Title

Understanding Anti-tuberculous Drug Resistance Detection, Prevalence of Resistance-Conferring Mutations, and Population Dynamics of Mycobacterium Tuberculosis complex in a High HIV and Tuberculosis Setting

Author

wang, qiao

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Understanding Anti-tuberculous Drug Resistance Detection, Prevalence of Resistance-Conferring Mutations, and Population Dynamics of *Mycobacterium Tuberculosis* complex in a High HIV and Tuberculosis Setting

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy in Epidemiology

by

Qiao Wang

2024

ABSTRACT OF THE DISSERTATION

Understanding Anti-tuberculous Drug Resistance Detection, Prevalence of Resistance-Conferring Mutations, and Population Dynamics of *Mycobacterium Tuberculosis* complex in a High HIV and Tuberculosis Setting

by

Qiao Wang

Doctor of Philosophy in Epidemiology

University of California, Los Angeles, 2024

Professor Timothy F. Brewer, Chair

Tuberculosis (TB) remains a global health challenge, particularly in regions with high human immunodeficiency virus (HIV) prevalence. HIV weakens the immune system, alters the natural history of TB disease, and is an important risk factor for both drug-sensitive and drug-resistant forms of TB. However, understanding of the transmission dynamics of various TB lineages in high HIV settings is limited. Additionally, the influence of HIV on the diagnosis of drug-resistant TB and the mutations that confer resistance to anti-tuberculous drugs remain unclear. Using whole genome sequencing (WGS) data from *Mycobacterium tuberculosis* complex (Mtbc) isolates collected from a population-based TB study in Botswana, we sought to 1) evaluate the diagnostic accuracy of WGS-based method in detecting resistance to first-line anti-tuberculous

drugs in HIV-positive and HIV-negative individuals; 2) determine whether Mtbc variants with high-fitness cost mutations occurred more frequently among individuals with HIV-infection; and 3) use Bayesian phylodynamic analysis to comprehensively delineate the spread of Mtbc lineages underlying the TB endemic in Botswana. Between 2012 and 2016, 4,331 participants with presumed TB in the greater Gaborone and Ghanzi districts were enrolled into the research study, of whom 2,162 participants had bacteriologically-confirmed TB. Mtbc were successfully cultured and sequenced from 1,426 participants. WGS was performed with the Illumina NextSeq 500 platform and data were analyzed with MTBseq pipeline. Overall, 1,350 participants had available WGS and culture-based phenotypic drug susceptibility testing (pDST; reference standard) results. Of those, 702 (52.0%) were HIV-positive participants, of whom 228 (32.5%) had advanced immunosuppression ($\leqslant$200 CD4+ cells/mm^3). Culture-based pDST (reference standard) indicated 104 (7.7%), 72 (5.3%), 58 (4.3%), and 27 (2.0%) Mtbc isolates had resistance to isoniazid, rifampin, ethambutol, and pyrazinamide, respectively; 49 (3.6%) had resistance to at least isoniazid and rifampin (multidrug-resistant TB). While the overall distribution of anti-tuberculous resistance-conferring mutations among isolates appeared to be similar between HIV-positive and HIV-negative participants, WGS-based detection may have lower sensitivity for identifying isoniazid resistance among participants who had $\leqslant$200 CD4+ cells/mm^3 as compared to participants without HIV (37.5% vs. 71.7%; p-value=0.019). Furthermore, of the Mtbc isolates analyzed, we found the predominant lineage in Botswana was L4 (87.8%), followed by L1 (6.0%), L2 (5.3%), and L3 (0.8%). We identified multiple circulating L4 sublineages, with L4.3.4 being the most prevalent (29.5%). Temporally resolved phylogenies inferred the most recent common ancestor (MRCA) to have emerged around year 1727 (95% highest posterior density [HPD], 1607–1828) and 1900 (95% HPD 1854–1938) for

L1 and L2, respectively. Time to MRCA varied among L4 sublineages; ranging from 1695 for L4.1.2 (95% HPD 1546–1817) to 1941 for L4.3.2 (95% HPD 1911–1966). Phylodynamic analysis revealed L4.3.4 expanded steadily from late 1800s to early 2000s. Conversely, L1, L4.4, and L4.3.2 showed population trajectories closely aligned with the HIV epidemic in the region. Meanwhile, L2 saw rapid expansion throughout most of the 20th century but declined sharply in early 1990s. Additionally, pairwise genome comparison of Mtbc highlighted differences in clustering proportions due to recent transmission at the sublineage level. These findings emphasize the diverse transmission dynamics of Mtbc lineages and highlight the potential for phylodynamic analysis of routine sequences to refine our understanding of lineage-specific behaviors, thereby enabling tailored public health interventions for more effective tuberculosis control in high burden settings.

