb, suggesting at least a 10-fold improvement in the contiguity of the genome sequences by ONT sequencing (Fig. S7).

Histoplasma strains are important fungal pathogens with clinical significance. The previously available genomes of the most highly studied *Histoplasma* strains, G217B and G186AR, were highly fragmented. There are a number of analytical techniques (e.g., evolutionary and taxonomic studies, transcriptional profiling, forward and reverse genetics, and mapping chemical DNA modifications and structural genomic changes) that would benefit from fully assembled genomes. Therefore, in this study, we have utilized ONT sequencing to fully assemble *Histoplasma* genomes.

With additional polishing using short reads from Illumina sequencing, we assembled genomes of five *Histoplasma* strains (WU24, G217B, G186AR, G184AR, and H88) at the chromosomal level. The sizes of the genomes vary from 31 Mb (G186AR and G184AR) to 40 Mb (G217B). The size differences among the strains were attributable to the amount of repeat content. After accounting for the repeat content, the nonrepeat genome content (~30 Mb) of *Histoplasma* is consistent with other Onygenales, such as *Paracoccidioides*, *Blastomyces* and *Emmonsia* (all four within family Ajellomycetaceae) and *Coccidioides* (within family Onygenaceae) (24, 38–40).

Having nearly complete chromosomes in hand, the first analysis we performed was investigating the synteny among *Histoplasma* strains, other close relatives (*Blastomyces dermatitidis*, *Paracoccidioides brasiliensis* and *Coccidioides immitis*), and a distant relative *Aspergillus nidulans*. Despite the variability in the repeat content among *Histoplasma* strains, the four strains WU24, G217B, G186AR, and H88 are highly syntenic, whereas G184AR seems to be the most rearranged strain. This was an unexpected observation, as G186AR and G184AR are both Panamanian strains and are the most similar to each other at the nucleotide level. This indication that geographic location and nucleotide-level resemblance does not correlate with patterns of genomic rearrangement suggests that in some cases, phylogenetic divergence may not be related to large-scale rearrangement. Additional synteny analyses around the previously studied genes (*viz.*, a virulence factor *CBP1*, sulfur assimilation genes, a regulator of iron acquisition *SRE1*, and a histidine kinase *DRK1*) showed a conservation of gene order throughout the *Histoplasma* strains. Moreover, a region around *DRK1* is highly syntenic throughout the Onygenales with substantial conserved gene order even in *Aspergillus* (Fig. S5D). Chromosome-level assemblies revealed these cases of synteny in addition to larger-scale conserved regions that correspond to chromosomes 2, 3, 4, and 5 in H88.

Another important observation from the synteny analysis between *Histoplasma* strains was a duplication in G186AR of a region that contains the important developmental regulator *RYP2*. *RYP2* was previously identified as required for yeast phase in a screen in G217B (11) and, in complex with the velvet transcription factor *RYP3*, directly associates with promoters of yeast-enriched genes (9). The duplicated region also contains *HPD1*, a yeast-enriched gene (41) which has been shown to have a role in morphology in *P. brasiliensis* (42) and *Talaromyces marneffei* (43). Despite the duplication of these regulators in the G186AR strain, the yeast and hyphal transcriptional programs of G186AR (35) are grossly similar to those of G184AR (18), where the region appears to be in single copy. However, a detailed transcriptional analysis of the transition between yeast and hyphae has not been performed in these strains and could uncover kinetic differences that correspond to the duplication of developmental regulators.

One major advantage of having fully assembled genomes is to be able to visualize and analyze whole-genome-level data (e.g., transcriptional profiling) at the chromosomal level. Mapping of existing RNA-Seq data sets for G217B, G186AR, G184AR, and

H88 strains revealed that the transposon-adjacent genes are not distinguishable in their expression patterns from the transposon-distant genes. However, the transposon-embedded genes have significantly increased expression in yeast versus hyphae comparisons in two of the large, transposon-enriched *Histoplasma* genomes; *viz.*, the G217B and H88 strains. While easily missed when analyzing the data relative to the previously available 261 contig G217B assembly, this large-scale pattern of transcriptional regulation is immediately apparent when plotted on the new near-chromosomal assembly. Further studies will reveal the importance of transposon-rich regions in *Histoplasma*. We predict that the new *Histoplasma* assemblies will likewise empower analysis of further genome-scale data sets going forward.

