OXFORD GENETICS

Impact of pathogen genetics on clinical phenotypes in a population of *Talaromyces marneffei* from Vietnam

Poppy Sephton-Clark,[1] Thu Nguyen,[2] Ngo Thi Hoa,[3,4,5] Philip Ashton,[6] H. Rogier van Doorn,[4,7] Vo Trieu Ly,[4,8] Thuy Le,[2,9] Christina A. Cuomo[1,*]

[1]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[2]Division of Infectious Diseases and International Health, Duke University School of Medicine, Durham, NC 27710, USA
[3]Oxford University Clinical Research Unit, Oxford University, Ho Chi Minh City 749000, Vietnam
[4]Centre for Tropical Medicine and Global Health, University of Oxford, Oxford OX37LG, UK
[5]Microbiology department and Biological Research Center, Pham Ngoc Thach University of Medicine, Ho Chi Minh City 740500, Vietnam
[6]Veterinary and Ecological Sciences, Institute of Infection, University of Liverpool, Liverpool CH647TE , UK
[7]Oxford University Clinical Research Unit, Oxford University, Hanoi 113000, Vietnam
[8]Department of Medicine and Pharmacy, Hospital for Tropical Diseases, Ho Chi Minh City 749000, Vietnam
[9]Tropical Medicine Research Center for Talaromycosis, Pham Ngoc Thach University of Medicine, Ho Chi Minh City 740500, Vietnam

*Corresponding author: Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.
Email: cuomo@broadinstitute.org

Abstract

Talaromycosis, a severe and invasive fungal infection caused by *Talaromyces marneffei*, is difficult to treat and impacts those living in endemic regions of Southeast Asia, India, and China. While 30% of infections result in mortality, our understanding of the genetic basis of pathogenesis for this fungus is limited. To address this, we apply population genomics and genome-wide association study approaches to a cohort of 336 *T. marneffei* isolates collected from patients who enrolled in the Itraconazole vs Amphotericin B for Talaromycosis trial in Vietnam. We find that isolates from northern and southern Vietnam form two distinct geographical clades, with isolates from southern Vietnam associated with increased disease severity. Leveraging longitudinal isolates, we identify multiple instances of disease relapse linked to unrelated strains, highlighting the potential for multistrain infections. In more frequent cases of persistent talaromycosis caused by the same strain, we identify variants arising over the course of patient infections that impact genes predicted to function in the regulation of gene expression and secondary metabolite production. By combining genetic variant data with patient metadata for all 336 isolates, we identify pathogen variants significantly associated with multiple clinical phenotypes. In addition, we identify genes and genomic regions under selection across both clades, highlighting loci undergoing rapid evolution, potentially in response to external pressures. With this combination of approaches, we identify links between pathogen genetics and patient outcomes and identify genomic regions that are altered during *T. marneffei* infection, providing an initial view of how pathogen genetics affects disease outcomes.

Keywords: *Talaromyces marneffei*, talaromycosis, penicilliosis, Vietnam, whole genome sequencing, population genomics, GWAS

Introduction

The thermally dimorphic fungal pathogen, *Talaromyces marneffei*, causes talaromycosis, a severe fungal infection that affects immunocompromised people living in, or traveling to, endemic regions spanning Southeast Asia, India, and China (Qin *et al.* 2020; Narayanasamy, Dougherty, *et al.* 2021). People with advanced HIV disease are particularly affected, in whom the prevalence of talaromycosis can reach 26.5% in some regions (Qin *et al.* 2020). The impact of this disease is an estimated 4,900 (95% CI 2,500–7,300) deaths annually (Narayanasamy, Dat, *et al.* 2021). Talaromycosis is difficult to treat, with mortality rates of up to 30% (Le *et al.* 2011; Jiang *et al.* 2019), and rates of infection increase throughout the rainy season, disproportionately impacting farmers and agricultural workers (Chariyalertsak *et al.* 1997; Narayanasamy, Dougherty, *et al.* 2021).

T. marneffei infection originates primarily through inhalation (Narayanasamy, Dougherty, *et al.* 2021). Once in the lungs, *T. marneffei* can disseminate to multiple organs, including the liver, spleen, lymph nodes, blood, bone marrow, bone, and skin (Supparatpinyo *et al.* 1994; Vanittanakom *et al.* 2006; Wong and Wong 2011; Narayanasamy, Dougherty, *et al.* 2021). The infectious propagule includes aerosolized conidia, which may transition to the pathogenic yeast form upon inhalation (Pruksaphon *et al.* 2022). Effective containment by lung-resident primary immune cells is critical to preventing dissemination, as *T. marneffei* possesses multiple virulence traits that enable its persistence within the human host. These virulence factors include the ability to grow at 37°C (Vanittanakom *et al.* 2006), the production of reactive oxygen species detoxifying enzymes that may promote survival within host macrophages (Pongpom *et al.* 2013), the secreted galactomannoprotein Mp1 that suppresses the proinflammatory host immune responses (Lam *et al.* 2019), and

multiple laccases important for resisting external stressors and killing by phagocytes (Sapmak *et al.* 2016). The recommended treatment for Talaromycosis includes amphotericin B and itraconazole. Whilst there are no established breakpoints described for antifungal resistance, minimum inhibitory concentrations of azoles (0.008–16 μg/ml) and amphotericin B (0.12–1 μg/ml) have been reported against *T. marneffei* (Stott *et al.* 2021; Fang *et al.* 2022). Further identification of genes important for virulence and resistance in *T. marneffei* is still needed to enable a better understanding of how *Talaromyces* adapts and survives within the host.

Recently generated complete genome assemblies for representative isolates of *T. marneffei* from north and south Vietnam (Cuomo *et al.* 2020) will enable further discovery of genes essential for virulence, as well as the evaluation of genetic variability across clinical isolates. Genetic variation in populations of many pathogenic fungi has been surveyed (Desjardins *et al.* 2017; Rhodes, Desjardins, *et al.* 2017, 2022; Chow *et al.* 2020; O'Brien *et al.* 2021), and our initial understanding of variation across populations of *T. marneffei* has been guided by work that uncovered the population delineation of country-linked clades, based on a multilocus sequence typing (MLST) study (Henk *et al.* 2012). This prior work supplied an overview of the population structure of *T. marneffei* and provided a foundation for further analysis, based on whole genome sequencing, to better understanding of the genetic variation within clinical populations of *T. marneffei*. Characterizing genetic heterogeneity in populations of *Aspergillus*, *Cryptococcus*, and *Candida* has enabled a better understanding of resistance mechanisms and virulence traits through the study of aneuploidy (Selmecki *et al.* 2006; Poláková *et al.* 2009; Stone *et al.* 2019), copy number variation (Chow *et al.* 2020), and variant–phenotype association studies (Desjardins *et al.* 2017; Gerstein *et al.* 2019; Rhodes *et al.* 2022; Sephton-Clark *et al.* 2022). Population studies leveraging whole genome data have also allowed for the discovery of clade specific adaptation and evolution, such as the loss of subtelomeric adhesins in the only non-outbreak causing *Candida auris* clade (clade II) (Muñoz *et al.* 2021). In addition to understanding heterogeneity across populations, these approaches have enabled the interrogation of variability evolving in isolates over time, through the study of patient longitudinal and relapse isolates (Ormerod *et al.* 2013; Chen *et al.* 2017; Helmstetter *et al.* 2022), and microevolution of isolates passaged through infection models (Ene *et al.* 2018; Fu *et al.* 2021; Sephton-Clark *et al.* 2023).

In this study, we leverage a cohort of 336 clinical *T. marneffei* isolates from patients living with HIV in northern and southern Vietnam (Le *et al.* 2017) to explore population diversity, determine levels of recombination, identify genomic regions under selection, and assess genomic predictors of clinical phenotypes. With this approach, we identify two geographically distinct populations showing similar levels of recombination within and between clades. Notably, we find that clinical aspects of disease severity differ between isolates from these two clades. We employ longitudinal samples isolated from blood and skin lesions to identify variants arising over the course of invasive infection. We observe five instances of disease recrudescence/relapse that are linked to the independent introduction of an unrelated isolate, in addition to the more frequent detection of persistent infection by the same isolate. We identify multiple genomic regions that appear to be under selection across these two populations, with copy number variation arising throughout the population, and multiple genetic variants showing a significant association with relevant clinical parameters including a high initial fungal burden in the blood, mortality, relapse, immune reconstitution inflammatory syndrome (IRIS), and poor treatment response.

