Genome Diversity, Recombination, and Virulence across the Major Lineages of Paracoccidioides

José F. Muñoz,a,b,c @ Rhys A. Farrer,c Christopher A. Desjardins,c Juan E. Gallo,a,d Sean Sykes,c Sharadha Sakthikumar,c Elizabeth Misas,a,b Emily A. Whiston,e Eduardo Bagagli,f Celia M. A. Soares,g Marcus de M. Teixeira,h,i John W. Taylor,e Oliver K. Clay,a,j Juan G. McEwen,a,k Christina A. Cuomo,c

Cellular and Molecular Biology Unit, Corporación para Investigaciones Biológicas, Medellín, Colombiaa; Institute of Biology, Universidad de Antioquia, Medellín, Colombiab; Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA;c; Doctoral Program in Biomedical Sciences, Universidad del Rosario, Bogotá, Colombia;c; Department of Plant and Microbial Biology, University of California, Berkeley, California, USA; Instituto de Biociências, Universidade Estadual Paulista, Botucatu, São Paulo, Brazil;f; Laboratório de Biologia Molecular, Instituto de Ciências Biológicas, ICBI, Goiânia, Brazil;g; Instituto de Ciências Biológicas, Universidade de Brasília, Brasília, Distrito Federal, Brazil;h; Division of Pathogen Genomics, Translational Genomics Research Institute North, Flagstaff, Arizona, USA;i; School of Medicine and Health Sciences, Universidad del Rosario, Bogotá, Colombia; School of Medicine, Universidad de Antioquia, Medellín, Colombiaa

ABSTRACT The Paracoccidioides genus includes two species of thermally dimorphic fungi that cause paracoccidioidomycosis, a neglected health-threatening human systemic mycosis endemic to Latin America. To examine the genome evolution and the diversity of Paracoccidioides spp., we conducted whole-genome sequencing of 31 isolates representing the phylogenetic, geographic, and ecological breadth of the genus. These samples included clinical, environmental and laboratory reference strains of the S1, PS2, PS3, and PS4 lineages of P. brasiliensis and also isolates of Paracoccidioides lutzii species. We completed the first annotated genome assemblies for the PS3 and PS4 lineages and found that gene order was highly conserved across the major lineages, with only a few chromosomal rearrangements. Comparing whole-genome assemblies of the major lineages with single-nucleotide polymorphisms (SNPs) predicted from the remaining 26 isolates, we identified a deep split of the S1 lineage into two clades we named S1a and S1b. We found evidence for greater genetic exchange between the S1b lineage and all other lineages; this may reflect the broad geographic range of S1b, which is often sympatric with the remaining, largely geographically isolated lineages. In addition, we found evidence of positive selection for the GP43 and PGA1 antigen genes and genes coding for other secreted proteins and proteases and lineage-specific loss-of-function mutations in cell wall and protease genes; these together may contribute to virulence and host immune response variation among natural isolates of Paracoccidioides spp. These insights into the recent evolutionary events highlight important differences between the lineages that could impact the distribution, pathogenicity, and ecology of Paracoccidioides.

IMPORTANCE Characterization of genetic differences between lineages of the dimorphic human-pathogenic fungus Paracoccidioides can identify changes linked to important phenotypes and guide the development of new diagnostics and treatments. In this article, we compared genomes of 31 diverse isolates representing the major lineages of Paracoccidioides spp. and completed the first annotated genome sequences for the PS3 and PS4 lineages. We analyzed the population structure and characterized the genetic diversity among the lineages of Paracoccidioides, including a deep split of S1 into two lineages (S1a and S1b), and differentiated S1b, associated
with most clinical cases, as the more highly recombining and diverse lineage. In addition, we found patterns of positive selection in surface proteins and secreted enzymes among the lineages, suggesting diversifying mechanisms of pathogenicity and adaptation across this species complex. These genetic differences suggest associations with the geographic range, pathogenicity, and ecological niches of *Paracoccidioides* lineages.

**KEYWORDS:** *Paracoccidioides*, evolution, genetic recombination, genome analysis, mycology, population genetics

*Paracoccidioides* spp. are the cause of paracoccidioidomycosis (PCM), a systemic mycosis that mainly affects people in Latin America. In this region where PCM is endemic, PCM has an estimated incidence of 1 to 3 cases per 100,000 inhabitants (1, 2). The vast majority of PCM cases (roughly 80%) occur in Brazil, while Colombia and Venezuela have the next highest numbers of infections (3). *Paracoccidioides* is a thermally dimorphic fungus closely related to *Histoplasma* and *Blastomyces*, which cause similar infections worldwide or predominantly in regions of North America, respectively.

Multilocus sequencing studies elucidated species boundaries within the *Paracoccidioides* genus and supported the existence of two distinct species, *P. brasiliensis* and *P. lutzii* (4). *P. lutzii* is a single monophyletic and recombining population found to date in central, southwest, and north Brazil and Ecuador (4). *P. brasiliensis* is monophyletic and is comprised of distinct lineages classified as S1, PS2, PS3, and PS4 (4–6). The S1 lineage is associated with the majority of PCM cases and is widely distributed in South America (4–6). PS2 has been identified to date only in Brazil and Venezuela, whereas PS3 is mainly found in regions of endemcity in Colombia (4, 5). Recently, a novel lineage, PS4, was described from a region of Venezuela (6). Evidence of recombination was noted for *P. brasiliensis* S1 and *P. lutzii*, but not other lineages, based on a small number of genomic loci (4, 5).

Isolates from each of these phylogenetic lineages of *Paracoccidioides* can infect humans; however, different lineages can vary in virulence and culture adaptation and can induce different immune responses by the host (7, 8). One feature that is correlated with the differential rates of infection is variation in the number of infective conidia. For example, isolates from S1 produce many more conidia than PS2 isolates, which could be related to the disproportional 9:1 rate of S1 to PS2 infection in both human and armadillo isolates (8). In addition to interspecific variation between lineages and between species, *Paracoccidioides* isolates have been shown to contain extensive intraspecific genetic variability between strains of the same lineage (9–11).

To enable genome-based studies of this medically important fungus, isolates of *P. brasiliensis* S1 and PS2 and *P. lutzii* were previously sequenced and compared to related dimorphic and nondimorphic fungi (12). Notably, *Paracoccidioides* and related dimorphic pathogens have a reduced number of genes involved in carbohydrate metabolism, protein metabolism, and synthesis of secondary metabolites (12), an observation that allows new insights into the differences between these related fungi and their physiological potential for pathogenicity. Recently, the genome assemblies and gene annotations of those reference strains were improved using Illumina resequencing, increasing the overall accuracy of assembly bases and gene structures (13). These improved reference genomes of *Paracoccidioides* spp. provide an opportunity to map the population structure and examine variation with finer resolution.

In this study, we used genome sequences of 31 isolates for a comprehensive comparison of gene conservation, genetic diversity, and genome evolution across the major lineages of *Paracoccidioides*. The panel of isolates sequenced in this study included clinical isolates from acute and chronic PCM cases and environmental isolates from soil or two species of armadillos (*Dasypus novemcinctus* and *Cabassous centralis*). We assembled the first reference genomes for the PS3 and PS4 lineages: compared to the previously assembled references from other lineages, gene content and order are...
highly conserved, with few rearrangements. We characterized genetic diversity among the lineages, as well as lineage-specific evolutionary patterns within the *Paracoccidioides* genus and found evidence of recombination and ancestral hybridization patterns between some of the lineages. Additionally, we identified genomic regions or genes that are highly diverse within or between lineages; these include genes with potential roles in virulence. We found that genes with the strongest evidence of positive selection include the GP43 antigen gene and genes coding for other secreted proteins and proteases (e.g., *PGA1*, *CBP1*, *SOD3*, and *ENG1*), as well as loss-of-function mutation in genes that are specific to some lineages. Our analyses provide insight into the recent evolutionary events highlighting genetic differences between the lineages that could impact the distribution, pathogenicity, and ecology of *Paracoccidioides*. These potential virulence factors and genetic differences at the population level will be important for future studies to compare with the infectious potency of *Paracoccidioides* clinical isolates and clinical symptoms attributable to paracoccidioidomycosis.

**RESULTS**

**Conserved genome organization across *Paracoccidioides spp.*** To compare genome structures and gene contents between the major *Paracoccidioides* lineages, we sequenced, assembled, and annotated a representative isolate from the PS3 lineage (strain PbCnh) and from the PS4 lineages (strain Pb300) (see Fig. S1A and Text S1 in the supplemental material). The 29.4-Mb assemblies of both strains are intermediate in size between those of the 29.95-Mb assembly of Pb18 (S1b) and the 29.06-Mb assembly of Pb03 (PS2) (see Fig. S1B) (13). The gene annotation of strains PbCnh and Pb300 resulted in 8,324 and 8,070 predicted protein-coding genes, respectively. High representation of core eukaryotic genes provides evidence that those genomes are nearly complete; 96 to 98% of these conserved genes are found in all assemblies (see Fig. S1C). Predicted gene contents were highly similar across all four *P. brasiliensis* genomes, comparing the new assemblies to the previously sequenced genomes (see Fig. S1B), which suggests that the gene content is very consistent across the lineages.