MATERIALS AND METHODS

Sequencing of *Histoplasma* strains. Genomic DNA of *H. capsulatum* G217B, H88, G184AR, G186AR, and WU24 strains was harvested using phenol-chloroform extraction from yeast cultures grown to stationary phase in *Histoplasma* Macrophage Medium (44) at 37°C with 5% CO_2 and shaking at 150 rpm. High molecular weight (HMW) gDNA was used to construct libraries for Illumina and ONT sequencing.

For G217B, G184AR, and H88, DNA libraries were generated using a sequencing kit SQK-LSK108 and sequenced on the ONT MinION R9.4 Flow Cell (FLO-MIN106). For G217B, 207,770 reads were generated with an N_{50} of 24.7 kb; for G184AR, 139,371 reads were generated with an N_{50} of 30.2 kb; and for H88, 119,871 reads were generated with an N_{50} of 24.7 kb. The raw fast5 files were basecalled using Guppy v3.2.2 (ONT) and the resulting fastq reads were utilized to build the assemblies, following the pipelines most appropriate for each strain. Preliminary genome assemblies were created prior to the release of Guppy v3.2.2 and were built using the reads basecalled with Albacore v2.1.3 (ONT). However, all final assemblies only contain data from Guppy-basecalled reads. Additional libraries were generated with the TruSeq PCR-Free kit and sequenced on the Illumina MiSeq, generating paired-end 150-bp reads. Coverage for G217B, G184AR, and H88 was 28×, 16× and 25×, respectively.

For ONT sequencing of G186AR and WU24, libraries were constructed using the ligation sequencing kit (SQK-LSK109) and each library was loaded on a single flowcell (FLO-MIN106D) and run on a GridIon. Basecalling was performed with Guppy using MinKNOW 19.06.8. For WU24, a total of 578,143 reads were generated with an N_{50} of 29.2 kb and estimated genome coverage of 202×. For G186AR, a total of 6,660,775 reads were generated with an N_{50} of 4.8 kb and estimated genome coverage of 456×. For Illumina sequencing of G186AR and WU24, 100 ng of genomic DNA was sheared to ~250 bp using a Covaris LE instrument and prepared for sequencing as previously described (45); libraries were pooled and sequenced on a HiSeq2000 to generate paired 100 base reads, generating 150× coverage for WU24 and 152× coverage for G186AR.

Genome Assemblies. Several methods of genome assembly were utilized to construct the final sequences (Fig. S2). The appropriate method for each genome was selected using an iterative process during which various assembly and polishing tools designed for error-prone long reads were used in different combinations. The methods reported below reflect those that resulted in the most contiguous final genome assemblies with read coverage across any manually joined contigs. The minimap/miniasm/racon (MMR) assembly protocol referenced below entails the following: ONT reads all-vs-all mapped using minimap2 v2.13-r850 with the parameter -x ava-ont (46). Overlapping regions assembled with miniasm v0.3-r179 (47). ONT reads mapped back onto the overlap assembly using minimap2 -x map-ont. Racon v1.3.1 (48) used to build a consensus assembly from this mapping and the ONT reads.

G217B ONT reads basecalled with Albacore were assembled following the MMR protocol, creating assembly *G217B_v1*. This assembly was polished using Illumina reads with 15 rounds of Pilon v1.22 (49). Assembly *G217B_v2* was created using the Guppy basecalled reads. These were first assembled using Flye v2.3.4 (50), then polished using Illumina reads with 16 rounds of Pilon. Alignment of *G217B_v1* and *G217B_v2* using NUCmer v3.1 (30) indicated that one more join could be made between two contigs in *G217B_v2*. Manually stitching this join in *G217B_v2* resulted in assembly *G217B_v3*.