The dissertation of Qiao Wang is approved.

Sanghyuk Shin

Marc Adam Suchard

Akihiro Nishi

Timothy F. Brewer, Committee Chair

University of California, Los Angeles

2024

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would like to first express my deepest gratitude to my committee members, Drs. Timothy Brewer, Sanghyuk Shin, Marc Suchard, and Akihiro Nishi, for their guidance, support, and expertise throughout this journey. Their insights and encouragement have been instrumental in the completion of this work. A heartfelt appreciation goes to the Kopanyo Study team and participants, whose dedication and participation made this research possible. I would also like to extend my appreciation to Research Center Borstel, who handled all the sequencing for the study, especially Ivan Barilar and Dr. Stefan Niemann, for their guidance on understanding the bioinformatics and sequencing data. Special thanks to Dr. Volodymyr Minin and Isaac Goldstein at UC Irvine for their advice and support, especially during my early stages of learning and using BEAST. I'm also deeply thankful to Ms. Joy Miller and Ms. Shirley Rayner in our department, whose administrative support and assistance have been essential. Lastly, on a personal note, I'm most grateful to my family and friends for their patience, love and encouragement.

This research was supported by the National Institute of Allergy and Infectious Diseases (R01AI097045, R01AI147336).

Chapters 1 – 4 of this dissertation are in preparation for publication as two separate manuscripts.

VITA

EDUCATION

2017 University of California, Los Angeles

Master of Public Health

2011 University of California, San Diego

Bachelor of Science

HONORS & AWARDS

2022 – 2023 Graduate Research Mentorship Fellowship, UCLA Graduate Division

2023 – 2024 Dissertation Year Fellowship, UCLA Graduate Division

PUBLICATIONS

Wang Q, Modongo C, Allender C, Engelthaler DM, Warren RM, Zetola NM, Shin SS. Utility of targeted, amplicon-based deep sequencing to detect resistance to first-line tuberculosis drugs in Botswana. Antimicrob. Agents Chemother. 2019; 63(11). doi:10.1128/AAC.00982-19.

Wang Q, Dima M, Ho-Foster A, Molebatsi K, Modongo C, Zetola N, Shin SS. The association of household food insecurity and HIV infection with common mental disorders among newly diagnosed tuberculosis patients in Botswana. Public Health Nutr. 2020 Oct 19;1-9.

Chapter 1. Introduction

Tuberculosis (TB) is a leading cause of death from a single infectious agent, killing almost 1.5 million lives each year worldwide.[1] Large-scale comparative genomic analysis have provided valuable insights into the origin and genetic diversity of human-adapted TB bacteria, collectively known as the *Mycobacterium tuberculosis* complex (Mtbc).[2, 3] There are currently a total of 9 major lineages of Mtbc and 64 sublineages that characterize their genetic diversity according to a 62-single nucleotide polymorphism (SNP) barcode.[4-6] Strain diversity in Mtbc can lead to important phenotypic variations in virulence, antibiotic resistance patterns, and transmissibility.[7, 8] For example, Lineages 2 and 4 are widespread globally and noted for their enhanced transmissibility, as evidenced by comparing Mtbc genomes (*e.g.* genotypic clustering proportions) collected from patients in many settings.[9, 10] By contrast, Lineage 1 is less likely to cause disease shortly after infection and may have a reduced ability to spread.[7, 9, 10] However, most epidemiological studies and surveillance efforts still do not distinguish between these Mtbc lineages, and the conventional approach of treating TB as a uniform epidemic with a "one size fits all" strategy fails to account for the heterogeneity of the disease, which varies significantly across different regions, populations, and Mtbc lineages.

Located in southern Africa, Botswana continues to face one of the most severe TB endemics globally and remains one of the 30 high TB/HIV burden countries.[11] Despite the critical public health implications, the evolutionary history and transmission dynamics of Mtbc lineages remain poorly understood in this high burden region. Understanding the specific dynamics of Mtbc lineages may help tailor public health interventions to unique challenges presented by different TB strains, ultimately leading to more effective control and elimination strategies.