The *Paracoccidioides* genomes of both species and all lineages are highly conserved in terms of whole-genome sequence similarity and gene synteny (see Text S1 and Fig. S1D). The genomes of *P. brasiliensis* share an average of 98.5% identity, whereas the genome of the more distant species *P. lutzii* shares an average of 94.8% with *P. brasiliensis*. The genomes of PbCnh (PS3) and Pb300 (PS4) share the highest percentage aligned (98.9%), which correlates with the phylogenetic and population structure relationships (see below). We identified syntenic regions of conserved gene order and found an average of 6,907 genes within syntenic blocks among the *Paracoccidioides* lineages (see Fig. S1D to F). The percentage of genes in syntenic blocks ranged from 75.3% (interspecies, *P. brasiliensis* versus *P. lutzii*) to 89.9% (intrasppecies, *P. brasiliensis* versus *P. brasiliensis*). In contrast, the dimorphic fungus Blastomyces has only ~69% genes in syntenic blocks due to the presence of isochores-like structures of repeat-rich GC-poor and GC-rich blocks rarely observed in *Paracoccidioides* (14).

While the *Paracoccidioides* genomes were largely colinear, a few chromosomal rearrangements were detected. A large rearrangement was detected between *P. brasiliensis* S1/PS3 and *P. lutzii*/*P. brasiliensis* PS2 in chromosome 4, where the regions at the beginning and the end of the supercontig 4 are inverted, and the gene order across the middle of the supercontig is conserved (see Fig. S1F). A large chromosomal rearrangement was also detected between *P. brasiliensis* and *P. lutzii*, where syntenic blocks in *P. brasiliensis* chromosomes 3 and 5 (Pb18, supercontigs 2, 13, and 10) are combined into a single supercontig in *P. lutzii* (see Fig. S1F). We found no evidence of assembly errors across the junctions of these rearrangements based on even coverage of aligned reads across these regions. In addition, we called structural variants based on the read alignments to the Pb18 assembly, and recovered each of the rearrangements present in the assemblies (see Text S1 and Data Set S1 in the supplemental material). Chromosomal rearrangements may impact the capacity for genetic exchange, as some crossover events will generate missing chromosomal regions or other aneu-
ploidies and nonviable progeny. The rearrangement between PS2 and the other lineages of \textit{P. brasiliensis} could potentially prevent genetic exchange between these groups.

In addition to chromosomal rearrangements, we looked for evidence of copy number and ploidy variation. We calculated the normalized read alignment density in 10-kb nonoverlapping windows (see Materials and Methods) and examined the variation across the \textit{Pb18} chromosomes. The alignment density showed no large regions of higher sequencing depth, supporting that \textit{Paracoccidioides} species do not maintain aneuploid chromosomes or segments (see Fig. S2 in the supplemental material) such as those found in other fungi.

\textbf{Phylogenetic for S1 group split and allele sharing between lineages.} To examine the \textit{Paracoccidioides} phylogenetic relationships, we identified polymorphisms across an extended panel of 31 isolates (Table 1). Using 614,570 positions (single-nucleotide polymorphisms [SNPs]) in the sequenced isolates, maximum likelihood and Bayesian phylogenies were constructed to examine intralineage relationships (Materials and Methods). Both maximum likelihood and Bayesian analyses highly supported that \textit{P. brasiliensis} isolates are clustered into five distinct lineages (Fig. 1; see Fig. S3 in the supplemental material). Of the four previously identified lineages, S1 is the most highly variable, with two distinct clades we denoted as the S1a and S1b lineages. The S1b lineage includes the reference strain Pb18, along with three clinical isolates from Brazil and Argentina. The S1a lineage includes clinical and environmental isolates and was split into two subclades: one includes clinical isolates from Argentina and central-west regions of Brazil, while the second includes three clinical and four armadillo isolates.

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|c|}
\hline
\textbf{Isolate ID} & \textbf{Other name(s)} & \textbf{Origin} & \textbf{Source} & \textbf{Lineage} & \textbf{Provider} & \textbf{Reference} \\
\hline
\textit{P. brasiliensis} & & & & & & \\
\textbf{Pb18}\textsuperscript{a} & B17 & Sao Paulo, Brazil & Chronic PCM\textsuperscript{b} & S1b & R. Puccia & 44 \\
\textbf{PbCaz} & Cazon; A1 & Chaco, Argentina & Acute PCM & S1b & R. Negroni & 45 \\
\textbf{Pb113} & & Manaus-AM, Brazil & PCM & S1b & C. de Almeida Soares & 46 \\
\textbf{PbBlo} & & Brazil & PCM & S1b & C. de Almeida Soares & This study \\
\textbf{MS1} & & Mato Grosso do Sul, Brazil & PCM & S1a & M. Sueli Felipe & 47 \\
\textbf{DO3} & & Piracicaba, SP, Brazil & PCM & S1a & E. Bagagli & This study \\
\textbf{MS2} & & Mato Grosso do Sul, Brazil & PCM & S1a & M. Sueli Felipe & 47 \\
\textbf{Pb1445} & A5 & Argentina & Chronic PCM & S1a & R. Negroni & 5 \\
\textbf{Pb337} & T15L1; B10 & Brazil & \textit{D. novemcinctus} & S1a & E. Bagagli & 48 \\
\textbf{Pb66} & & Brazil & PCM & S1a & C. de Almeida Soares & 49 \\
\textbf{PbBer} & & Bercelli; A3 & Argentina & PCM & S1a & R. Negroni & 5 \\
\textbf{D02} & & Laranjal Paulista, SP, Brazil & PCM & S1a & E. Bagagli & This study \\
\textbf{T1F1} & B1 & Pratanea, SP, Brazil & \textit{D. novemcinctus} & S1a & E. Bagagli & 48 \\
\textbf{T1S1} & & Botucatu, Brazil & \textit{D. novemcinctus} & S1a & E. Bagagli & This study \\
\textbf{T16B1} & & Brazil & \textit{D. novemcinctus} & S1a & E. Bagagli & This study \\
\textbf{Pb300}\textsuperscript{a} & V1 & Miranda, Venezuela & Soil & PS4 & M. B. Albornoz & 50 \\
\textbf{EPM83} & & Bogotá, Colombia & Chronic PCM & PS3 & A. Restrepo & 49 \\
\textbf{Pb339} & B18 & Sao Paulo, Brazil & PS3 & A. Restrepo & 51 \\
\textbf{Pb60855} & C4 & Antioquia, Colombia & Chronic PCM & PS3 & A. Restrepo & 52 \\
\textbf{PbBac} & & Colombia & Chronic PCM & PS3 & A. Restrepo & This study \\
\textbf{PbCab} & P196; C6 & Caldas, Colombia & \textit{C. centralis} & PS3 & A. Restrepo & 53 \\
\textbf{PbCnh}\textsuperscript{a} & & Colombia & Chronic PCM & PS3 & A. Restrepo & This study \\
\textbf{PbJam} & & Colombia & Chronic PCM & PS3 & A. Restrepo & This study \\
\textbf{Pb02} & V2 & Caracas, Venezuela & Chronic PCM & PS2 & R. Puccia & 54 \\
\textbf{Pb03}\textsuperscript{a} & B26 & Sao Paulo, Brazil & Chronic PCM & PS2 & R. Puccia & 54 \\
\textbf{Pb262} & & Uberlândia, MG, Brazil & Dog food & PS2 & Z. Pires de Camargo & 55 \\
\textbf{T10B1} & B7 & Botucatu, Brazil & \textit{D. novemcinctus} & PS2 & E. Bagagli & 5 \\
\hline
\textit{P. lutzii} & & & & & & \\
\textbf{Pb01}\textsuperscript{a} & & Goiás, Brazil & PCM & \textit{P. lutzii} & R. Puccia & 56 \\
\textbf{PI1578} & & Goiás, Brazil & PCM & \textit{P. lutzii} & C. de Almeida Soares & 47 \\
\textbf{ED01} & & Goiás, Brazil & PCM & \textit{P. lutzii} & C. de Almeida Soares & 47 \\
\textbf{PIEE} & EE & Mato Grosso, Brazil & PCM & \textit{P. lutzii} & M. Sueli Felipe & 4 \\
\hline
\textsuperscript{a}Reference strain assembled and annotated genome. \\
\textsuperscript{b}PCM, paracoccidioidomycosis.
from southeast Brazil. This phylogenetic analysis also supports that PS3 is a monophyletic group with very limited diversity of mostly isolates from Colombia, including both chronic PCM isolates and one isolate from armadillo. While previous phylogenies using short loci had suggested that PS3 only includes isolates from Colombia (4, 5, 8), we found that the Pb339 isolate from southeast Brazil was placed in this lineage (Fig. 1), suggesting that this lineage may be more widespread than previously described. In addition, our phylogenetic analysis provides strong evidence for the separation of the PS4 lineage, as recently proposed (6), sharing a common ancestor with PS3. This was also supported by population structure analyses (see below). To confirm that the phylogeny was not influenced by the use of a reference genome from one lineage, we identified SNPs from reads aligned to reference genomes representing each lineage (Pb03, PbCnh, Pb300, and Pb01) and found the same topology in phylogenies of each set (see Fig. S3).