G184AR ONT reads basecalled with Albacore were assembled using wtdbg2 v2.3 and polished with Pilon using Illumina reads to create assembly *G184AR_v1*. Reads basecalled with Guppy were assembled following the MMR protocol and polished with Pilon using Illumina reads to create assembly *G184AR_v2*. *G184AR_v1* was used to guide contig joins on *G184AR_v2*, producing assembly *G184AR_v3*.

H88 ONT reads basecalled with Albacore were assembled using Flye v2.3.4 to construct the genomic DNA sequence *H88_v1*. The mitochondrial DNA sequence *H88_v2* was built from the same reads assembled using wtdbg2 v2.3, then polished with 15 rounds of Pilon using Illumina reads. These two assemblies were combined to form *H88_v3*. Assembly *H88_v4* was created with the Guppy basecalled reads following the MMR protocol, then polished with four rounds of Pilon using the Illumina reads. *H88_v3* and *H88_v4* were aligned with NUCmer v3.1, which implied one join in *H88_v4*. This join in *H88_v4* was manually assembled to create *H88_v5*.

Contigs from *G217B_v3*, *G184AR_v3*, and *H88_v5* were then extended through the telomeres. This was accomplished by using minimap2 to align the ONT reads to the contigs. Reads containing telomeric repeats which extended past the ends of the contigs were extracted, and wtdbg2 (51) was used to assemble the telomeric ends of each chromosome.

G186AR and WU24 ONT reads were assembled into *G186AR_v1* and *WU24_v1* using Canu (version v1.6 with parameters genomeSize = 30000000 stopOnReadQuality=false correctedErrorRate = 0.075)

(52). The resulting assemblies were inspected and refined using alignments of contig ends and spanning reads, and two assemblies, *G186AR_v2* and *WU24_v2*, were generated using Flye (version 2.7b-b1526 with parameter–genome-size = 30000000). These alignments were used to identify and fix misassemblies (including building out the second copies of two large collapsed duplications in G186AR), extend contig ends to telomeric repeats, and make joins between contig ends. The mitochondrial contig in WU24 appears complete based on end overlap, which was trimmed. This refinement produced assemblies *G186AR_v3* and *WU24_v3*.

To produce *G217B_final*, *G184AR_final*, *H88_final*, *G186AR_final*, and *WU24_final*, all five preliminary assemblies were polished first with Medaka (version 0.8.1, using only the Nanopore reads) and then the Medaka-polished contigs were polished using Illumina reads aligned with BWA mem and Pilon (version 1.22 and 1.23), with default settings for both Medaka and Pilon. *G217B_final*, *G184AR_final*, *H88_final*, *G186AR_final*, and *WU24_final* underwent 6, 5, 6, 3, and 3 rounds of Pilon, respectively.

Assembly quality control. Mitochondrial DNA contigs in all assemblies were determined to be complete based on end overlap indicating circularity. Genome assembly quality and completeness was assessed using BUSCO version 4.0.4 (53) with data set eurotiomycetes_odb10.

Gene annotation. Transcript sequences for G217B, H88, and G184AR were taken from Gilmore et al. (18). The G184AR transcripts were used for G186AR as well. Transcript sequences for WU24 were downloaded from the Broad Institute on 6/15/2011 and are available at https://histo.ucsf.edu/downloads/histoplasma_capsulatum_nam1_1_transcripts.fasta.

Transcripts were mapped to the new assemblies using BLAT version 35 invoked as blat -q = dna -t = dna -fine -noTrimA -maxIntron = 10000 -minIdentity = 98, retaining the top scoring hit for each transcript. Genes with CDS corrupted by the BLAT transfer were repaired, where possible, based on TBLASTN of the original protein sequence to the new genome assembly prior to submission to GenBank.