Materials and methods
Sample collection and antifungal susceptibility testing
The 336 clinical *T. marneffei* isolates were obtained from cultures of blood and/or skin lesions of 272 patients from five hospitals across Vietnam who participated in the Itraconazole vs Amphotericin B for Talaromycosis (IVAP) trial conducted between October 2012 and December 2016 in Vietnam (Le *et al.* 2017). The in vitro antifungal susceptibility of all 336 *T. marneffei* yeast isolates against itraconazole and amphotericin B was determined using the M27-A3 reference method for testing yeasts as per the clinical and laboratory standards institute (CLSI) (Rex *et al.* 2008). The minimum inhibitory concentration (MIC) was determined by both visual and optical endpoints with growth inhibition of 90% for amphotericin B and 50% for itraconazole according to the CLSI. The study was approved by the ethics and scientific committees of the University of Oxford, all five participating hospitals in Vietnam, and Duke University. All patients gave informed consent for their *T. marneffei* clinical isolates to be stored and used for this research.

Clinical metadata
De-identified clinical metadata was collected and shared by the IVAP trial investigators. These data included a range of clinical measures collected during the IVAP trial and deemed to be non-identifiable. These measures include presence of fever, measures of baseline fungal burden in colony forming units (CFUs) per ml of blood, rates of fungal clearance in CFUs/ml/day, 24-week mortality, time to clinical resolution (defined as resolution of fever, skin lesions, and positive blood culture), incidence of relapse and IRIS, and respiratory failure over 24 weeks of follow-up. These patient outcome variables were originally defined in the IVAP trial (Le *et al.* 2017). All statistical analyses were performed in R 3.6.0.

Genomic DNA isolation and sequencing
DNA was purified using the yeast genomic DNA purification kit (VWR). Briefly, cells were disrupted with sodium dodecyl sulphate (SDS) (VWR) at 65°C. The pellet was then treated with ammonium acetate and RNase A (VWR). DNA was precipitated with isopropanol and stored in Tris-EDTA buffer at a pH of 8.0 (VWR). DNA purity was assessed via Nanodrop (Thermo Fisher Scientific) and 1% agarose gel electrophoresis. The DNA concentration was measured via Qubit fluorometer (Thermo Fisher Scientific). Initial library construction for the isolates entailed tagmentation via the Nextera XT DNA library preparation protocol (Illumina). Isolates were initially sequenced on a HiSeq 3000 (Illumina), generating 100-bp paired reads. Samples with poor (<10X) or biased coverage resulting from this first method then underwent additional sequencing, with libraries generated in accordance with the NEBNext Ultra II DNA Library Prep protocol (New England Biolabs). Libraries were sequenced on a HiSeq X10 (Illumina), generating 150-bp paired reads (minimum coverage 10X per sample).

Variant analysis
To identify variants for all samples, reads were aligned to the *T. marneffei* (11CN-03-130) reference genome (GCA_009650675.1) (Cuomo *et al.* 2020) with BWA-MEM version 0.7.17 (Li 2013). GATK v4 variant calling was carried out as documented in our publicly available cloud-based pipeline (Van der Auwera *et al.* 2013) (https://github.com/broadinstitute/fungal-wdl/tree/master/gatk4). Briefly, this includes identifying adapter sequences via GATK MarkIlluminaAdapters. The unaligned BAM files are

then aligned to the reference genome. Alignments are sorted, then duplicates are marked via GATK MarkDuplicates. Variants are called via GATK HaplotypeCaller, and variant sets are combined for sample cohorts. Post calling, variants were filtered on the following parameters: QD < 2.0, QUAL < 30.0, SOR > 3.0, FS > 60.0 (indels > 200), MQ < 40.0, GQ < 50, alternate allele percentage = 0.8, DP < 10. Variants were annotated with a custom annotation database generated for 11CN-03-130 in SNPeff, version 4.3t (Cingolani *et al.* 2012).

Population genomics and genome-wide association studies

A maximum likelihood phylogeny was estimated using 281,773 segregating SNP sites present in one or more isolates, allowing ambiguity in a maximum of 10% of samples, with RAxML version 8.2.12 GTRCAT rapid bootstrapping (1,000 replicates) (Stamatakis 2014). The phylogeny generated was visualized with ggtree (Yu *et al.* 2017). These SNP sites were also used to compute an unrooted phylogenetic network with splitstree4 (Huson and Bryant 2006). To assess support for clades identified via phylogenetic methods, MavericK performed estimation of K. Due to compute and convergence constraints, 100 iterations of rMavericK MCMC (Verity and Nichols 2016) was run for all samples, with a random subset of 1,000 variants selected per iteration. To assess recombination, linkage disequilibrium (LD) decay was estimated with vcftools version 0.1.16, based on all variants passing the filtering steps outlined above, for all isolates included in the population being measured (northern clade, southern clade, and all isolates). To do this, we calculated LD for 1000-bp windows, with a minimum minor allele frequency of 0.1, and the –hap-r2 option (Danecek *et al.* 2011). LD50 values were identified based on the number of bases at which the R^2 values, for each population, halved from the highest initial values at 1 kb. To identify regions of the genome under selection within these populations, PopGenome 2.7.5 was used to perform composite likelihood ratio (CLR) analysis, assess nucleotide diversity and divergence, and calculate Tajima's D and F-Statistic scores (per chromosome, by 5-kb windows) (Pfeifer *et al.* 2014). Copy number variation was calculated with CNVnator v0.3 (Abyzov *et al.* 2011). Association analysis between clinical data, in vitro phenotypes, and variants was carried out using PLINK v1.08p formatted files and Gemma version 0.94.1 (Zhou and Stephens 2012) (options: centered relatedness matrix gk 1, linear mixed model), as previously described (Desjardins *et al.* 2017).

Centromere identification

To identify centromere locations, the GC% was calculated across individual chromosomes with a sliding window approach in R 3.6.0. Repeat motifs across the proposed centromere regions were identified with RepeatModeler 2.0 (Flynn *et al.* 2020). Recombination rates across individual chromosomes were estimated with a random subset of 50 isolates spanning north and southern clades, via Ldhelmet 1.10 per the best practices workflow, which recommends using a window size of 50 SNPs, computing Padé coefficients, and calculating a mutation matrix (Chan *et al.* 2012).

Results

Characterization of genome structure

For this study, we leveraged a complete *T. marneffei* genome generated from isolate 11-CN-03-130. This assembly consists of eight chromosomes, with telomeric sequences present at both ends of all chromosomes excepting one ending in ribosomal DNA repeats (Cuomo *et al.* 2020). This haploid genome encodes 10,025 genes, including rRNA content on chromosome 3, with a BUSCO completeness score of 97.7% (Cuomo *et al.* 2020). To determine the location of the centromeric regions of these chromosomes, we calculated GC% across all chromosomes and detected single regions displaying notable reductions in GC% for each chromosome, corresponding to likely centromeric positions (Fig. 1a). To determine the composition of these proposed centromeric regions, we identified repetitive sequences based both on similarity to known elements and de novo self-alignments, identifying long interspersed nuclear elements (LINE) and DNA/TcMar-Fot1 transposons throughout these putative centromeric regions (Fig. 1b). The repetitive content across these regions is consistent with the repetitive nature of centromere regions observed across other species of filamentous fungi (Smith *et al.* 2012). These proposed regions range from 15.5 to 28 kb in length, with an average centromere length for all chromosomes of 23 kb. To assess whether rates of recombination across these proposed centromeric regions are lower when compared to the non-centromeric regions, we calculated recombination rates per chromosome, with variant information from a subset of 50 isolates sequenced in this study. We saw similar rates of recombination across chromosomes (Supplementary Figure 1), and while we did not observe any reduction in recombination rate that might be associated with centromeric locations, the proposed centromeric regions displayed rates that were in line with those observed across the entire chromosome.

Population structure and geographical clades

To finely characterize genetic variation of clinical populations, we compared the genomes of *T. marneffei* isolated from HIV-positive patients enrolled in the IVAP trial across northern and southern Vietnam (Le *et al.* 2017). All individuals were HIV-positive and had *T. marneffei* isolates obtained from blood cultures taken at enrollment (Day 0). For a subset of individuals, *T. marneffei* isolates were obtained from both cultures taken at enrollment (Day 0), and from longitudinal time points for both blood and skin sites. Longitudinal samples were collected between Day 4 and Day 240, with the median longitudinal sample collection time being 1 month. We performed whole genome sequencing on all isolates collected, requiring a minimum genome coverage of 10X. We then called variants for these isolates, using a reference genome assembly (Cuomo *et al.* 2020) of isolate 11CN-03-130, also collected from southern Vietnam as part of the IVAP trial.