To better understand the evolutionary history of *Paracoccidioides*, we estimated the divergence and most common ancestor dates (time to most recent common ancestor [TMRCA]). This analysis suggests that *P. lutzii* and *P. brasiliensis* diverged about 22.5 million years ago (MYA). The *P. brasiliensis* lineages separated more recently at
1.47 MYA (Fig. 1). Within *P. brasiliensis* lineages, the shortest time of divergence was found for PS3 (TMRCA, 23,000 years ago [23 KYA]) and the longest was found in S1b (TMRCA, 575 KYA), with PS2 and S1a showing intermediate values (Fig. 1; see Fig. S4 in the supplemental material). We estimated that *P. lutzii* diverged slightly earlier than the *P. brasiliensis* groups, around 833 KYA. Using whole-genome SNPs, the separation of *P. brasiliensis* and *P. lutzii* is very similar to a previously reported estimate (8); however, we estimate more recent separation within each lineage.

While the SNP data strongly supported a single tree, we also examined the relationship of the sequenced isolates using a network approach to look for evidence of alternative topologies. The NeighborNet algorithm identified five major groups within the *P. brasiliensis* species, including the clear separation of S1 into the S1a and S1b lineages. In addition, the network suggests some level of ancestral recombination between the groups as well as more recent recombination involving the S1b lineage (Fig. 1B). We tested for evidence of phylogenetic heterogeneity using the pairwise homoplasy index test (15) and found statistically significant support for recombination (*P* = 4.3 e - 13).

Next, we compared the distribution of SNPs across the genome and examined how sites were shared between lineages. We found a similar pattern of polymorphism frequency across the genome for isolates of the same lineage; differences between lineages include two distinct patterns for the related phylogenetic lineages S1a and S1b, supporting this subdivision (see Fig. S2 in the supplemental material). To more finely compare variant sites between lineages, we classified SNP alleles based on populations as fixed, shared, or private based on pairwise comparisons between lineages (S1a, S1b, PS2, PS3, PS4, and *P. lutzii*). S1b has the highest shared SNP allele component, ranging from 12.4% to 21.1% compared with the S1a, PS2, PS3, and PS4 lineages and 2.5% compared with *P. lutzii* (Fig. 2). In contrast, the next highest shared frequency is for S1a, which ranged from 0.17% to 1.1% compared with PS2, PS3, PS4, and *P. lutzii*. To determine whether the large fraction of shared alleles was due to recombination rather than just a higher genetic diversity in S1b, we combined SNPs for PS3 and PS4 to make an artificial lineage with diversity relatively equivalent to that of S1b. In this combined set, we did not observe the fraction of shared alleles increase from the values of PS3 and PS4 alone, suggesting the effect in S1b is not due to high diversity alone. While recombination with other lineages is also supported by the network tree analysis, it is unclear whether this recombination was relatively recent or ancestral.

**Recombination and hybridization between the major lineages of *Paracoccidioides*.** To further examine the *Paracoccidioides* lineages for evidence of recomb-
nation, we performed a principal-component analysis (PCA) of the SNP data. For all *Paracoccidioides* isolates (*n* = 31), we found clear separation of the two species; PC1 cleanly separates *P. brasiliensis* and *P. lutzii* isolates as two distinct species. Within *P. brasiliensis*, PC2 separates PS2 from all other groups and divides S1 into two distinct groups (S1a and S1b lineages) (Fig. 3A). In this comparison, the S1b, PS3, and PS4 lineages appeared more closely related, suggesting less genetic divergence between these lineages. Comparing only the *P. brasiliensis* isolates (*n* = 26), PC2 separates the S1b, PS4, and PS3 lineages, where PS4 is located between S1b and PS3 (Fig. 3A). This analysis supports the major subdivisions in the *P. brasiliensis* population and suggests that some lineages have similar relationships to multiple other lineages, appearing more centrally in PCA plots, suggesting the impact of recombination.

To examine the relationships of these groups, we predicted the population ancestry for each strain using a model-based clustering algorithm implemented by the Structure software (16). We identified populations within *Paracoccidioides* spp. using (i) all isolates and (ii) only *P. brasiliensis* isolates. There is clear separation between *P. brasiliensis* and *P. lutzii*, with only two major clusters of ancestry (Fig. 3B). When only *P. brasiliensis* isolates are examined, four distinct ancestry clusters are most highly supported (see Materials and Methods), corresponding to S1a, PS2, PS3, and partially S1b. Evaluation

---

**FIG 3** Genetic population structure of *Paracoccidioides*. (A) PCA of genetic variants in *Paracoccidioides* is shown as a two-dimensional plot for all isolates (*n* = 31) (left) or for all *P. brasiliensis* isolates (*n* = 26) (right). In each plot, circles indicate isolates and colors indicate the lineage. (B) Population structure of *Paracoccidioides* spp. (left) and *P. brasiliensis* (right) inferred from 476,589 and 339,966 SNPs, respectively, using the Structure software program with different k values of 2 and 4, respectively. An admixture model with correlated allele frequencies and site-by-site analysis was used. Each isolate is represented by a single vertical line broken into k-colored segments, with lengths proportional to each of the k inferred clusters. (C) Whole-genome plot for Pb300 (PS4) deducted from Structure site-by-site analysis showing the chromosomal ancestry.
of different values of $k$ showed overall support for these major subdivisions; however, *P. brasiliensis* could also be inferred to have 3 primary clusters of S1a, PS2, and PS3 (see Materials and Methods and Fig. S5 in the supplemental material). In both the 3- and 4-cluster analyses, isolates from S1b and PS4 have a subset of sites found in different clusters; there is a unique set of S1b/PS4 sites in the 4-cluster analysis, similar to the private alleles found for S1b (Fig. 2). These patterns support a hybrid ancestry for the S1b and PS4 *P. brasiliensis* isolates, which share SNP markers in different proportions with the S1a and PS3 groups (Fig. 3B). Plotting SNPs colored by ancestry across the genome for the PS4 isolate Pb300 revealed a highly intermixed pattern of small blocks of S1b and PS3 ancestry, suggesting a relatively ancient hybridization event (Fig. 3C).

**Equal frequencies of mating types in each lineage.** Although the sexual phase of *Paracoccidioides* had not been completely characterized in nature or in the laboratory, the patterns of recombination and ancestral hybridization we found may be most parsimoniously explained by sexual reproduction. We identified the mating type of each isolate based on the genomic sequenced data. While the previously sequenced reference genomes Pb18 (S1b) and Pb03 (PS2) contain the mating type HMG (*MAT1-2*), here we generated the first assembled genome of mating type /H9251 (*MAT1-1*) for *P. brasiliensis* (strain PbCnh [PS3]). By aligning the assemblies, we observed that there are not any chromosomal rearrangements near the mating locus that could impact the capacity for interlineage genetic exchange (see Fig. S1F in the supplemental material). The assembly of Pb300 (PS4) also contains the mating type HMG (*MAT1-2*). As previously noted, conservation of mating- and meiosis-specific genes suggests that *P. brasiliensis* has the necessary machinery for sexual reproduction (12), and we see no additional gene loss in these lineages. All of the 31 sequenced *Paracoccidioides* isolates were heterothallic with a population ratio of 1:1 of each mating type, both in the population as a whole as well as among each lineage, with a total of 15 isolates /H9251 (*MAT1-1*) and 16 isolates HMG (*MAT1-2*) (see Fig. S6 in the supplemental material). The evidence for genetic exchange and recombination and these roughly equal numbers of both mating types in each lineage support the potential for sexual reproduction in *Paracoccidioides*.

**Genome-wide population genetic variation between the major lineages of Paracoccidioides.** Both polymorphism and phylogenetic analyses suggest that PS3 is a monophyletic group derived from a shared common ancestor with limited diversity, and S1b is the most variable lineage, reflecting a more widespread geographic distribution. We found the highest level of nucleotide diversity in S1b ($\pi = 0.00256$), which was 30-fold greater than the very low nucleotide diversity found in PS3 ($\pi = 0.00008$). Nucleotide diversities in S1a, PS2, and *P. lutzii* were intermediate: 0.00053, 0.00066, and 0.00015, respectively (Fig. 4A). In addition, we tested for genome-wide allele frequency distribution using Tajima’s $D$ to scan for signatures of demography and selection. We found that all populations showed genome-wide Tajima’s $D$ values not significantly different from the null expectation. This suggests that each lineage is evolving neutrally at the whole-genome level (Fig. 4A).