Transposon annotation. Transposons were identified by mapping the translated open reading frames of a representative full-length MAGGY transposon to each genome assembly using TBLASTN from NCBI BLAST+ 2.6.0 with default parameters. Transposons were additionally identified by LTRharvest from GenomeTools 1.5.9 with default parameters. For the purpose of classifying genes by proximity to transposons, transposon-rich regions were defined as contiguous bases within 50 kb of a TBLASTN or LTRharvest identified transposon.

Synteny analysis. Genes orthologous among G217B, H88, and G184AR were taken from Gilmore et al. Orthologs between G184AR and G186AR were assigned based on top BLAT hits to the same query sequence, as described above in "Gene annotation." Orthologs between WU24 and the remaining genomes were assigned by INPARANOID 1.35.

Global synteny patterns were explored by coloring genes according to their chromosome in a reference genome (as in Fig. 2 and Fig. S4) and by ortholog-based dotplots (as in Fig. S3). This method gave comparable results to nucleotide based dotplots generated by NUCmer from MUMmer 3.23.

Detailed synteny plots for a given ordered set of genomes and a given query region on the first genome (as in Fig. S5) were generated as follows. First, we identified all genes in the query region with an ortholog in all of the remaining genomes. Then, for each genome, we identified all regions with at least 5 such orthologs separated by no more than 500 kb. These regions were then plotted in alternating colors with lines connecting orthologous genes between adjacent genomes. The plotted regions were oriented to minimize crossovers between the ortholog-connecting lines.

Analysis of G186AR duplications. Two large (~250 kb) duplications in G186AR were initially observed as mapping artifacts in minimap2 mapping of ONT reads (as minimap2 generates incomplete mappings for reads with multiple high identity matches in the target genome). These duplications were also evident in NUCmer dotplots of G186AR versus G184AR and were explored in detail with BLASTN. Lastly, ~2× coverage of the unduplicated G184AR regions by G186AR reads were confirmed using BWA MEM 0.7.15 for Illumina reads and minimap2 (git commit d90583b83cd81a) for ONT reads.

Fungal phylogeny analysis. The Pfam Gcd10p domain of Gcd10p was identified in each gene set using HMMSEARCH and aligned with HMMALIGN from HMMER3. FASTTREE2 was used to estimate a phylogeny from the resulting protein multiple alignment.

Fungal genome query. Two searches were executed on 19 October 2020 in order to determine the contig N_{50} for all fungal genomes compared to fungal genomes sequenced using Oxford Nanopore technology. Fungal genome assemblies were queried by searching NCBI genome records with the search term *fungi[Organism]*. A total of 1,395 related assembly records were extracted from this search. ONT-based fungal genome assemblies were queried by searching NCBI assembly records with the search term *fungi[Organism] AND nanopore[Sequencing Technology]*. The output of this search was 137 total assembly records. The assembly record XML summaries were downloaded from NCBI and the contig N_{50} values were extracted from them. The median contig N_{50} for the 1,395 fungal assemblies was 140,278 bp. The median contig N_{50} for the 137 Nanopore fungal assemblies was 1,400,166 bp.

Data availability. Sequencing reads and annotated genome assemblies were submitted to GenBank under BioProjects PRJNA682643, PRJNA682644, PRJNA682645, PRJNA682647, and PRJNA682648.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

DATA SET S1, TXT file, 4 MB.

FIG S1, PDF file, 0.02 MB.

FIG S2, PDF file, 0.2 MB.

FIG S3, PDF file, 0.2 MB.
FIG S4, PDF file, 1 MB.
FIG S5, PDF file, 0.3 MB.
FIG S6, PDF file, 0.9 MB.
FIG S7, PDF file, 0.2 MB.
TABLE S1, DOCX file, 0.02 MB.
TABLE S2, DOCX file, 0.02 MB.

REFERENCES

1. Sil A, Andrianopoulos A. 2015. Thermally dimorphic human fungal pathogens: polyphyletic pathogens with a convergent pathogenicity trait. Cold Spring Harb Perspect Med 5:a019794. https://doi.org/10.1101/cshperspect.a019794.