To assess population structure, we generated a maximum likelihood phylogeny from segregating SNP site information (Fig. 2). Isolates clearly separate into two clades based on bootstrap support, SplitsTree analysis, and K = 2 determination via MavericK for this population. There is a significant difference between evidence scores for K = 1 and K = 2, but overlapping distributions (IQR) of -$\log_2$ evidence scores for K = 2 and K = 3 (K = 2: Q1 −18347, Q3 −17,632. K = 3: Q1 −17901, Q3 −17,337). These two clades represent the northern and southern regions of Vietnam from which they were isolated. This pattern is consistent with prior geographical substructure with clade separation based on region of collection (Henk *et al.* 2012). Clinical and in vitro susceptibility phenotypes appear well distributed across the phylogeny for these isolates (Fig. 2, Supplementary Figure 2). Of these isolates, 168 harbor the MAT1-2 locus, found across both northern and southern clades. Northern isolates show slightly higher levels of diversity (Table 1) and possess significantly longer terminal branch lengths when compared to southern clade isolates (Mann–Whitney U-Test *P* < 0.00001).
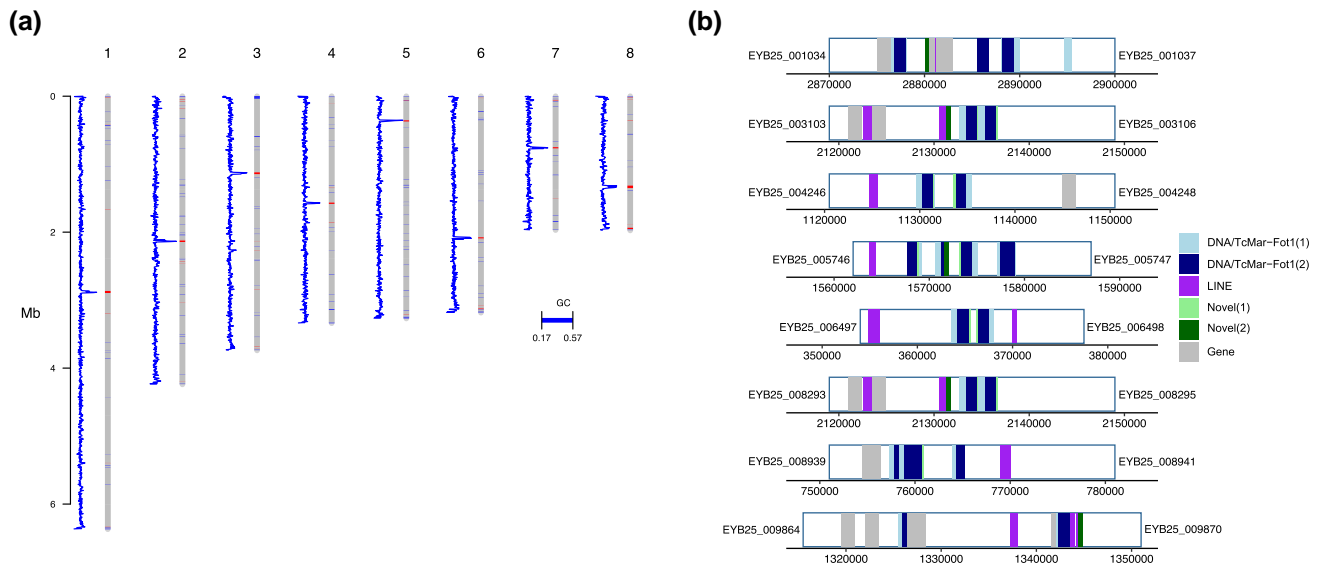
Fig. 1. *T. marneffei* centromere identification. A sliding window analysis of GC% and repetitive element content was used to predict candidate centromere locations. a) GC% for each chromosome (line plot), calculated and plotted in 500-bp windows beside each chromosome (bars). Regions of low GC (17–40%) are shown as bands on each chromosome. b) Proposed centromere regions with repetitive elements colored by color representing class, with the flanking gene IDs listed on either side (EYB25 prefix).
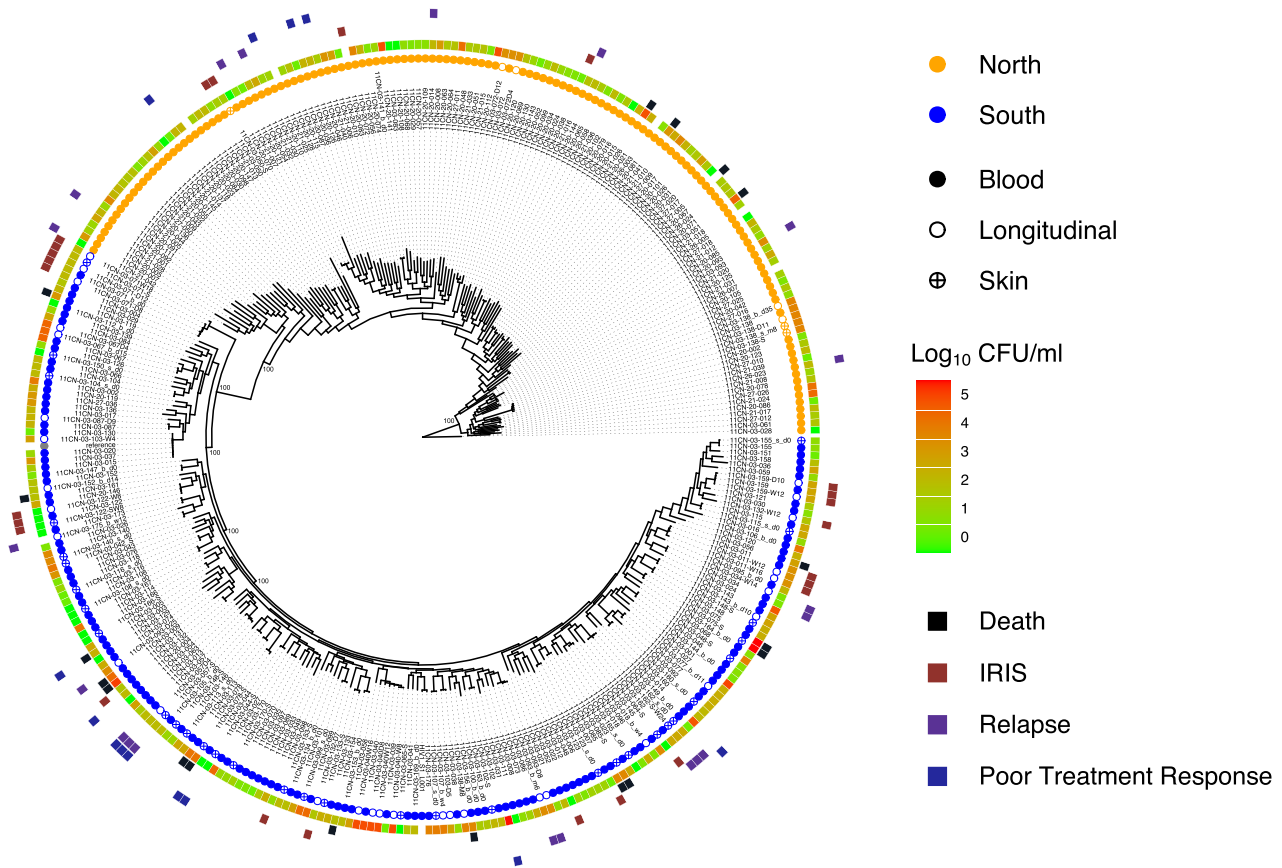

Fig. 2. Phylogeny of patient isolates reveals no genetic clusters of severe outcomes. A maximum likelihood phylogeny was estimated using segregating SNP sites. Isolates separate distinctly into northern and southern clades, with both clades having 100% bootstrap support. The colored circles correspond to clade and sample type. Metadata for the clinical isolates includes initial fungal burden for a patient, indicated by the Log_{10} CFU/ml heatmap, and patient outcomes that are indicated by colored squares in the outer perimeter of the phylogeny.

Table 1. Genomic and population characteristics of isolates sampled.

	Northern Vietnam	Southern Vietnam
Sample size (baseline blood, skin, and longitudinal)	141	195
LD50	12 kb	12 kb
π	0.00077	0.00051

To assess recombination levels between and across these two clades, we calculated LD decay for the northern clade, the southern clade, and all isolates combined (Supplementary Figure 3 and Table 1). We see similar rates of decay, and identical LD50 values (12 kb) across all groups, indicating that neither clade is recombining exclusively within their group.