Additionally, we analyzed interspecific divergence between *P. brasiliensis* and *P. lutzii* and between *P. brasiliensis* lineages (S1a, S1b, PS2, and PS3) across the genome using Wright’s $F_{ST}$ statistic and a 10-kb sliding-window approach. Across the *P. brasiliensis* pairwise comparisons, values ranged between 0.70 and 0.94, suggesting limited genetic exchange between the lineages (Fig. 4C; see Fig. S7 in the supplemental material). Furthermore, there were no regions of very low $F_{ST}$ that would indicate recent genetic exchange (see Fig. S7). Comparisons of *P. lutzii* and *P. brasiliensis* lineages had values ranging from 0.97 to 0.99, suggesting strong local divergence between these species, similar to the 0.95 $F_{ST}$ value observed for most regions of *Coccidioides immitis* compared to *Coccidioides posadasii* (17). S1b had the lowest pairwise $F_{ST}$ values (0.70 to 0.76) compared to the remaining lineages (0.91 to 0.94) (Fig. 4B; see Fig. S7). This supports the hypothesis of recombination between S1b and the other lineages, which correlates with the intermediate position of S1b in the population structure analysis and higher degree of shared alleles (Fig. 3 and 2, respectively). This may have clinical implications
since the S1b lineage is widely distributed in South America, includes highly virulent strains, and has been associated with the vast majority of cases of PCM (Fig. 1).

**Lineage-specific gain and loss and rapid evolution of virulence-associated genes.** Comparing the *P. brasiliensis* lineages, a total of 6,670 core ortholog clusters had representative genes from all five reference genomes. We found 720 ortholog groups in at least two strains and 459 ortholog groups that were present in all *P. brasiliensis* strains but absent in *P. lutzii*. We did not find any significant enrichment of functional categories among *P. lutzii* and *P. brasiliensis* strains or among their lineages, which suggests that the phenotypic differences between *P. lutzii* and *P. brasiliensis* and between the lineages are not due to large protein family expansion or contraction. However, unique genes (those without orthologs in other lineages) could contribute to different phenotypic differences observed in each lineage, as well as genes duplicated in only one species. Among the unique genes we identified, there were genes coding for several protein kinases, proteases, transcription factors, and transporters (see Data Set S1 in the supplemental material). More specifically, in *P. lutzii* there were several unique genes coding for anhydrolases, glycosyl hydrolases, peptidases (M24 and C12), and methyltransferases, while in *P. brasiliensis* there were several unique genes coding for actin, transporters, aspartyl proteases, peptidases (M16 and M28), and transcription factors. These unique proteins may provide a more diverse

---

**FIG 4** Genome-wide nucleotide diversity ($\pi/\theta$), Tajima's $D$ (TD), and population divergence analysis ($F_{ST}$) in *Paracoccidioides*. (A) Genome-wide average of nucleotide diversity ($\pi$) and TD for sliding windows of 10-kb regions within the main *Paracoccidioides* lineages. (B) Average of genome-wide (10-kb windows) variation in ($\theta$), Weir’s formulation of Wright’s fixation index ($F_{ST}$), for pairwise comparisons in each lineage. (C) Distribution of the average nucleotide diversity ($\pi$), TD, and $F_{ST}$ for sliding windows of 10-kb regions for the S1a, S1b, PS3, and PS3 *Paracoccidioides* lineages.
functional repertoire to each \textit{P. lutzii} or \textit{P. brasiliensis} lineage, enabling the fungus different mechanisms and strategies to produce infection and disease.

In addition to identifying strain-specific genes, we identified nonsense mutations that were specific to each lineage. For example, the serine carboxypeptidase gene (\textit{CPDS}; \textit{PADG} \_07980), the predicted mannan endo-1,6-\(\alpha\)-mannosidase gene (\textit{PADG} \_00193), and the predicted transmembrane gene (\textit{PADG} \_08161) have nonsense mutations in all lineages of \textit{P. brasiliensis} (S1a, S1b, PS2, PS3, and PS4) but not in \textit{P. lutzii}, suggesting these genes are not functional in \textit{P. brasiliensis}. Other genes that do not have nonsense mutations in \textit{P. brasiliensis} (S1a, S1b, PS2, PS3, and PS4) but have nonsense mutations in all \textit{P. lutzii} strains included genes coding for the separase protease (\textit{peptidase \_C50}; \textit{PADG} \_07698), a DNA-binding protein protease (\textit{peptidase \_M24}; \textit{PADG} \_01032), other secreted proteins (e.g., \textit{PADG} \_00436), and other genes coding for proteins involved in different cellular processes (e.g., sugar transporter, \textit{PADG} \_04202; sterigmatocystin biosynthesis monoxygenase \textit{StcW}, \textit{PADG} \_04703; and a glucose-6-phosphate isomerase, \textit{PADG} \_00451) (see Data Set S1). Differences in gene content between the species, including genes coding for predicted proteases and cell wall genes, could impact how each species/lineage interacts with the host or environment, although other genes with overlapping functions could compensate for loss of these genes.

To more finely examine gene sequence differences and the impact of selection, we calculated the ratio of nonsynonymous to synonymous evolutionary changes \((dN/dS)\) for each gene across the lineages of \textit{Paracoccidioides}. An average of 485 genes were found to be under positive selection \((dN/dS \geq 1)\) in each of the lineages (Fig. 5; see Data Set S1). The set of genes evolving under positive selection includes the surface antigen gene \textit{GP43} (\textit{PADG} \_07615), the superoxide dismutase gene \textit{SOD3} (\textit{PADG} \_02842), the alternative oxidase gene \textit{AOX} (\textit{PADG} \_03747), and the thioredoxin gene (\textit{PADG} \_05504), virulence-associated genes of central importance in \textit{Paracoccidioides} and other dimorphic fungi (Table 2). Other notable genes under positive selection included several protease genes, including a glutamate carboxypeptidase gene (\textit{PADG} \_00686), a multifunctional tryptophan biosynthesis protein gene (\textit{PADG} \_07274), the \(\alpha\)-pheromone processing metallopeptidase gene \textit{STE23} (\textit{PADG} \_07053), and the subtilase-type proteinase gene \textit{PSP3} (\textit{PADG} \_07422). In addition, many genes coding for secreted proteins appear under positive selection, including the calcium binding protein gene \textit{CBP1} (\textit{PADG} \_02399) and the mannan endo-1,6-\(\alpha\)-mannosidase gene \textit{DCW1} (\textit{PADG} \_01494). Among the 88 total secreted proteins under positive selection, seven were found to be significantly differentially upregulated during a mouse model of infection in \textit{Blastomyces} (14), including a
TABLE 2 Selection of candidate virulence factors in *Paracoccidioides* isolates found in highly diverse regions and/or under selection