2. Bahr NC, Antinori S, Wheat LJ, Sarosi GA. 2015. Histoplasmosis infections worldwide: thinking outside of the Ohio River Valley. Current Tropical Medicine Rep 2:70–80. https://doi.org/10.1007/s40475-015-0044-0.

3. Damasceno-Escoura AH, Mora DJ, Cardeal AC, Berto-Nascimento JC, Etchebehere RM, de Meneses ACO, Adad SJ, Micheletti AMR, Silva-Vergara ML. 2020. Histoplasmosis in HIV-infected patients: epidemiological, clinical and necropsy data from a Brazilian teaching hospital. Mycopathologia 185:339–346. https://doi.org/10.1007/s11046-020-00435-y.

4. Brown GD, Denning DW, Gow NAR, Levitz SM, Netea MG, White TC. 2012. Hidden killers: human fungal infections. Sci Transl Med 4:165rv13–165rv13. https://doi.org/10.1126/scitranslmed.3004404.

5. Kasuga T, White TJ, Koenig G, Mcewen J, Restrepo A, Castañeda E, Da Silva Lacaz CDA, Heins-Vaccari EM, De Freitas RS, Zancopé-Oliveira RM, Qin Z, Negroni R, Carter DA, Mikami Y, Tamura M, Taylor ML, Miller GF, Poonwan N, Taylor JW. 2003. Phylogeography of the fungal pathogen *Histoplasma capsulatum*. Mol Ecol 12:3383–3401. https://doi.org/10.1046/j.1365-294x.2003.01995.x.

6. Kasuga T, Taylor JW, White TJ. 1999. Phylogenetic relationships of varieties and geographical groups of the human pathogenic fungus *Histoplasma capsulatum* Darling. J Clin Microbiol 37:653–663. https://doi.org/10.1128/JCM.37.3.653-663.1999.

7. Teixeira M de M, Patané JSL, Taylor ML, Gómez BL, Theodoro RC, de Hoog S, Engelthaler DM, Zancopé-Oliveira RM, Felipe MSS, Barker BM. 2016. Worldwide phylogenetic distributions and population dynamics of the genus *Histoplasma*. PLoS Negl Trop Dis 10:e0004732–20. https://doi.org/10.1371/journal.pntd.0004732.

8. Sepúlveda VE, Márquez R, Turissini DA, Goldman WE, Matute DR. 2017. Genome sequences reveal cryptic speciation in the human pathogen *Histoplasma capsulatum*. mBio 8:e01339–17. https://doi.org/10.1128/mBio.01339-17.

9. Beyhan S, Gutierrez M, Voorhies M, Sil A. 2013. A temperature-responsive network links cell shape and virulence traits in a primary fungal pathogen. PLoS Biol 11:e1001614. https://doi.org/10.1371/journal.pbio.1001614.

10. Nguyen VQ, Sil A. 2008. Temperature-induced switch to the pathogenic yeast form of *Histoplasma capsulatum* requires Ryp1, a conserved transcriptional regulator. Proc Natl Acad Sci U S A 105:4880–4885. https://doi.org/10.1073/pnas.0710448105.

11. Webster RH, Sil A. 2008. Conserved factors Ryp2 and Ryp3 control cell morphology and infectious spore formation in the fungal pathogen *Histoplasma capsulatum*. Proc Natl Acad Sci U S A 105:14573–14578. https://doi.org/10.1073/pnas.0806221105.

12. Rodriguez L, Voorhies M, Gilmore S, Beyhan S, Myint A, Sil A. 2019. Opposing signaling pathways regulate morphology in response to temperature

in the fungal pathogen *Histoplasma capsulatum*. PLoS Biol 17:e3000168. https://doi.org/10.1371/journal.pbio.3000168.

13. Salazar AN, Gorter de Vries AR, van den Broek M, Wijsman M, de la Torre Cortés P, Brickwedde A, Brouwers N, Daran J-MG, Abeel T. 2017. Nanopore sequencing enables near-complete de novo assembly of *Saccharomyces cerevisiae* reference strain CEN.PK113-7D. FEMS Yeast Res 17:fox074. https://doi.org/10.1093/femsyr/fox074.