Clinical outcomes and longitudinal samples

To assess whether clinical phenotypes or in vitro antifungal susceptibility values significantly differed by clade, we performed a series of statistical analyses to assess associations between clade and phenotype. We did not observe statistically significant differences between clades in considering 24-week mortality, rates of relapse, IRIS, or MICs of itraconazole and amphotericin B. However, some striking differences were detected. While isolates from the southern clade have comparable rates of fungal clearance to isolates from the northern clade (Supplementary Figure 4a), the southern isolates exhibit significantly higher baseline fungal burden (Supplementary Figure 4b) (Wilcoxon rank sum test, $P = 0.024$), significantly higher prevalence of fever (Fisher exact probability test, $P = 0.0002$), and significantly decreased incidence rates of respiratory failure (Fisher exact probability test, 0.0021) (Table 2). Given the increased presence of fever and baseline fungal burden across isolates from the southern clade, we sought to assess whether additional factors including patient age, antiretroviral (ARV) treatment status, or CD4 counts were confounders in these analyses. For patients included in this analysis, the median age was 33, the median CD4 count was 10, and 44% of patients had previously received ARV treatment. Neither age, ARV treatment status, or CD4 counts were significantly correlated with the presence of fever or baseline fungal burden, indicating that isolate clade is a contributing factor to some aspects of disease severity in patients from Vietnam.

To assess the genetic relationships between isolates sampled from different body sites (skin vs blood) of a patient at the same time point, and isolates sampled at different time points from the same patient (pre vs post-treatment), we calculated the number of SNP differences between each isolate. Over 80% of isolates sampled at longitudinal time points show a small number of SNP differences (≤ 10) when compared to their baseline counterparts (Tables 3 and 4), indicating that these longitudinal samples represent the original infecting isolates, having undergone a small number of genetic changes during infection. Some of the genes that were impacted by SNPs arising during infection include a predicted Myb-like binding domain (EYB25_007616), a transcription factor (EYB25_005954), an AMP-binding protein (EYB25_000740), an ORC DNA replication protein (EYB25_004486), and a cytochrome P450 (EYB25_003095). We performed additional tests to assess differences in the MICs between incident and post-drug exposure isolates. We observed a general trend of longitudinal isolates maintaining (62%) or increasing (12%) MIC values over time when compared to their baseline counterparts.

Table 2. Frequency of clinical phenotypes associated with northern and southern clades.

Clinical phenotypes	Northern Vietnam	Southern Vietnam	P value[a]
Median baseline fungal burden (Log_{10} CFU/ml)	2.1	2.5	0.024
Median rate of clearance (EFA)	0.429	0.347	0.15
Fever	10	34	0.0002
Respiratory failure	45	24	0.0021

[a] Fishers exact test and Wilcoxon rank sum test were used to calculate P values.

We also observed that a subset of longitudinal isolates from four patients appear unrelated to the primary infecting isolate, based on phylogenetic placement and SNP differences. In total, we observed four longitudinal blood isolates and two contemporaneous skin isolates (Table 5) that appeared unrelated to the initial blood isolate, or matched time point blood isolate, respectively. With SNP differences between these isolates and their initial or blood counterparts ranging from 1,072 to 2,810 (Table 5), it is likely that these cases represent novel strain introductions or pre-existing multistrain infections, as opposed to a persistent infection of the same isolate over time, suggesting that patients may be exposed to multiple strains of *Talaromyces* prior to and during active infections, resulting in multistrain infections.

Genomic variants associated with clinical outcomes

Given the range of clinical phenotypes and outcomes associated with isolates from both clades, we leveraged this variability to assess significant associations between clinical phenotypes and pathogen genetics. We took a genome-wide association study (GWAS) approach to test for significance between loss-of-function (LoF) (frameshift and stop-codon gain) mutations and the clinical metadata suggestive of severe infections. Four patients with multiple infecting strains were excluded from this analysis. We saw LoF variants associate significantly ($P < 5 \times 10^{-6}$, gemma score test) with poor clinical outcomes including mortality, relapse, IRIS, poor clinical response, and high baseline fungal burden (CFU/ml) (Table 6). These variants impacted genes with predicted roles regulating gene expression (methyltransferase and GAL4 transcription factor homolog), metabolism and fermentation (alcohol acetyltransferase, GTPase activating protein, and malate synthase), cellular signaling (calcineurin-like phosphoesterase), protein processing (oligosaccharyltransferase, ubiquitin-protein ligase, and exocyst complex component Sec3), and transport (HCO^3 transporter, sugar transporter, and major facilitator superfamily transporter). These isolates also varied in the MIC of itraconazole and amphotericin B, with a small number of isolates displaying notably elevated MIC values (Supplementary Figure 5). While a GWAS did not identify any LoF variants significantly associated with higher MICs to these antifungals, there was very limited power in this analysis due to the low number of isolates with high MIC values. Genetic variation across a population may not only contribute to virulence but may also be selected for through environmental pressures. The variants identified here may give some advantage to survival within the environment of the human host and may be working in tandem with other forms of genetic variation, such as copy number, to elicit more severe clinical phenotypes.

Table 3. Genomic changes between related isolates taken from the same patients over time, or from different body sites at the same time point.

Baseline	Longitudinal	Time (days)	SNPs	Gene-impacted
11CN-03-011	11CN-03-011-W12, 11CN-03-011-W16	84, 112	0	
11CN-03-018	11CN-03-018_b_w4	28	0	
11CN-03-034	11CN-03-034-W14	98	0	
11CN-03-040	11CN-03-040D8, 11CN-03-040-W8, 11CN-03-040-W12, 11CN-03-040-W16	8, 56, 84, 112	0	
11CN-03-057	11CN-03-057D4	4	0	
11CN-03-067	11CN-03-067D4	4	**1**	EYB25_004707
11CN-03-067	11CN-03-067_b_d15	15	0	
11CN-03-071_b_d0	11CN-03-071-D8	8	**3**	EYB25_004793, EYB25_007616
11CN-03-071_b_d0	11CN-03-071-D12	12	**2**	EYB25_004798, EYB25_004800
11CN-03-071_b_d0	11CN-03-071W18	126	7	
11CN-03-072	11CN-03-072D4	4	2	
11CN-03-072	11CN-03-072-D12	12	0	
11CN-03-077	11CN-03-077_b_d11	11	0	
11CN-03-087	11CN-03-087-D9	9	**1**	EYB25_005954
11CN-03-103-D5	11CN-03-103-W4	23	0	
11CN-03-122	11CN-03-122-W8	56	0	
11CN-03-138	11CN-03-138-D11, 11CN-03-138_b_d35	11, 35	0	
11CN-03-138-S	11CN-03-138_s_m8	240	0	
11CN-03-152	11CN-03-152_b_d14	14	**2**	EYB25_000740, EYB25_004486
11CN-03-159	11CN-03-159-D10	10	1	
11CN-03-159	11CN-03-159-W12	84	1	
11CN-03-162_s_d0	11CN-03-162-S-W24	168	**3**	EYB25_002369, EYB25_003095, EYB25_003335

Coding SNPs are bolded.

Table 4. Genomic changes between related isolates taken from the same patients at different body sites.

Blood	Skin	Time (days)	SNPs
11CN-03-044	11CN-03-044-S	0	0
11CN-03-046	11CN-03-046-S	0	0
11CN-03-050	11CN-03-050-S	0	0
11CN-03-057	11CN-03-057-S	0	3
11CN-03-071	11CN-03-071-S	0	2
11CN-03-075	11CN-03-075-S	0	0
11CN-03-078	11CN-03-078-S	0	0
11CN-03-085	11CN-03-085-S	0	1
11CN-03-099	11CN-03-099_s_d0	0	0
11CN-03-102	11CN-03-102-S	0	0
11CN-03-104	11CN-03-104_s_d0	0	0
11CN-03-108	11CN-03-108_s_d0	0	0
11CN-03-113	11CN-03-113_s_d0	0	1
11CN-03-115	11CN-03-115_s_d0	0	1
11CN-03-116	11CN-03-116_s_d0	0	1
11CN-03-122-W8	11CN-03-122-SW8	0	0
11CN-03-123	11CN-03-123_s_d0	0	0
11CN-03-133	11CN-03-133-S	0	10
11CN-03-138	11CN-03-138-S	0	0
11CN-03-140	11CN-03-140_s_d0	0	0
11CN-03-146	11CN-03-146-S	0	0
11CN-03-148	11CN-03-148-S	0	0
11CN-03-155	11CN-03-155_s_d0	0	1
11CN-03-160	11CN-03-160_s_d0	0	1
11CN-03-162	11CN-03-162_s_d0	0	0
11CN-03-166	11CN-03-166-S	0	1
11CN-03-170	11CN-03-170-S	0	1

Table 5. SNP differences in unrelated isolates taken from the same patient sets.