| Locus ID | Description of protein | π | TD | SNPs | dN/dS >1 | FST |
|----------|------------------------|---|----|-----|----------|-----|
| PADG_02498 | 3-Hydroxyanthranilate 3,4-dioxygenase | + |  |  |  |   |
| PADG_00940 | Acetate kinase; *B. dermatitidis* ortholog induced during macrophage interaction | − | + | − | + |   |
| PADG_01835 | Aldehyde reductase; *B. dermatitidis* ortholog induced during macrophage interaction | − | − | − | + |   |
| PADG_07461 | α-1,3-Glucanase | + | + | + | + |   |
| PADG_03747 | Alternative oxidase; AOX gene | − | − | − | + |   |
| PADG_02460 | Antigenic GPI-protein; secreted; antigen; PGA1 gene | + | + | + | + |   |
| PADG_04167 | Aspartyl aminopeptidase; peptidase family M18 | + | + | + | + |   |
| PADG_02521 | Antigenic GPI-protein; vesicles | − | − | − | + |   |
| PADG_12450 | Antigenic GPI-protein; vesicles | − | − | − | + |   |
| PADG_00948 | Oxidoreductase; *B. dermatitidis* ortholog induced during in vivo infection | − | − | − | + |   |
| PADG_01238 | Hypothetical protein; *B. dermatitidis* ortholog induced during macrophage interaction | − | − | − | + |   |
| PADG_01283 | Hypothetical protein; *B. dermatitidis* ortholog induced during macrophage interaction | − | − | − | + |   |
| PADG_03908 | Hypothetical protein; *B. dermatitidis* ortholog induced during macrophage interaction | − | − | − | + |   |
| PADG_07534 | Hypothetical protein; *B. dermatitidis* ortholog induced during macrophage interaction | − | − | − | + |   |
| PADG_11963 | Hypothetical protein; *B. dermatitidis* ortholog induced during macrophage interaction | − | − | − | + |   |
| PADG_02535 | Hypothetical protein; secreted; vesicles; *B. dermatitidis* ortholog induced during *in vivo* infection | + | + | + | + |   |
| PADG_02542 | Hypothetical protein; vesicles | − | − | − | + |   |
| PADG_12450 | Hypothetical protein; vesicles | + | − | + | + |   |
| PADG_04741 | Hypothetical protein; vesicles | + | + | − | + |   |
| PADG_02521 | Hypothetical protein; vesicles | − | − | − | + |   |
| PADG_12101 | Hypothetical protein; vesicles | − | − | − | + |   |
| PADG_01494 | Mannan endo-1,6-α-mannosidase; secreted; DCW1 gene | − | − | − | + |   |
| PADG_00948 | Oxidoreductase; *B. dermatitidis* ortholog induced during *in vivo* infection | − | − | − | + |   |
| PADG_07460 | Predicted aminopeptidase; peptidase family M18; induced during macrophage interaction | + | + | + | + |   |
| PADG_05820 | Predicted aminopeptidase; peptidase family M24; induced during macrophage interaction | − | − | − | + |   |
| PADG_07369 | Predicted dehydrogenase | + | + | + | + |   |
| PADG_02527 | Predicted dehydrogenase | + | − | − | + |   |
| PADG_02562 | Predicted dehydrogenase | + | + | − | + |   |
| PADG_02492 | Predicted dehydrogenase | + | + | − | + |   |
| PADG_07365 | Predicted dehydrogenase | + | + | − | + |   |
| PADG_07411 | Predicted dehydrogenase | + | + | − | + |   |
| PADG_02575 | Predicted nonribosomal peptide synthetase | + | + | − | + |   |
| PADG_06309 | Predicted oxidoreductase | + | + | + | + |   |
| PADG_02592 | Predicted oxidoreductase | + | + | + | + |   |
| PADG_06322 | Predicted peroxidase | + | + | − | + |   |
| PADG_02507 | Predicted peroxidase; secreted | + | − | − | + |   |
| PADG_07053 | Predicted protease; peptidase family 16 | − | − | − | + |   |
| PADG_00686 | Predicted protease; peptidase family M28 | − | − | − | + |   |
| PADG_06314 | Predicted protease; peptidase family S10; secreted; induced during macrophage interaction | + | + | − | + |   |
| PADG_06167 | Predicted protease; peptidase family S24 | − | − | − | + |   |
| PADG_02492 | Predicted protease; peptidase family S8; secreted | − | − | − | + |   |

(Continued on following page)
high-affinity nickel transporter (PADG_05345), an oxidoreductase (PADG_00948) and the SOD3-encoded protein. Other genes that were found significantly induced during the interaction of Blastomyces with macrophages (14) were also found to evolve under positive selection in Paracoccidioides, including the two most highly upregulated in infected macrophages, a gene coding for a secreted protein with unknown function (PADG_01283) and a secreted endo-1,3(4)-β-glucanase gene, ENG1 (PADG_12370); other genes found in both analyses include genes coding for GATA-binding protein (PADG_05497), an aldehyde reductase (PADG_01835), and a glucan 1,3-β-glucosidase (PADG_06699) (Table 2; see Data Set S1). These genes therefore may be generally important for host interactions of dimorphic pathogens and make good candidates for future studies on their contribution to virulence.

Combining diversity measures reveals lineage-specific targets of positive selection. Examination of the local variation in nucleotide diversity (π) and Tajima’s D (TD) across the genome revealed small regions with high lineage-specific diversity, which also typically showed high TD (Fig. 4C; see Data Set S1 in the supplemental material). PS2 had the most lineage-specific high-diversity windows, with 14 in total, coinciding with the more ancient divergence of this lineage and in agreement with the phylogenetic analysis. The windows with significantly high π and TD highlighted lineage-specific regions of high diversity and selection, which in some cases were regions with low interspecific divergence (FST) (Fig. 4B).

These high-diversity regions highlighted different sets of variable and rapidly evolving genes within lineage. These genes encompass diverse cellular functions, including coding for secreted and cell wall proteins, transport, transcription regulation, oxidative stress, and proteolysis (Table 2; see Data Set S1), some of which have experimental evidence of a role in virulence and pathogenicity (see Text S1 in the supplemental material). Two high-diversity regions found for S1b include two aminopeptidases (PADG_05820 and PADG_07460) previously noted to be upregulated in Paracoccidioides during macrophage infection (18). In the low-diversity PS3 isolates, one large highly variable region was identified that includes 125 genes (23 with a dN/dS of >1, 12 secreted protein genes, and 9 genes involved with the oxidation reduction process). Among these genes are genes encoding two glycosylphosphatidylinositol (GPI)-anchored proteins (including PGA1) and the secreted protein PADG_02535, which was identified in extracellular vesicles of Paracoccidioides (19), and it was recently shown in B. dermatitidis that an ortholog was more highly expressed during mouse pulmonary infection (14). In two regions found for PS2, 11 of 48 genes show evidence of positive selection (dN/dS of >1) and include genes coding for a serine carboxypeptidase
(PADG_06314) induced in Paracoccidioides during macrophage infection (18) and a peroxidase (PADG_06322), as well as the amino acid permease PADG_07440, which in B. dermatitidis was highly induced during mouse pulmonary infection (14). Differences in these genes might impact the virulence phenotypes observed across the lineages and make good candidates for future studies of their contribution to virulence variation between isolates of Paracoccidioides (Table 2).

**DISCUSSION**

Building on previous genomic analysis of Paracoccidioides (12, 13), here we reevaluate the major lineages and provide new reference genomes for two lineages, PS3 and PS4. Together these data enable a more comprehensive view of the genome content of Paracoccidioides and help trace the genome-level variation following the recent divergence into well-defined lineages. We find clear support for the separation of P. brasiliensis into distinct lineages; however, we also find evidence of recombination and highest diversity within a single lineage, the newly described S1b. Dating the timing of the separation of each lineage supported S1b as the earliest diverging P. brasiliensis lineage, with the later emergence of PS2, S1a, and PS3.

Our study provides additional support for P. brasiliensis and P. lutzii as separate species, with some amount of incomplete lineage sorting suggested by intermediate FST values between the P. brasiliensis lineages. Species definitions have been reevaluated in many fungal groups, including the Onygenales (20), based on improved phylogenies from wider sets of isolates compared or incorporation of additional loci, or even whole genomes, in addition to the analysis of morphological data of sexual and asexual structures and the evidence of sexual reproduction. Using multiple approaches, we see clear separation of P. brasiliensis and P. lutzii and evidence of recombination within P. brasiliensis; however, this recombination did not appear to be recent. While the phylogenetic separation suggests the P. brasiliensis lineages are largely genetically isolated, increased sample size is required to compare variation within and between the lineages to support their separation into different species.

We found support for a split in the S1 group into S1a and S1b lineages; in addition to genetic separation, signatures of recombination clearly differentiate S1b from S1a. While the phylogenetic analysis suggests that S1a and S1b are both monophyletic, we also find evidence from multiple analyses that S1b, the most geographically widespread lineage, has undergone recombination with each of the other lineages. This appears to be mostly ancestral, as FST while variable did not reveal any large recently introgressed regions between lineages. The mixed ancestry of PS4 also suggests that more ancient recombination created this lineage, as the pattern of SNP ancestry across the genome revealed an intermixed distribution of both S1b and PS3 alleles that were likely shuffled by extensive recombination over time. The sequence of additional isolates of PS4 as well as S1b could help further explore how alleles within these groups are distributed within each lineage and across geographic regions.

All lineages, including S1a and S1b, contain roughly equal numbers of both mating types based on the sequenced isolates; an equal ratio was also previously noted in a wider set of P. brasiliensis isolates (21). The equal representation suggests that other differences between the strains explain how S1b has undergone higher levels of recombination compared to S1a. Direct testing of mating potential for different strains requires the further development of laboratory mating experiments, which appear to only initiate but not complete mating to date (22). In addition, further studies comparing larger numbers of isolates from the same geographic region but different lineage groups would be more sensitive to detecting recent exchange between groups in nature.

This genome-wide comparison of the major lineages of Paracoccidioides revealed that while the genome organization and gene content are highly conserved, specific genes under positive selection include several previously known to be important for host interaction as well as new candidates. Two of the genes we identified (GP43 and CTS20) were noted in a previous study, which utilized expressed sequence tags (ESTs) and found evidence of positive selection in 11 of 32 examined genes implicated in
virulence (23). Here, we find evidence for positive selection in an average of 485 genes per strain using the updated genome sequences. In addition to GP43, we see evidence of selection in the gene coding for the secreted antigenic GPI-anchored protein, PGA1 (24). Other genes under positive selection include those coding for aminopeptidases highly upregulated during host-pathogen interactions (14), secreted proteins previously found in Paracoccidioides extracellular vesicles (19), and other proteins involved in mitigating oxidative stress that were also found induced during the interaction with macrophages, such as the AOX, SOD3, and CBP1 genes (14, 18, 25–27). Secreted proteins in Paracoccidioides have been associated with nutrient acquisition, cell defense, and modulation of the host defense machinery.