14. McGinty RJ, Rubinstein RG, Neil AJ, Dominska M, Kiktev D, Petes TD, Mirkin SM. 2017. Nanopore sequencing of complex genomic rearrangements in yeast reveals mechanisms of repeat-mediated double-strand break repair. Genome Res 27:2072–2082. https://doi.org/10.1101/gr.228148.117.

15. Yadav V, Sun S, Billmyre RB, Thimmappa BC, Shea T, Lintner R, Bakkeren G, Cuomo CA, Heitman J, Sanyal K. 2018. RNAi is a critical determinant of centromere evolution in closely related fungi. Proc Natl Acad Sci U S A 115:3108–3113. https://doi.org/10.1073/pnas.1713725115.

16. Muñoz JF, Welsh RM, Shea T, Batra D, Gade L, Litvintseva AP, Cuomo CA. 2019. Chromosomal rearrangements and loss of subtelomeric adhesins linked to clade-specific phenotypes in *Candida auris*. bioRxiv https://doi.org/10.1101/754143.

17. Kent WJ. 2002. BLAT: the BLAST-like alignment tool. Genome Res 12:656–664. https://doi.org/10.1101/gr.229202.

18. Gilmore SA, Voorhies M, Gebhart D, Sil A. 2015. Genome-wide reprogramming of transcript architecture by temperature specifies the developmental states of the human pathogen *Histoplasma*. PLoS Genet 11:e1005395. https://doi.org/10.1371/journal.pgen.1005395.

19. Fraser JA, Stajich JE, Tarcha EJ, Cole GT, Inglis DO, Sil A, Heitman J. 2007. Evolution of the mating type locus: insights gained from the dimorphic primary fungal pathogens *Histoplasma capsulatum*, *Coccidioides immitis*, and *Coccidioides posadasii*. Eukaryot Cell 6:622–629. https://doi.org/10.1128/EC.00018-07.

20. Kwon-Chung KJ. 1972. *Emmonsiella capsulata*: perfect state of *Histoplasma capsulatum*. Science 177:368–369. https://doi.org/10.1126/science.177.4046.368.

21. Kwon-Chung KJ. 1972. Sexual stage of *Histoplasma capsulatum*. Science 175:326. https://doi.org/10.1126/science.175.4019.326.

22. Bubnick M, Smulian AG. 2007. The MAT1 locus of *Histoplasma capsulatum* is responsive in a mating type-specific manner. Eukaryot Cell 6:616–621. https://doi.org/10.1128/EC.00020-07.

23. Voorhies M, Foo CK, Sil A. 2011. Experimental annotation of the human pathogen *Histoplasma capsulatum* transcribed regions using high-resolution tiling arrays. BMC Microbiol 11:216. https://doi.org/10.1186/1471-2180-11-216.

24. Muñoz JF, Gauthier GM, Desjardins CA, Gallo JE, Holder J, Sullivan TD, Marty AJ, Carmen JC, Chen Z, Ding L, Gujja S, Magrini V, Misas E, Mitreva M, Priest M, Saif S, Whiston EA, Young S, Zeng Q, Goldman WE, Mardis ER, Taylor JW, McEwen JG, Clay OK, Klein BS, Cuomo CA. 2015. The dynamic

genome and transcriptome of the human fungal pathogen *Blastomyces* and close relative *Emmonsia*. PLoS Genet 11:e1005493. https://doi.org/10.1371/journal.pgen.1005493.

25. Ellinghaus D, Kurtz S, Willhoeft U. 2008. LTRharvest, an efficient and flexible software for *de novo* detection of LTR retrotransposons. BMC Bioinformatics 9:18. https://doi.org/10.1186/1471-2105-9-18.

26. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402. https://doi.org/10.1093/nar/25.17.3389.