Baseline	Longitudinal	Time (days)	SNPs
11CN-03-093	11CN-03-093-D8	8	2764
11CN-03-093	11CN-03-093-D25	25	2810
11CN-03-093	11CN-03-093_b_m6	180	2801
11CN-03-132	11CN-03-132-D7	7	1427
11CN-03-132	11CN-03-132-W12	84	1614
11CN-03-138	11CN-03-138-M8	240	2721
Blood	Skin		
11CN-03-138-M8	11CN-03-138_s_m8	0	2142
11CN-03-153	11CN-03-153-S	0	1072

These unrelated isolates did not co-localize across the phylogeny.

Population diversity and copy number variation

To investigate larger scale genetic variation, we performed copy number variation analysis. We observed no evidence of chromosomal aneuploidy across these isolates but identified multiple genomic regions that appeared duplicated and deleted within this population, relative to the reference genome. Many of these regional duplications and deletions appeared unique to individual isolates (Fig. 3) with most isolates possessing both deletions and duplications across different loci; however, common deletions were also detected. The most commonly deleted region encodes a predicted afadin- and alpha-actinin-binding region that is deleted across 141 isolates from the northern clade and 125 isolates from the southern clade, relative to the reference strain. The largest number of isolates sharing a duplication across a single locus is 17, duplicated in isolates from the southern clade (this region encodes two reverse transcriptases) and a GAG-pre-integrase domain, likely corresponding to a retrotransposon integration. Of those regions deleted relative to the reference strain that are shared across more than 50 isolates, these encode genes predicted to function as transcription factors (six genes), REDOX proteins (five genes), protein kinases (four genes), methyltransferases (three genes), aspartic proteases (two genes), dehydrogenases (two genes), a cytochrome P450, a glycosyl hydrolase, a glycosyl transferase, a phospholipase, a phosphorylase, an adenylosuccinate lyase, an actin binding protein, a decarboxylase, an aminotransferase, an apoptosis regulating protein, an autophagy

Table 6. Variants significantly associated with clinical phenotypes.

Phenotype[a]	Gene	P value	Gene function (Pfam)	Loss-of-function variant type
Relapse	EYB25_006356	5.40E-07	Metallophosphoesterase	Stop gained
Relapse	EYB25_007086	1.39E-06	Sec3	Stop gained
Relapse	EYB25_007995	1.54E-06		Stop gained
Relapse	EYB25_006573	2.15E-06	Major facilitator superfamily	Frameshift
Mortality	EYB25_006471	4.44E-08		Stop gained
Mortality	EYB25_009504	2.50E-07		Frameshift
IRIS	EYB25_004020	2.37E-12	Methyltransferase	Frameshift
IRIS	EYB25_007265	8.05E-11	AEX-3	Stop gained
IRIS	EYB25_009427	6.71E-10	Alcohol acetyltransferase	Stop gained
IRIS	EYB25_009447	4.68E-08	Transcription factor	Frameshift
Poor clinical response	EYB25_006915	1.73E-18		Stop gained
Poor clinical response	EYB25_006356	2.93E-09	Metallophosphoesterase	Stop gained
Poor clinical response	EYB25_004962	3.84E-08	Oligosaccharyltransferase	Stop Gained
Poor clinical response	EYB25_001778	4.96E-08	Malate synthase	Stop gained
Poor clinical response	EYB25_004898	4.22E-07	HCO3⁻ transporter	Stop gained
Poor clinical response	EYB25_008150	6.30E-07	Sugar transporter	Frameshift
Poor clinical response	EYB25_006081	1.06E-06	Haloacid dehalogenase hydrolase	Frameshift
Poor clinical response	EYB25_007654	2.32E-06		Frameshift
High CFU	EYB25_007106	4.56E-06	Ubiquitin transferase	Stop gained

[a]Significant variants ($P < 5 \times 10^{-6}$) identified through multiple GWAS tests are combined in this table. Column 1 lists the phenotype that the variant is significantly associated with.
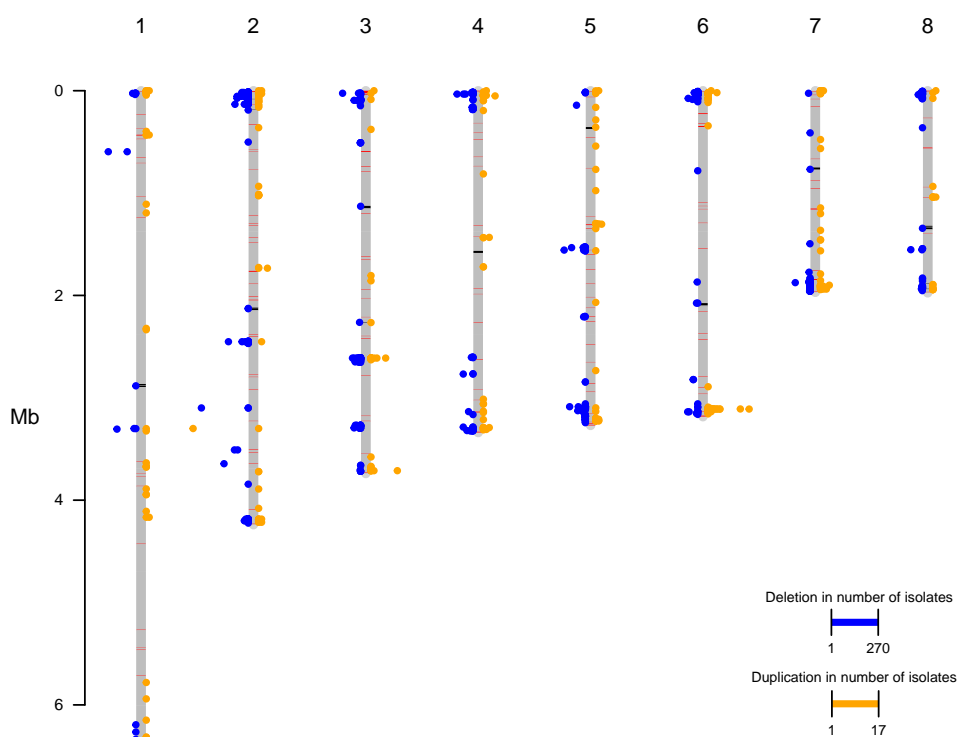

Fig. 3. Frequencies of regional genomic duplications and deletions per chromosome. Frequency of deleted regions (left of chromosome axis) and duplicated regions (right of chromosome axis) assessed across all isolates, with frequency of deletion or duplication represented as distance from axis of the chromosomes (bars). 1000-bp windows were used to assess copy number variation. GC content across chromosomes represented by black (GC 17–40%) and red (GC 50–57%) bars plotted on the chromosome bars.