Better knowledge of the differences between the Paracoccidioides lineages can help guide the development of new diagnostics, treatments, and models of pathogen evolution. Delineation of the prevalent S1 group into the well-separated S1a and S1b lineages allowed us to differentiate S1b as the more highly recombining and diverse lineage. The higher diversity present in S1b may contribute to its dispersion and survival in a wider geographical range than the other lineages, perhaps enabling local adaptation such as to temperature variation. In our data, all of the S1b strains are clinical isolates and include two highly virulent strains, Pb18 and Pb113, and two strains from acute PCM cases (PbBl0 and PbCaz). In contrast, the S1a lineage includes clinical and armadillo isolates, with very few sequence differences between strains from these sources. Sequence of isolates from a wider geographical range, including Brazil, Argentina, Paraguay, and Mexico, would likely increase the sampled genetic diversity of Paracoccidioides and allow further study of these trends; additional isolates could also be used to increase genome-wide association studies of clinically relevant phenotypes. Development of sequence-based diagnostics will need to take into consideration that S1b strains share a higher percentage of alleles with all other strains; as these sites could contribute to misidentification of lineages, sufficient power or selection of sites unique to each group that also take into account the higher diversity in S1b would improve diagnostic power.

### MATERIALS AND METHODS

**Selection and sequencing of Paracoccidioides isolates.** A total of 31 isolates of *P. brasiliensis* and *P. lutzii* were sequenced and included in the analyses (Table 1): 11 isolates from lineage S1a, 4 isolates from S1b, 7 isolates from PS3, 4 isolates from PS2, 1 isolate from PS4, and 4 isolates from *P. lutzii*. Selected isolates included clinical isolates from acute and chronic PCM cases, environmental isolates from armadillos, and the previously published and updated reference genomes of *P. brasiliensis* (Pb03 from PS2 and Pb18 from S1) and Pb01 from *P. lutzii* (12, 13). Genomic DNA for sequencing was prepared from yeast culture, using phenol-chloroform extraction method. The genomes were sequenced using Illumina sequencing platforms and using different insert sizes (~620, ~180, or ~650 bp), paired-end-read lengths (150, 101, or 100 bp) and sequence coverage ranging from 68 to 419 (see Table S1 in the supplemental material).

**Identification and analysis of gene orthologs and selection analysis.** For a detailed description of the genome assembly and gene annotation of *P. brasiliensis* Pb300 and PbCnh strains, see the supplemental material. The five assembled and annotated genomes representing each lineage of the *Paracoccidioides* species were used for comparative analysis. Genes were functionally annotated by assigning Pfam domains, GO terms, KEGG classification, SignalP, and TMHMM. OrthoMCL (28) was used to cluster the protein-coding genes of the four chosen genomes by similarity and create sets of genes that have a high probability of being orthologous to each other.

For selection analysis, variant call format (VCF) files were used to align the reference transcript set in coding triplets in the PHYLIP multiple sequence alignment format. Then we calculated the ratio of nonsynonymous to synonymous substitutions (dN/dS) within the *Paracoccidioides* genus on properly aligned genes. We employed the yn00 program in PAML (29), implementing the Yang and Nielsen method (30). To eliminate the possibility that fast-evolving genes detected in this analysis were biased for false-positive SNP calls, we calculated the mean base quality, mean mapping quality, and quality normalized to depth and compared these parameters for all fast-evolving genes and all other genes.

**SNP variant detection and analysis.** To detect polymorphisms, we used reference assemblies representing each lineage (Pb18, Pb03, PbCnh, Pb300, and Pb01). Each of the 31 Illumina data sets was independently aligned to the genome assemblies using the short read component aln of BWA version 0.5.9 (31) with default settings. SNPs and indels were called with Pilon version 1.4 using the haploid plodpy default setting (32). An average of 38 million reads was aligned per strain, with an average quality and error rate of 33.2 and 2.0E−02, respectively (see Table S1 in the supplemental material). Variant call format (VCF) files were filtered using VCFTools version 0.1.12 (33) or according to further analyses—e.g.,
considering SNP calls in all strains with genotype 1/1 and minimum depth 4. For SNP positions, the total mapping depth was 125 and the mean base quality was 34, averaged across all variants (see Table S1).

To address if any lineage could be uniformly diploid, we examined candidate heterozygous positions predicted by Pilon. The low frequency of such positions (~0.04%), which often overlap repetitive sequence (68% of these positions), suggests there is little evidence for diplomy. This suggests that all sequenced genomes are homozygous haploids, as expected from prior work establishing that *Paracoccidioides* species are haploid (34, 35).

The false discovery rate (FDR) was estimated as an additional parameter of the mapping and SNP calling accuracy (36). As a truth set, we simulated 670,000 random mutations in the reference genome (Pb18). The number of random mutations was the maximum number of mutations detected with the sequenced strains. Next, we aligned the raw reads to the random-mutated reference, called SNPs, and compared them to the known truth set of simulated mutations to calculate the accuracy of our data and process both the number of true positives (658,566; precision, 99.8%; and sensitivity, 98.3%) and the number of false positives (386; 0.06%) show the high accuracy of the overall process, including read quality, alignment accuracy, and the SNP calls.

To determine the chromosomal copy number variation (CCNV) and the distribution of SNPs across the genome, we calculated the alignment density (depth of read coverage) and SNP density (frequency of SNPs per site), respectively. All VCF files were summarized in 10-kb nonoverlapping windows. The alignment density was normalized by the average genomic coverage and the window length. The SNP density was normalized by the window length. The genome assembly of Pb18 is anchored to chromosomes (12); we used this reference to map the alignment density and SNP density into chromosomes. Pb03, PbCnh, Pb300, and Pb01 were mapped into scaffolds.

To more finely compare variant sites between lineages, we classified the distribution of biallelic loci between two populations categorized as “shared,” “fixed,” or “private.” If one allele is present in all members of one population and the other allele is present in all members of the other population, that locus was considered “fixed.” If both alleles are present in both populations, that locus was considered “shared.” If one allele is present in all members of one population and the other population has both alleles, that locus was considered “private.”

To analyze each SNP in the context of gene annotations of the reference genomes we used VCFannotator (http://sourceforge.net/projects/vcfannotator/). Variants were placed in the context of genome feature annotations and indicated as “intergenic,” “intronic,” or “coding.” The type of coding mutation was further characterized by its impact on the protein-coding sequence as “synonymous,” “nonsynonymous,” “frameshift,” “nonsense,” or “read-through.” To evaluate the support for each mutation, the mean base quality, the mean mapping quality, and the quality normalized to depth were retrieved and compared.

**Phylogenetic analysis using whole-genome SNPs.** Alignments were constructed from SNP matrices extracted from the VCF format. To consider a position in the alignment matrix, we used the minimum depth of coverage of 4, and we kept those positions with at least one variant site in all the sequenced isolates. We built an SNP alignment matrix for each *Paracoccidioides* reference strain used here representing each lineage to obtain four trees and compare their topologies. Maximum likelihood phylogenies were constructed using RAxML version 8.0.20 (37) using the GTRCAT nucleotide substitution model and bootstrap analysis based on 1,000 replicates. To infer Bayesian phylogenetic trees and to estimate the divergence and most common ancestor dates, we used BEAST v1.8.2 (38). Using the SNP matrices, we normalized the mutation rate for genome-wide variable sites (normalized, 4.43E−8; genome-wide, 1E−9 [39]) that was used as the clock rate along with the coalescent model, relaxed clock/uncorrelated lognormal clock model. For nonpartitioned variant sites, we used the strict clock model, with the constant pattern model, general time reversible (GTR) (G+I) substitution model, and 20,000,000 Markov chain Monte Carlo (MCMC) chains. Tracer v1.6.0 (http://beast.bio.ed.ac.uk/Tracer) was used to visualize traces and inspect the effective sample size and MCMC. We determined the relationship of the sequenced isolates using the NeighborNet algorithm with SplitsTree4 (40).