27. Canteros CE, Zuiani MF, Ritacco V, Perrotta DE, Reyes-Montes MR, Granados J, Zúñiga G, Taylor ML, Davel G. 2005. Electrophoresis karyotype and chromosome-length polymorphism of *Histoplasma capsulatum* clinical isolates from Latin America. FEMS Immunol Med Microbiol 45:423–428. https://doi.org/10.1016/j.femsim.2005.05.015.

28. Steele PE, Carle GF, Kobayashi GS, Medoff G. 1989. Electrophoretic analysis of *Histoplasma capsulatum* chromosomal DNA. Mol Cell Biol 9:983–987. https://doi.org/10.1128/mcb.9.3.983-987.1989.

29. Remm M, Storm CE, Sonnhammer EL. 2001. Automatic clustering of orthologs and in-paralogs from pairwise species comparisons. J Mol Biol 314:1041–1052. https://doi.org/10.1006/jmbi.2000.5197.

30. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12. https://doi.org/10.1186/gb-2004-5-2-r12.

31. Sebghati TS, Engle JT, Goldman WE. 2000. Intracellular parasitism by *Histoplasma capsulatum*: fungal virulence and calcium dependence. Science 290:1368–1372. https://doi.org/10.1126/science.290.5495.1368.

32. Hwang LH, Seth E, Gilmore SA, Sil A. 2012. SRE1 regulates iron-dependent and -independent pathways in the fungal pathogen *Histoplasma capsulatum*. Eukaryot Cell 11:16–25. https://doi.org/10.1128/EC.05274-11.

33. Nemecek JC, Wüthrich M, Klein BS. 2006. Global control of dimorphism and virulence in fungi. Science 312:583–588. https://doi.org/10.1126/science.1124105.

34. Laskowski-Peak MC, Calvo AM, Rohrssen J, Smulian AG. 2012. VEA1 is required for cleistothecial formation and virulence in *Histoplasma capsulatum*. Fungal Genet Biol 49:838–846. https://doi.org/10.1016/j.fgb.2012.07.001.

35. Edwards JA, Chen C, Kemski MM, Hu J, Mitchell TK, Rappleye CA. 2013. *Histoplasma* yeast and mycelial transcriptomes reveal pathogenic-phase and lineage-specific gene expression profiles. BMC Genomics 14:6951. https://doi.org/10.1186/1471-2164-14-695.

36. Dong S, Raffaele S, Kamoun S. 2015. The two-speed genomes of filamentous pathogens: waltz with plants. Curr Opin Genet Dev 35:57–65. https://doi.org/10.1016/j.gde.2015.09.001.

37. Rhodes J, Abdolrasouli A, Farrer RA, Cuomo CA, Aanensen DM, Armstrong-James D, Fisher MC, Schelenz S. 2018. Genomic epidemiology of the UK outbreak of the emerging human fungal pathogen *Candida auris*. Emerg Microbes Infect 7:1–13. https://doi.org/10.1038/s41426-018-0045-x.

38. Desjardins CA, Champion MD, Holder JW, Muszewska A, Goldberg J, Bailão AM, Brigido MM, Ferreira Me da S, Garcia AM, Grynberg M, Gujja S, Heiman DI, Henn MR, Kodira CD, León-Narváez H, Longo LVG, Ma L-J, Malavazi I, Matsuo AL, Morais FV, Pereira M, Rodríguez-Brito S, Sakthikumar S, Salem-Izacc SM, Sykes SM, Teixeira MM, Vallejo MC, Walter MEMT, Yandava C, Young S, Zeng Q, Zucker J, Felipe MS, Goldman GH, Haas BJ, McEwen JG, Nino-Vega G, Puccia R, San-Blas G, Soares CM, de A, Birren BW, Cuomo CA. 2011. Comparative genomic analysis of human fungal pathogens causing paracoccidioidomycosis. PLoS Genet 7:e1002345. https://doi.org/10.1371/journal.pgen.1002345.