regulating protein, an arrestin, and a meiotic nuclear division protein. The range in functionality of genes lost across more than 50 isolates highlights the diversity within this clinical dataset. Of those duplicated regions shared across more than two isolates, these can be characterized by their predicted functions as heat shock and stress response genes, including heat shock protein 70; cell membrane and wall modulating genes, including a chitin synthase and a phosphoylcholine transferase; metabolic regulatory genes, including a cytochrome P450, a thiolase, a glycosyl hydrolase, an aspartyl protease, and a citrate synthase; and gene expression regulatory genes, including a transcription factor and an mRNA capping protein. These results suggest that genes with roles in heat shock, cell wall structure, and metabolic response could provide an advantage to *T. marneffei* in this setting.
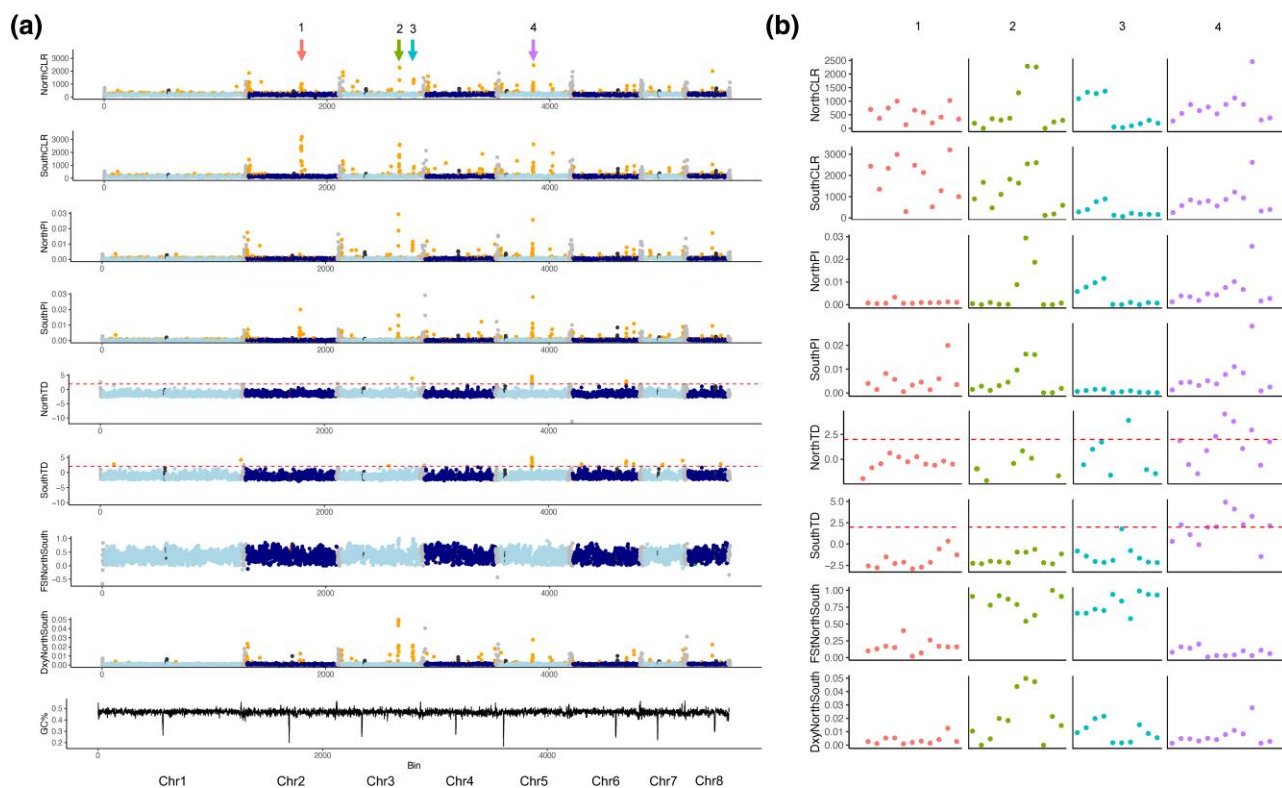
Fig. 4. Genomic regions under selection across northern and southern clades. a) Signals of selection (CLR and TD), nucleotide diversity (Pi), and divergence (Fst and Dxy) across all chromosomes, per clade. b) Enlarged view of regions displaying elevated signals of selection, including region 1–4.

We did not observe any duplications of the CYP51 ortholog within these isolates, suggesting alternate mechanisms may be responsible for the range in MIC values seen to itraconazole.

Population genomics identify regions under selection

To assess regions that may be under selection within these populations, we performed a population genetics analysis to identify rapidly evolving regions across these isolates. We noted strong signals present across many telomeric and sub-telomeric regions (Fig. 4a), in tests for selective pressure (CLR) (Nielsen *et al.* 2005), nucleotide diversity (Pi), and population divergence (Dxy). To determine if these regions encode common functions, we performed enrichment analysis on all genes within the sub-telomeric regions. The sub-telomeric regions displaying strong signals for selective pressure were enriched (hypergeometric test, $P < 0.05$) for genes predicted to play roles in thiamine diphosphate metabolism and energy derivation by oxidation of organic compounds (Supplementary Table 1). When northern and southern clades were assessed independently via CLR analysis, non-sub-telomeric regions exhibiting strong signals for selective pressure across both populations were enriched (hypergeometric test, $P < 0.05$) for genes predicted to function in ion and organic anion transport, respiration, regulation of gene expression, and energy derivation by oxidation of organic compounds (Supplementary Table 1). Non-sub-telomeric regions displaying signals of population divergence between clades were enriched (hypergeometric test, $P < 0.05$) for genes predicted to function in anaerobic respiration (Supplementary Table 1).

To assess specific genes under positive selective pressures across these clades, we performed a CLR analysis and found multiple regions with strong signals of selection across northern and southern clades, of which we will highlight four regions of interest (Fig. 4b). Region 1 is under selection across isolates in both clades, but with relatively lower nucleotide diversity and CLR scores in the smaller northern clade (Fig. 4a, region 1). This region shows lower levels of divergence between clades (Fst and Dxy), and appears to be rapidly evolving across all isolates with little clade specificity. This region encodes genes predicted to function as transcription factors (three genes), protein kinases (three genes), a ferric reductase, a methyltransferase, an adenylosuccinate lyase, and a phospholipase. Two additional regions displayed strong selective scores (CLR) across both clades in addition to elevated scores of divergence between clades (Fst and Dxy) (Fig. 4a, regions 2 and 3). While region 2 shows high levels of nucleotide diversity for both clades, region 3 shows increased nucleotide diversity restricted to the northern clade, accompanied by an increased Tajima's *D* score for this population. When region 2 and region 3 are interrogated further (Fig. 4b), these two regions divide based on diversity within clade, with tracts of low diversity (Pi) within clades but high divergence (Dxy and Fst) observed between clades, indicating the potential for clade-specific adaptations in these regions. Region 2 encodes genes predicted to function as a cytochrome P450 and a glycosyl hydrolase. Region 3 encodes genes predicted to function as a transcription factor, a protein kinase, an aspartyl protease, and a pyridoxal-dependent decarboxylase. To determine whether any of these signals may be due to a recent selective sweep, or balancing selection, we assessed scores for Tajima's D. We identified one region that shared a strong signal in both our CLR analysis and our Tajima's D analysis (Fig. 4a, region 4), indicating that this region may have recently undergone a selective sweep. We note that this region also shows an elevated

Dxy score, but no increase in Fst between the two populations, indicating increased levels of divergence between the two populations at this region, accompanied by high levels of within-population diversity, as reflected in our measurements of nucleotide diversity (Pi) for each clade at this region. This region encodes 14 genes and multiple ankyrin repeats (EYB25_006920-006938) in total. The predicted functions of these genes include a transcription factor, an endoplasmic reticulum membrane protein, an arrestin, a meiotic nuclear division protein, a glycosyl transferase, a cell surface protein, an alcohol dehydrogenase, and an autophagy protein. These regions feature genes that may be playing important roles across and between clades, highlighting candidate genes for future functional analysis.

Discussion

This large-scale genomic study of *T. marneffei* has begun to shed light on the population variability that exists within a large endemic area of talaromycosis. Phylogenetic analysis of 336 clinical isolates revealed a population of two major clades, consistent with a previous MLST study that identified geographically structured clades of *T. marneffei* (Henk *et al.* 2012). The clades representing both northern and southern Vietnam identified through MLST are recapitulated here using WGS based SNP phylogenies. Across both clades, the MAT1-2 locus is present, occurring in 50% of isolates; however, we are unable to confirm the presence of the MAT1-1 locus in the non-MAT1-2 isolates, due to the reference genomes used in this study (Cuomo *et al.* 2020) possessing the MAT1-2 locus. We calculate similar rates of LD decay for both clades individually as for together, indicating that recombination is taking place at similar levels within and between clades. Southern isolates show slightly increased clonality, harboring alleles with fewer SNPs and displaying significantly shorter terminal branches than isolates from the northern clade, highlighting the possibility of a more recent introduction.

Leveraging the clinical metadata available for these isolates, we investigated the possibility of clade-based differences with regards to clinical phenotypes and outcomes. We did not observe significant differences between the northern and southern clades with respect to 24-week mortality, rates of relapse, IRIS, clinical response to treatment, or MICs of itraconazole and amphotericin B. However, we detected significant associations between the southern clade and aspects of disease severity, including higher baseline fungal burden and presence of fever. In addition, we identified multiple loss-of-function variants significantly associated with mortality, relapse, IRIS, poor clinical response, and high baseline fungal burden. We noted that many of these clinical phenotypes may occur at low rates for the group sizes, likely impacting our ability to detect both significant differences between clades, and significant associations between variants and phenotypes. In addition, there are many confounding factors that contribute to these clinical outcomes, including host immune status, host response, timeliness of treatment, and antifungal treatment regimen. While we did not find that age, CD4 count, or ARV treatment status confounded associations between the southern clade and higher baseline fungal burden or presence of fever, larger studies of talaromycosis outcomes should be completed to provide increased power for estimations of clinical differences between clades to confirm these findings. Further work using larger group sizes for genotype-phenotype association studies will also be essential for building on our GWAS findings.