Furthermore, the emergence and spread of antimicrobial resistance poses a significant challenge to effective TB control.[12, 13] In 2008, the Botswana National Drug Resistance Survey found 9% of diagnosed TB patients had drug-resistance to isoniazid (INH), rifampin (RIF), or ethambutol (EMB) – almost 2-fold increase from the previous survey in 2002.[14] Multidrug-resistant tuberculosis (MDR-TB), characterized by resistance to at least INH and RIF, remains a major concern due to frequent misdiagnosis, low treatment success rate, and increased mortality.[13, 15] Mtbc acquires drug resistance through spontaneous mutations, including single nucleotide polymorphisms (SNPs), insertions, and deletions (indels).[16] Laboratory studies have demonstrated that Mtbc strains with INH- and RIF-resistant mutations may experience decreased competitive fitness (*i.e.,* reduced growth) relative to drug-susceptible strains, and these fitness cost associated with different resistance-conferring mutations may vary.[17-20] Mutations associated high fitness cost are less prevalent in clinical populations. However, HIV-associated immunosuppression may facilitate the proliferation and transmission of drug-resistant strains with high-fitness cost mutations, especially among those with severe immunosuppression (*e.g.,* CD4+ <200 cells/mm^3). This might offer partial insight into the association between HIV infection and increased MDR-TB prevalence,[21, 22] as well as the occurrence of deadly MDR-TB outbreaks in regions with a high prevalence of HIV.[23-25] Nonetheless, investigations of the differential occurrence of resistance mutation patterns in Mtbc strains from people with and without HIV-infection have yielded inconsistent results.[26, 27]

Accurate and timely diagnosis of drug resistance is critical to guide selection of effective treatment regimens and prevent further spread of drug-resistant strains. Recently, the World Health Organization (WHO) endorsed the use of next-generation sequencing (NGS; specifically targeted deep sequencing) method to detect anti-tuberculous drug resistance.[28] Compared to

current commercial molecular assays that target limited resistance-conferring mutations, NGS offers high coverage by characterizing more gene regions simultaneously.[29, 30] However, the reliability of NGS depends on the inclusion of an extensive database of mutations responsible for drug resistance, including those observed less frequently in endemic populations due to potential high fitness costs they might carry. There is some evidence indicating that targeted deep sequencing (tNGS) may exhibit reduced accuracy in detection of drug-resistant TB among people living with HIV compared to those without HIV, possibly due to the presence of uncommon resistance-conferring mutations with high fitness costs not targeted by tNGS.[31]

In this research, we leveraged whole genome sequencing (WGS) data from archived Mtbc isolates collected from a population-based TB study in Botswana and addressed the following study aims:

Aim 1: Evaluate the clinical performance of WGS-based methods in detecting drug resistance to first-line anti-TB drugs (sensitivity and specificity) among HIV-positive and HIV-negative TB patients. We hypothesize that sensitivity of WGS-based algorithm for resistance detection is reduced (vs. reference standard culture-based drug susceptibility testing) among HIV-positive individuals.

Aim 2: Describe genetic mutations associated with first-line anti-TB drugs in HIV-positive and HIV-negative TB patients. We hypothesize that high-fitness cost variants associated with resistance to INH and RIF are overrepresented in HIV coinfected TB patients.

Aim 3: Use Bayesian phylodynamic analysis to comprehensively delineate the spread of Mtbc lineages underlying the TB endemic in Botswana. We hypothesize that TB in Botswana is composed of a collection of heterogeneous epidemics driven by distinct transmission dynamics of Mtbc lineages.

Chapter 2. Methods

2.1 Study setting and data collection procedures

We obtained data from a population-based TB study conducted between 2012 and 2016 in Botswana (R01AI097045; referred to as the Kopanyo study hereinafter).[32] Botswana has one of the highest HIV prevalence rates (20.7% in the general population in 2019) and TB incidence rates (236 per 100,000 in 2021) in the world.[1, 33] In addition, about 50% of people with TB also have HIV/AIDS.[34]

Between 2012 and 2016, the Kopanyo study enrolled persons of all ages diagnosed with pulmonary TB (N = 4,331) in the greater Gaborone and Ghanzi, two largest districts in the country.[35] Gaborone, located in the southeast and bordering South Africa, is the capital city with a high population density of 1,370 persons/km^2 in 2011. In contrast, Ghanzi is a rural district located in western Botswana with the lowest population density in the country according to the 2011 census, at 0.37 persons/km^2.[36] Despite Ghanzi's low population density, it consistently has the highest TB incidence in the country. Prior to the study, the TB incident rate per 100,000 population in Botswana, and Gaborone and Ghanzi districts were approximately 445, 440, and 722 cases, respectively.[37]