39. Sharpton TJ, Stajich JE, Rounsley SD, Gardner MJ, Wortman JR, Jordar VS, Maiti R, Kodira CD, Neafsey DE, Zeng Q, Hung CY, McMahan C, Muszewska A, Grynberg M, Mandel MA, Kellner EM, Barker BM, Galgiani JN, Orbach MJ, Kirkland TN, Cole GT, Henn MR, Birren BW, Taylor JW. 2009. Comparative genomic analyses of the human fungal pathogens *Coccidioides* and their relatives. Genome Res 19:1722–1731. https://doi.org/10.1101/gr.087551.108.

40. Neafsey DE, Barker BM, Sharpton TJ, Stajich JE, Park DJ, Whiston E, Hung C-Y, McMahan C, White J, Sykes S, Heiman D, Young S, Zeng Q, Abouelleil A, Aftuck L, Bessette D, Brown A, FitzGerald M, Lui A, Macdonald JP, Priest M, Orbach MJ, Galgiani JN, Kirkland TN, Cole GT, Birren BW, Henn MR, Taylor JW, Rounsley SD. 2010. Population genomic sequencing of *Coccidioides* fungi reveals recent hybridization and transposon control. Genome Res 20:938–946. https://doi.org/10.1101/gr.103911.109.

41. Hwang L, Hocking-Murray D, Bahrami AK, Andersson M, Rine J, Sil A. 2003. Identifying phase-specific genes in the fungal pathogen *Histoplasma capsulatum* using a genomic shotgun microarray. Mol Biol Cell 14:2314–2326. https://doi.org/10.1091/mbc.e03-01-0027.

42. Nunes LR, Costa de Oliveira R, Leite DB, da Silva VS, dos Reis Marques E, da Silva Ferreira ME, Ribeiro DCD, de Souza Bernardes LA, Goldman MHS, Puccia R, Travassos LR, Batista WL, Nóbrega MP, Nobrega FG, Yang D-Y, de Bragança Pereira CA, Goldman GH. 2005. Transcriptome analysis of *Paracoccidioides brasiliensis* cells undergoing mycelium-to-yeast transition. Eukaryot Cell 4:2115–2128. https://doi.org/10.1128/EC.4.12.2115-2128.2005.

43. Boyce KJ, McLauchlan A, Schreider L, Andrianopoulos A. 2015. Intracellular growth is dependent on tyrosine catabolism in the dimorphic fungal pathogen *Penicillium marneffei*. PLoS Pathog 11:e1004790. https://doi.org/10.1371/journal.ppat.1004790.

44. Worsham PL, Goldman WE. 1988. Quantitative plating of *Histoplasma capsulatum* without addition of conditioned medium or siderophores. J Med Vet Mycol 26:137–143. https://doi.org/10.1080/02681218880000211.

45. Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey TM, Young G, Fennell TJ, Allen A, Ambrogio L, Berlin AM, Blumenstiel B, Cibulskis K, Friedrich D, Johnson R, Juhn F, Reilly B, Shammas R, Stalker J, Sykes SM, Thompson J, Walsh J, Zimmer A, Zwirko Z, Gabriel S, Nicol R, Nusbaum C. 2011. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biol 12:R1. https://doi.org/10.1186/gb-2011-12-1-r1.

46. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

47. Li H. 2016. Minimap and miniasm: fast mapping and *de novo* assembly for noisy long sequences. Bioinformatics 32:2103–2110. https://doi.org/10.1093/bioinformatics/btw152.

48. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate *de novo* genome assembly from long uncorrected reads. Genome Res 27:737–746. https://doi.org/10.1101/gr.214270.116.

49. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. https://doi.org/10.1371/journal.pone.0112963.

50. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2018. Assembly of long error-prone reads using repeat graphs. bioRxiv https://doi.org/10.1101/247148.

51. Ruan J, Li H. 2020. Fast and accurate long-read assembly with wtdbg2. Nat Methods 17:155–158. https://doi.org/10.1038/s41592-019-0669-3.

52. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.

53. Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. Methods Mol Biol 1962:227–245.