Applying a range of population genetics approaches with the variant data generated for these clinical isolates, we interrogated signals of selective pressure, increased nucleotide diversity, and

population divergence for these clades. We observed shared signals of rapidly evolving regions present in northern and southern clades, with the strongest shared signals spanning regions containing genes with roles in gene expression regulation, iron sequestration, cellular signaling, and metabolism. Regions exhibiting signals of divergence between populations encoded genes with roles in metabolism and adaptation to environmental niches, sugar metabolism, transcriptional regulation, and protein cleavage. Of note, we observed strong signals of selection at the telomeric and subtelomeric regions, in line with observations for other fungal pathogens such as *Cryptococcus neoformans* (Desjardins *et al.* 2017; Sephton-Clark *et al.* 2022). While sugar transporters are known to be enriched in these regions of *C. neoformans*, the subtelomeric regions displaying signals of selection in *T. marneffei* are enriched for genes involved in energy derivation by oxidation of organic compounds, and thiamine diphosphate metabolism.

The phylogenetic analysis of both longitudinal samples and samples obtained from multiple body sites allowed for the inference of infections resulting from single, or multiple infecting isolates. We observed six instances in which longitudinal isolates obtained from a patient appeared unrelated to the primary infecting isolate, hinting at novel strain introductions, or the presence of multiple infecting strains. This is the first time this approach has been applied to *T. marneffei* to determine whether patients with protracted talaromycosis or infection recrudescence have been re-infected with novel isolates. This approach could also enable the identification of patients that have been chronically exposed, resulting in multiple contained primary infections that break up in the event of becoming immunocompromised. Sequencing of additional isolates at each time point would enable increased resolution to explore this possibility. The majority of patients that underwent longitudinal isolate collection showed continued infection with the basal isolate, in line with similar observations for *C. neoformans* (Chen *et al.* 2017; Rhodes, Beale, *et al.* 2017).

Combining whole genome sequencing data and patient data for this clinical cohort has enabled a better understanding of the relatedness of infecting isolates, within individual patients, and within the broader population. This approach enables identification of pathogen loci that appear under selection across both geographic clades, and highlights multiple genes that have been repeatedly, and independently, lost or duplicated across these isolates. Patient metadata allowed for association analyses that have given us insights into genetic signatures of the pathogen that are significantly associated with specific patient outcomes. Combining data types such as these, on an ever-larger scale, will afford a better understanding of the genetic basis of pathogenicity in fungal pathogens such as *T. marneffei*.

Data availability

All sequencing data are available in NCBI via the accession PRJNA949141.

Supplemental material available at GENETICS online.

Acknowledgements

The authors thank the Broad Institute Genomics Platform for sequencing isolates as part of this study.

Funding

Conflicts of interest statement

The author(s) declare no conflict of interest.

Literature cited

Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. Genome Res. 2011;21(6):974–984. doi:10.1101/gr.114876.110.

Chan AH, Jenkins PA, Song YS. Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster.* PLoS Genet. 2012;8-(12):e1003090. doi:10.1371/journal.pgen.1003090.

Chariyalertsak S, Sirisanthana T, Supparatpinyo K, Praparattanapan J, Nelson KE. Case-control study of risk factors for *Penicillium marneffei* infection in human immunodeficiency virus-infected patients in northern Thailand. Clin Infect Dis. 1997;24(6): 1080–1086. doi:10.1086/513649.

Chen Y, Farrer RA, Giamberardino C, Sakthikumar S, Jones A, Yang T, Tenor JL, Wagih O, Van Wyk M, Govender NP, *et al.* Microevolution of serial clinical isolates of *Cryptococcus neoformans* var. *Grubii* and c. *Gattii.* mBio. 2017;8(2):e00166-17. doi:10.1128/mBio.00166-17.

Chow NA, Muñoz JF, Gade L, Berkow EL, Li X, Welsh RM, Forsberg K, Lockhart SR, Adam R, Alanio A, *et al.* Tracing the evolutionary history and global expansion of *Candida auris* using population genomic analyses. mBio. 2020;11(2):e03364-19. doi:10.1128/mBio.03364-19.

Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w^{1118}; iso-2; iso-3. Fly (Austin). 2012;6(2):80–92. doi:10.4161/fly.19695.

Cuomo CA, Shea T, Nguyen T, Ashton P, Perfect J, Le T. Complete genome sequences for two *Talaromyces marneffei* clinical isolates from northern and southern Vietnam. Microbiol Resour Announc. 2020;9(2):e01367-19. doi:10.1128/MRA.01367-19.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, *et al.* The variant call format and VCFtools. Bioinformatics. 2011;27(15): 2156–2158. doi:10.1093/bioinformatics/btr330.

Desjardins CA, Giamberardino C, Sykes SM, Yu C-H, Tenor JL, Chen Y, Yang T, Jones AM, Sun S, Haverkamp MR, *et al.* Population genomics and the evolution of virulence in the fungal pathogen *Cryptococcus neoformans.* Genome Res. 2017;27(7):1207–1219. doi: 10.1101/gr.218727.116.

Ene IV, Farrer RA, Hirakawa MP, Agwamba K, Cuomo CA, Bennett RJ. Global analysis of mutations driving microevolution of a heterozygous diploid fungal pathogen. Proc Natl Acad Sci USA. 2018; 115(37):E8688–E8697. doi:10.1073/pnas.1806002115

Fang L, Liu M, Huang C, Ma X, Zheng Y, Wu W, Guo J, Huang J, Xu H. MALDI-TOF MS-based clustering and antifungal susceptibility tests of *Talaromyces marneffei* isolates from Fujian and Guangxi (China). IDR. 2022;15(1):3449–3457. doi:10.2147/IDR.S364439

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. Repeatmodeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci USA. 2020;117(17): 9451–9457. doi:10.1073/pnas.1921046117.

Fu MS, Liporagi-Lopes LC, Dos Santos SR, Tenor JL, Perfect JR, Cuomo CA, Casadevall A. Amoeba predation of *Cryptococcus neoformans* results in pleiotropic changes to traits associated with virulence. mBio. 2021;12(2):e00567-21. doi:10.1128/mBio.00567-21.

Gerstein AC, Jackson KM, McDonald TR, Wang Y, Lueck BD, Bohjanen S, Smith KD, Akampurira A, Meya DB, Xue C, *et al.* Identification of pathogen genomic differences that impact human immune response and disease during *Cryptococcus neoformans* infection. mBio. 2019;10(4):e01440-19. doi:10.1128/mBio.01440-19.

Helmstetter N, Chybowska AD, Delaney C, Da Silva Dantas A, Gifford H, Wacker T, Munro C, Warris A, Jones B, Cuomo CA, *et al.* Population genetics and microevolution of clinical *Candida glabrata* reveals recombinant sequence types and hyper-variation within mitochondrial genomes, virulence genes, and drug targets. Genetics. 2022;221(1):iyac031. doi:10.1093/genetics/iyac031.

Henk DA, Shahar-Golan R, Devi KR, Boyce KJ, Zhan N, Fedorova ND, Nierman WC, Hsueh P-R, Yuen K-Y, Sieu TPM, *et al.* Clonality despite sex: the evolution of host-associated sexual neighborhoods in the pathogenic fungus *Penicillium marneffei.* PLoS Pathog. 2012;8(10):e1002851. doi:10.1371/journal.ppat.1002851.

Huson DH, Bryant D. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol. 2006;23(2):254–267. doi:10.1093/molbev/msj030.

Jiang J, Meng S, Huang S, Ruan Y, Lu X, Li JZ, Wu N, Huang J, Xie Z, Liang B, *et al.* Effects of *Talaromyces marneffei* infection on mortality of HIV/AIDS patients in southern China: a retrospective cohort study. Clin Microbiol Infect. 2019;25(2):233–241. doi:10.1016/j.cmi.2018.04.018.

Lam W-H, Sze K-H, Ke Y, Tse M-K, Zhang H, Woo PCY, Lau SKP, Lau CCY, Xu S, Lai P-M, *et al. Talaromyces marneffei* Mp1 protein, a novel virulence factor, carries two arachidonic acid-binding domains to suppress inflammatory responses in hosts. Infect Immun. 2019; 87(4):e00679-18. doi:10.1128/IAI.00679-18.

Le T, Kinh NV, Cuc NTK, Tung NLN, Lam NT, Thuy PTT, Cuong DD, Phuc PTH, Vinh VH, Hanh DTH, *et al.* A trial of itraconazole or amphotericin B for HIV-associated talaromycosis. N Engl J Med. 2017; 376(24):2329–2340. doi:10.1056/NEJMoa1613306.