**Population genetic structure analyses.** We performed a principal-component analysis (PCA) on a matrix of SNP calls for all of the *Paracoccidioides* isolates (*n* = 31) and for only the *P. brasiliensis* isolates (*n* = 26), using SMARTPCA (41). Population structure was performed using the Bayesian model-based clustering program STRUCTURE v2.3 (16) in the site-by-site mode, with successive *k* values from 2 to 6. We identified populations within Paracoccidioides using 476,589 SNPs in all isolates and within *P. brasiliensis* isolates using 339,966 SNPs. We estimated the model evidence for *k* using thermodynamic integration method as implemented in Maverick v1.0 (42), using replicates for 10 randomly generated 1% subsamples of the *P. brasiliensis* SNP matrix; this supported *k* = 4 as the best choice for *P. brasiliensis* (see Fig. S5B in the supplemental material). Genome-wide nucleotide diversity (*π*) and Tajima’s *D* were computed for each identified *Paracoccidioides* population (S1a, S1b, PS2, PS3, and P. lutzii) using VCFTools v0.1.12 (33). The average nucleotide diversity (*π*) and Tajima’s *D* were computed for nonoverlapping sliding windows of 10 kb. We calculated Wright’s fixation index (*F*<sub>ST</sub> [43]) according to the equations given in VCFtools v0.1.12 (33) adjusted for low sample sizes, sliding-window *F*<sub>ST</sub> analyses were conducted using all SNPs found within 10-kb nonoverlapping windows. Variant call format (VCF) files were filtered using an in-house perl script to calculate *F*<sub>ST</sub> and clusters for comparison were chosen based on the whole-genome phylogenetic tree.

**Accession number(s).** The assemblies and annotations of the *P. brasiliensis* genomes have been deposited in DDBJ/ENA/GenBank under the following accession numbers: *Paracoccidioides brasiliensis* PbCnh, LYUC00000000; *Paracoccidioides brasiliensis* Pb300, LZY000000000. All of the whole-genome sequence (WGS) raw data for the 31 *Paracoccidioides* strains have been deposited in the NCBI Sequence
Read Archive (BioProject, PRJNA322632; SRA, SRP077566). BioSample and SRA accession numbers for individual strains are included in Table S1 in the supplemental material.

SUPPLEMENTAL MATERIAL
Supplemental material for this article may be found at http://dx.doi.org/10.1128/mSphere.00213-16.

Text S1, DOCX file, 0.1 MB.
Data Set S1, XLSX file, 0.6 MB.
Figure S1, PDF file, 1.9 MB.
Figure S2, PDF file, 2.1 MB.
Figure S3, PDF file, 0.5 MB.
Figure S4, PDF file, 1.1 MB.
Figure S5, PDF file, 0.2 MB.
Figure S6, PDF file, 0.3 MB.
Figure S7, PDF file, 0.5 MB.
Table S1, XLSX file, 0.1 MB.

ACKNOWLEDGMENTS
We thank Angela Restrepo, Rosana Puccia, Zoilo Pires de Camargo, and Maria Sueli Felipe for kindly providing the isolates for this study.

This project has been funded in whole or in part with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under contract no. HHSN272200900018C. This work was partly supported by Colciencias via the grants “A Gene Atlas for Human Pathogenic Fungi” (122256934875) and “A Comprehensive Genomic and Transcriptional Analysis of Dimorphic Human Pathogen Fungi and Its Relation with Virulence” (221365842971) and by the Universidad de Antioquia via a “Sostenibilidad 2015/2016” grant. Colciencias National Doctorate Program funding supported J.F.M.; Enlaza Mundos partly supported his fellowship. The Wellcome Trust supported R.A.F.

J.F.M., J.G.M., O.K.C., and C.A.C. conceived and designed the experiments. J.F.M., R.A.F., C.A.D., and C.A.C. analyzed the data. J.F.M. and C.A.C. assembled and annotated Paracoccidioides genomes. J.E.G., S.M.S., S.S., E.M., E.A.W., E.B., C.M.A.S., M.D.M.T., and J.W.T. contributed reagents, materials, and/or analysis tools. J.F.M., R.A.F., C.A.D., C.A.C., J.E.G., S.M.S., S.S., E.M., E.A.W., E.B., C.M.A.S., M.D.M.T., and J.W.T. performed the experiments.

FUNDING INFORMATION
This work, including the efforts of José F. Muñoz, Juan E. Gallo, Elizabeth Misas, Oliver K. Clay, and Juan G. McEwen, was funded by Colciencias (122256934875 and 221365842971). This work, including the efforts of José F. Muñoz, Rhys A. Farrer, Christopher A. Desjardins, and Christina A. Cuomo, was funded by HHS | NIH | National Institute of Allergy and Infectious Diseases (NIAID) (HHSN272200900018C). This work, including the efforts of Rhys A. Farrer, was funded by Wellcome Trust. This work, including the efforts of Marcus de M. Teixeira, was funded by Conselho Nacional de Desenvolvimento Científico e Tecnológico (460999/2014-1).

REFERENCES
1. Bocca AL, Amaral AC, Teixeira MM, Sato PK, Shikanai-Yasuda MA, Soares Felipe MS. 2013. Paracoccidioidomycosis: eco-epidemiology, taxonomy and clinical and therapeutic issues. Future Microbiol 8:1177–1191. http://dx.doi.org/10.2217/fmb.13.68.
2. Bellissimo-Rodrigues F, Vitali LH, Martinez R. 2010. Serological diagnosis of paracoccidioidomycosis in HIV-infected patients. Mem Inst Oswaldo Cruz 105:904–907. http://dx.doi.org/10.1590/S0074-02762010000700011.
3. Brummer E, Castaneda E, Restrepo A. 1993. Paracoccidioidomycosis: an update. Clin Microbiol Rev 6:89–117. http://dx.doi.org/10.1128/CMR.6.2.89.
4. Teixeira MM, Theodoro RC, de Carvalho MJ, Fernandes L, Paes HC, Hahn RC, Mendoza L, Bagagli E, San-Blas G, Felipe MS. 2009. Phylogenetic analysis reveals a high level of speciation in the Paracoccidioides genus. Mol Phylogenet Evol 52:273–283. http://dx.doi.org/10.1016/j.ympev.2009.04.005.
5. Matute DR, McEwen JG, Puccia R, Montes BA, San-Blas G, Bagagli E, Rauscher JT, Restrepo A, Morais F, Niño-Vega G, Taylor JW. 2006. Cryptic speciation and recombination in the fungus Paracoccidioides brasiliensis as revealed by gene genealogies. Mol Biol Evol 23:65–73. http://dx.doi.org/10.1093/molbev/msj008.
6. Teixeira MM, Theodoro RC, Niño-Vega G, Bagagli E, Felipe MSS. 2014. Paracoccidioides species complex: ecology, phylogeny, sexual reproduction, and virulence. PLoS Pathog 10:e1004397. http://dx.doi.org/10.1371/journal.ppat.1004397.
7. Macoris SA, Sugizaki MF, Peraçoli MT, Bosco SM, Hebele-Barbosa F,
Genome Diversity of Paracoccidioides Lineages

Simões LB, Theodoro RC, Trinca LA, Bagagli E. 2006. Virulence attenuation and phenotypic variation of Paracoccidioides brasiliensis isolates obtained from armadillos and patients. Mem Inst Oswaldo Cruz 101: 331–334. http://dx.doi.org/10.1590/S0070-02762006000300019.

8. Theodoro RC, Teixeira MdM, Felipe MS, dos Santos Paduán K, Ribolla PM, San-Blas G, Bagagli E. 2012. Genus Paracoccidioides: species recognition and biogeographic aspects. PLoS ONE 7e37694. http://dx.doi.org/10.1371/journal.pone.0037694.

9. Montoya AE, Alvarez AL, Moreno MN, Restrepo A, McEwen JG. 1999. Electrophoretic karyotype of environmental isolates of Paracoccidioides brasiliensis. Med Mycol 37:219–222. http://dx.doi.org/10.1080/1369378992011-0917.x.

10. Soares CM, Madlun EE, da Silva SP, Pereira M, Felipe MS. 1995. Characterization of Paracoccidioides brasiliensis isolates by random amplified polymorphic DNA analysis. J Clin Microbiol 33:505–507.

11. Puccia R, McEwen JG, Cisalpino PS. 2008. Diversity in Paracoccidioides brasiliensis. The PbgP43 gene as a genetic marker. Mycopathologia 165:275–287. http://dx.doi.org/10.1007/s11046-007-9055-2.

12. Desjardins CA, Champion MD, Holder JW, Muszewska A, Goldberg J, Bailão AM, Briggs MM, Ferreira ME, Garcia AM, Gryngberg G, Guja S, Heiman DI, Henn MR, Kodira CD, Léon-Narváez H, Longo LV, Ma LJ, Malavazi I, Matsuo AL, Morais FW, Pereira M, Roque-Barreira MC, Sebghati T, Smith WS, Teixeira MM, Valjeki MC, Walter ME, 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, Birren BW, Cuomo CA. 2011. Comparative genomic analysis of human fungal pathogens causing paracoccidioidomycosis. PLoS Genet 7:e1002345. http://dx.doi.org/10.1371/journal.pgen.1002345.