Le T, Wolbers M, Chi NH, Quang VM, Chinh NT, Huong Lan NP, Lam PS, Kozal MJ, Shikuma CM, Day JN, *et al.* Epidemiology, seasonality, and predictors of outcome of AIDS-associated *Penicillium marneffei* infection in Ho Chi Minh City, Viet Nam. Clin Infect Dis. 2011;52(7):945–952. doi:10.1093/cid/cir028.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio].

Muñoz JF, Welsh RM, Shea T, Batra D, Gade L, Howard D, Rowe LA, Meis JF, Litvintseva AP, Cuomo CA. Clade-specific chromosomal rearrangements and loss of subtelomeric adhesins in *Candida auris.* Genetics. 2021;218(1):iyab029. doi:10.1093/genetics/iyab029.

Narayanasamy S, Dat VQ, Thanh NT, Ly VT, Chan JF-W, Yuen K-Y, Ning C, Liang H, Li L, Chowdhary A, *et al.* A global call for talaromycosis to be recognised as a neglected tropical disease. Lancet Glob Health. 2021a;9(11):e1618–e1622. doi:10.1016/S2214-109X(21)00350-8.

Narayanasamy S, Dougherty J, van Doorn HR, Le T. Pulmonary talaromycosis: a window into the immunopathogenesis of an endemic mycosis. Mycopathologia. 2021b;186(5):707–715. doi:10.1007/s11046-021-00570-0.

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. Genome Res. 2005;15(11):1566–1575. doi:10.1101/gr.4252305.

O'Brien CE, Oliveira-Pacheco J, Cinnéide EÓ, Haase MAB, Hittinger CT, Rogers TR, Zaragoza O, Bond U, Butler G. Population genomics of the pathogenic yeast *Candida tropicalis* identifies hybrid isolates in environmental samples. PLoS Pathog. 2021;17(3): e1009138. doi:10.1371/journal.ppat.1009138.

Ormerod KL, Morrow CA, Chow EWL, Lee IR, Arras SDM, Schirra HJ, Cox GM, Fries BC, Fraser JA. Comparative genomics of serial isolates of *Cryptococcus neoformans* reveals gene associated with carbon utilization and virulence. G3 (Bethesda). 2013;3(4):675–686. doi:10.1534/g3.113.005660.

Pfeifer B, Wittelsbürger U, Ramos-Onsins SE, Lercher MJ. Popgenome: an efficient Swiss army knife for population genomic analyses in R. Mol Biol Evol. 2014;31(7):1929–1936. doi:10.1093/molbev/msu136.

Poláková S, Blume C, Zárate JÁ, Mentel M, Jørck-Ramberg D, Stenderup J, Piškur J. Formation of new chromosomes as a virulence mechanism in yeast *Candida glabrata*. Proc Natl Acad Sci USA. 2009;106(8):2688–2693. doi:10.1073/pnas.0809793106.

Pongpom M, Sawatdeechaikul P, Kummasook A, Khanthawong S, Vanittanakom N. Antioxidative and immunogenic properties of catalase-peroxidase protein in *Penicillium marneffei*. Med Mycol. 2013;51(8):835–842. doi:10.3109/13693786.2013.807445.

Pruksaphon K, Nosanchuk JD, Ratanabanangkoon K, Youngchim S. *Talaromyces marneffei* infection: virulence, intracellular lifestyle and host defense mechanisms. J Fungi (Basel). 2022;8(2):200. doi:10.3390/jof8020200.

Qin Y, Huang X, Chen H, Liu X, Li Y, Li Y, Hou J, Li A, Yan X, Chen Y. Burden of *Talaromyces marneffei* infection in people living with HIV/AIDS in Asia during ART era: a systematic review and meta-analysis. BMC Infect Dis. 2020;20(1):551. doi:10.1186/s12879-020-05260-8.

Rex BD, Alexander DA, Arthington-Skaggs B, Brown SD, *et al.* Reference Method for Broth Dilution Antifungal Susceptibility Testing of Yeasts: Approved Standard. 3rd edition. Wayne, Pa: CLSI; 2008.

Rhodes J, Abdolrasouli A, Dunne K, Sewell TR, Zhang Y, Ballard E, Brackin AP, van Rhijn N, Chown H, Tsitsopoulou A, *et al.* Population genomics confirms acquisition of drug-resistant *Aspergillus fumigatus* infection by humans from the environment. Nat Microbiol. 2022;7(5):663–674. doi:10.1038/s41564-022-01091-2.

Rhodes J, Beale MA, Vanhove M, Jarvis JN, Kannambath S, Simpson JA, Ryan A, Meintjes G, Harrison TS, Fisher MC, *et al.* A population genomics approach to assessing the genetic basis of within-host microevolution underlying recurrent cryptococcal meningitis infection. G3 (Bethesda): Genes, Genomes, Genetics. 2017a;7(4):1165–1176. doi:10.1534/g3.116.037499.

Rhodes J, Desjardins CA, Sykes SM, Beale MA, Vanhove M, Sakthikumar S, Chen Y, Gujja S, Saif S, Chowdhary A, *et al.* Tracing genetic exchange and biogeography of *Cryptococcus neoformans* var. *Grubii* at the global population level. Genetics. 2017b;207(1):327–346. doi:10.1534/genetics.117.203836.

Sapmak A, Kaewmalakul J, Nosanchuk JD, Vanittanakom N, Andrianopoulos A, Pruksaphon K, Youngchim S. *Talaromyces marneffei* laccase modifies THP-1 macrophage responses. Virulence. 2016;7(6):702–717. doi:10.1080/21505594.2016.1193275.

Selmecki A, Forche A, Berman J. Aneuploidy and isochromosome formation in drug-resistant *Candida albicans*. Science. 2006;313(5785):367–370. doi:10.1126/science.1128242.

Sephton-Clark P, McConnell SA, Grossman N, Baker RP, Dragotakes Q, Fan Y, Fu MS, Gerbig G, Greengo S, Hardwick JM, *et al.* Similar evolutionary trajectories in an environmental *Cryptococcus neoformans* isolate after human and murine infection. Proc Natl Acad Sci U S A. 2023;120(2):e2217111120. doi:10.1073/pnas.2217111120.

Sephton-Clark P, Tenor JL, Toffaletti DL, Meyers N, Giamberardino C, Molloy SF, Palmucci JR, Chan A, Chikaonda T, Heyderman R, *et al.* Genomic variation across a clinical *Cryptococcus* population linked to disease outcome. mBio. 2022;13(6):e02626-22. doi:10.1128/mbio.02626-22.

Smith KM, Galazka JM, Phatale PA, Connolly LR, Freitag M. Centromeres of filamentous fungi. Chromosome Res. 2012;20(5):635–656. doi:10.1007/s10577-012-9290-3.

Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–1313. doi:10.1093/bioinformatics/btu033.

Stone NRH, Rhodes J, Fisher MC, Mfinanga S, Kivuyo S, Rugemalila J, Segal ES, Needleman L, Molloy SF, Kwon-Chung J, *et al.* Dynamic ploidy changes drive fluconazole resistance in human cryptococcal meningitis. J Clin Invest. 2019;129(3):999–1014. doi:10.1172/JCI124516.

Stott KE, Le T, Nguyen T, Whalley S, Unsworth J, Ly VT, Kolamunnage-Dona R, Hope W. Population pharmacokinetics and pharmacodynamics of itraconazole for disseminated infection caused by *Talaromyces marneffei*. Antimicrob Agents Chemother. 2021;65(11):e00636-21. doi:10.1128/AAC.00636-21.

Supparatpinyo K, Khamwan C, Baosoung V, Nelson KE, Sirisanthana T. Disseminated *Penicillium marneffei* infection in Southeast Asia. Lancet. 1994;344(8915):110–113. doi:10.1016/s0140-6736(94)91287-4.

Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, *et al.* From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43(1):11.10.1–11.10.33. doi:10.1002/0471250953.bi1110s43.

Vanittanakom N, Cooper CR, Fisher MC, Sirisanthana T. *Penicillium marneffei* infection and recent advances in the epidemiology and molecular biology aspects. Clin Microbiol Rev. 2006;19(1):95–110. doi:10.1128/CMR.19.1.95-110.2006.

Verity R, Nichols RA. Estimating the number of subpopulations (K) in structured populations. Genetics. 2016;203(4):1827–1839. doi:10.1534/genetics.115.180992.

Wong SYN, Wong KF. *Penicillium marneffei* infection in AIDS. Patholog Res Int. 2011;10(1):764293. doi:10.4061/2011/764293

Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. Ggtree: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol. 2017;8(1):28–36. doi:10.1111/2041-210X.12628.

Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies. Nat Genet. 2012;44(7):821–824. https://doi.org/10.1038/ng.2310

Editor: A. Mitchell