13. Muñoz JF, Gauthier GM, Prieto M, Amimovic A, Young S, Zeng Q, Clay OK, McEwen JG, Cuomo CA. 2014. Genome update of the dimorphic human pathogenic fungi causing paracoccidioidomycosis. PLoS Negl Trop Dis 8:e33348. http://dx.doi.org/10.1371/journal.pntd.00033348.

14. Muñoz JF, Gauthier GM, Desjardins CA, Gallo JE, Holder J, Sullivan TD, Marty AJ, Carmen JC, Chen Z, Ding L, Guja S, Magrini V, Misas E, Mitreva M, Prieto M, Saif S, Whitson EA, Young S, Zeng Q, Goldman WE, McEwen JG, Taylor JW, Clay OK, McEwen JG, Cuomo CA. 2015. The dynamic genome and transcriptome of the human fungal pathogen Blastomyces and close relative Emmonsia. PLoS Genet 11:e1005493. http://dx.doi.org/10.1371/journal.pgen.1005493.

15. Bruen TC, Philippe H, Bryant D. 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics 172: 2665–2681. http://dx.doi.org/10.1534 genetics.105.048975.

16. Pritchard JK, Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

17. Neafsey DE, Barker BM, Sharpton TJ, Stajich JE, Park DJ, Sharpton TJ, Whiston E, Barker BM, McEwen JG, Nino-Vega G, Puccia R, San-Blas G, Soares CM, Birren BW, Cuomo CA. 2016. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9:e112963. http://dx.doi.org/10.1371/journal.pone.0112963.

18. Danecik F, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lieber MR, McVean G, Sherry ST, Abecasis GR, Durbin R, 1000 Genomes Project Analysis Group. 2011. The varaint call format and VCFTools. Bioinformatics 27:2156–2158.

19. Almeida AJ, Matute DR, Martins M, Torres I, McEwen JG, Restrepo A, Leão C, Ludovico P, Rodrigues F. 2007. Genome size and ploidy of Paracoccidioides brasiliensis reveals a haploid DNA content: flow cytometry and GP43 sequence analysis. Fungal Genet Biol 44: 25–31. http://dx.doi.org/10.1016/j.fgb.2006.06.003.

20. Feitosa LSDs, Cisalpino PS, dos Santos MR, Mortara RA, Barros TF, Morais AFS, Puccia R, da Silveira JF, de Camargo ZP. 2003. Chromosomal polymorphism, syntenic relationships, and ploidy in the pathogenic fungus Paracoccidioides brasiliensis. Fungal Genet Biol 39:60–69. http://dx.doi.org/10.1016/S1087-1845(03)00003-3.

21. Farrer RA, Henk DA, MacLean D, Stulholme DJ, Fisher MC. 2013. Using false discovery rates to benchmark SNP callers in next-generation sequencing projects. Sci Rep 3:1512. http://dx.doi.org/10.1038/srep01512.

22. Stamatakis A. 2006. RAxML-VI-HPC: maximum-likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690. http://dx.doi.org/10.1093/bioinformatics/bti446.

23. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29:1969–1973. http://dx.doi.org/10.1093/molbev/msw075.

24. Sharpton TJ, Stajich JE, Rounsley SD, Gardner JM, Wortman JR, Jordan VS, Maiti R, Kodira CD, Neafsey DE, Zeng Q, Hung CY, McManus C, Muszewska A, Gryngberg M, Mandel MA, Kellner EM, Barker BM, Galjani MJ, Oberbach MJ, Kirkland TN, Cole GT, Birren BW, Henn MR, Cuomo CA. 2009. Comparative genomic analyses of the human fungal pathogen and their relatives. Genome Res 19:1722–1731. http://dx.doi.org/10.1101/gr.087511.108.

25. Kloepper TH, Huson DH. 2008. Drawing explicit phylogenetic networks

Volume 1 Issue 5 e00213-16 msphere.asm.org 17 msphere.asm.org

Downloaded from http://msphere.asm.org on January 8, 2017 by guest
and their integration into SplitsTree. BMC Evol Biol 8:22. http://dx.doi.org/10.1186/1471-2148-8-22.

41. Patterson N, Price AL, Reich D. 2006. Population structure and eigenanalysis. PLoS Genet 2:e190. http://dx.doi.org/10.1371/journal.pgen.0020190.

42. Verity R, Nichols RA. 2016. Estimating the number of subpopulations (K) in structured populations. Genetics 203:1827–1839 http://dx.doi.org/10.1534/genetics.115.180992.

43. Hudson RR, Slatkin M, Maddison WP. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics 132:583–589.

44. Teixeira HC, Calich VL, Singer-Vermes LM, D’Imperio-Lima MR, Russo M. 1987. Experimental paracoccidioidomycosis: early immune-suppression occurs in susceptible mice after infection with pathogenic fungi. Braz J Med Biol Res 20:587–589.

45. Imai T, Sano A, Mikami Y, Watanabe K, Aoki FH, Branchini ML, Negroni R, Nishimura K, Miyaji M. 2000. A new PCR primer for the identification of Paracoccidioides brasiliensis based on rRNA sequences coding the internal transcribed spacers (ITS) and 5′-8S regions. Med Mycol 38:323–326. http://dx.doi.org/10.1080/mmy.38.4.323.326.

46. Hanna SA, Monteiro da Silva JL, Giannini MJ. 2000. Adherence and intracellular parasitism of Paracoccidioides brasiliensis in Vero cells. Microbes Infect 2:877–884. http://dx.doi.org/10.1016/S1286-4579(00)00390-7.

47. Teixeira MdM, Theodoro RC, Oliveira FF, Machado GC, Hahn RC, Bagagli E, San-Blas G, Soares Felipe MS. 2014. Paracoccidioides lutzii sp. nov.: biological and clinical implications. Med Mycol 52:19–28. http://dx.doi.org/10.1089/mmy.2013.794311.

48. Hebeler-Barbosa F, Montenegro MR, Bagagli E. 2003. Virulence profiles of ten Paracoccidioides brasiliensis isolates obtained from armadillos (Dasypus novemcinctus). Med Mycol 41:89–96. http://dx.doi.org/10.1080/02692830312331271090.

49. Machado GC, Moris DV, Arantes TD, Silva LR, Theodoro RC, Mendes RP, Vicentini AP, Bagagli E. 2013. Cryptic species of Paracoccidioides brasiliensis: impact on paracoccidiomycosis immunodiagnosis. Mem Inst Oswaldo Cruz 108:637–643. http://dx.doi.org/10.1590/0074-0276108052013016.

50. De Albornoz MB. 1971. Isolation of Paracoccidioides brasiliensis from rural soil in Venezuela. Sabouraudia 9:248–253. http://dx.doi.org/10.1080/0036217185190491.

51. Restrepo-Moreno A, Schneidau JD, Jr. 1967. Nature of the skin-reactive principle in culture filtrates prepared from Paracoccidioides brasiliensis. J Bacteriol 93:1741–1748.

52. Bustamante-Simon B, McEwen JG, Taboas AM, Arango M, Restrepo-Moreno A. 1985. Characteristics of the conidia produced by the mycelial form of Paracoccidioides brasiliensis. Sabouraudia 23:407–414. http://dx.doi.org/10.1080/00362175853806061.

53. Corredor GG, Peralta LA, Castaño JH, Zuluaga JS, Henao B, Arango M, Taboas AM, Matute DR, McEwen JG, Restrepo A. 2005. The naked-tailed armadillo Cabassous centralis (Miller 1899); a new host to Paracoccidioides brasiliensis. Medical identification of the isolate. Med Mycol 43:275–280. http://dx.doi.org/10.1080/13693780412331271090.

54. Morais FV, Barros TF, Fukada MK, Cisalpino PS, Puccia R. 2000. Polymorphism in the gene coding for the immunodominant antigen gp43 from the pathogenic fungus Paracoccidioides brasiliensis. J Clin Microbiol 38:3960–3966.

55. Ferreira MS, Freitas LH, Lacaz CD, de Melo NT, Garcia NM, de Assis CM, Salebian A, Heins-Vaccari EM. 1990. Isolation and characterization of a Paracoccidioides brasiliensis strain from a dog-food probably contaminated with soil in Uberlândia, Brazil. J Med Vet Mycol 28:253–256. http://dx.doi.org/10.1080/02681219080000311.

56. Carrero LL, Niño-Vega G, Teixeira MM, Carvalho MJ, Soares CM, Pereira M, Jesuino RS, McEwen JG, Mendoza L, Taylor JW, Felipe MS, San-Blas G. 2008. New Paracoccidioides brasiliensis isolate reveals unexpected genomic variability in this human pathogen. Fungal Genet Biol 45:605–612. http://dx.doi.org/10.1016/j.fgb.2008.02.002.