At the time of enrollment, demographic and clinical data were collected by standardized interview and medical record abstraction, including age, gender, education, history of TB and TB treatment, HIV status, CD4+ T-cell counts, and antiretroviral therapy (ART) status. In adherence to Botswana national guidelines, the study team extended the offer of HIV testing to all participants who lacked documented HIV test results or had received negative test results more than 12 months prior to enrollment.[32] One to two expectorated sputum samples (2mL – 5mL per

sample) were also collected from each participant. The sputum samples were decontaminated using the N-acetyl-L-cysteine and sodium hydroxide method, and then inoculated into Mycobacteria Grow Indicator Tubes (MGIT) 960 system (Becton Dickinson Microbiology Systems, Sparks, MD) at the Reference Laboratory in Gaborone. Weekly monitoring of mycobacteria growth occurred, and if no growth was observed within 8 weeks, it was classified as culture negative. For culture positive colonies, Mtbc DNA was extracted with GenoLyse kit (Hain Lifescience, Germany) following the manufacturer's protocol. Genomic extracts were stored at -80°C, and those containing a sufficient amount of DNA (at least 0.05ng/uL) were sent to the Research Center Borstel for WGS.

The Kopanyo Study was approved by the U.S. Centers for Disease Control and Prevention Institutional Review Board, the Botswana Ministry of Health and Wellness, and the Institutional Review Board of the University of Pennsylvania. Additionally, approval of the secondary data analysis of the Kopanyo study was obtained from UC Irvine Institutional Review Board. Written informed consent was obtained from all participants.

2.2 Culture-based drug susceptibility testing (DST)

Culture-based or phenotypic drug susceptibility testing (pDST) for first-line anti-TB drugs was performed with MGIT 960 system and/or conventional culture method with Lowenstein-Jensen (LJ) medium. The critical concentrations for INH, RIF, ethambutol (EMB), and pyrazinamide (PZA) using MGIT DST were 0.1 μg/ml, 1.0 μg/ml, 5.0 μg/ml, and 100 μg/ml, respectively.[38, 39] In cases where MGIT DST outcomes were unavailable, conventional culture method with Lowenstein-Jensen (LJ) medium was used. The standard critical concentrations for INH, RIF, and EMB using LJ DST were 0.2 μg/ml, 40 mg/mL, and 2 μg/ml, respectively.[39]

2.3 Whole genome sequencing and bioinformatics

Genomic libraries were prepared with Illumina Nextera XT kit to generate 2x150bp paired-end reads for sequencing using the Illumina NextSeq 500 platform. An automated bioinformatics pipeline, MTBseq,[40] was used for WGS data analysis. MTBseq combined all necessary steps needed for analysis of WGS data. Briefly, raw sequences were aligned to the pan-susceptible, H37Rv reference genome (GenBank NC000962.2) with BWA-mem software and Samtools. Next, base call recalibration and realignment of reads around insertions and deletions was performed with GATKv3. Variant calling was performed using Samtools mpileup and in-house scripts, and positions were considered reliable with 75% allele frequency, at least 4 calls with a phred score of 20 or more, and a minimum of 4 reads mapped in each direction (forward and reverse orientation). Libraries were considered as high quality and further analyzed when at least 50x coverage was achieved and 95% of the reference genome covered. Next, genomes were annotated with known associations to antibiotic resistance. Resistance was called for any known resistance-conferring mutation appearing at an at least 5% allele frequency. MTBseq has shown superior sensitivity and similar specificity in resistance prediction compared to other popular tools such as TBProfiler, PhyResSE, and Mykrobe Predictor TB.[40] Finally, phylogenetically informative SNPs based on existing literature were used for lineage classification of the input samples,[6] excluding genes associated with antibiotic resistance, insertions and deletions, repetitive regions (PPE and PE-PGRS gene families), and consecutive variants in a 12bp window.

Finally, manual curation of called resistances was performed based on the frequency of the resistance-conferring mutation and the percentage of reads mapping to the reference Mtbc

genome. This was done to eliminate false resistance calls introduced by mapping of non-Mtbc reads. In short, mutations with a lower or similar frequency to the percentage of non-Mtbc reads in the sample were excluded.

2.4 Phylogenetic reconstruction

We extracted variable sites from whole genome alignments of Mtbc isolates to create the fasta files used for phylogenetic analysis. To illustrate the overall population structure of Mtbc lineages in Botswana, we used IQ-TREE v1.6.12 to reconstruct maximum likelihood phylogenies for each major Mtbc lineage, with isolates possibly containing mixed Mtbc strains excluded from the dataset.[41] The Hasegawa, Kishino, and Yano (HKY) model of nucleotide evolution was assumed, and ascertainment bias correction was specified. Additionally, time-measured phylogenetic trees for specific lineages included in the coalescent-based analysis were estimated using Bayesian method with details described in Section 2.5.3.

2.5 Statistical analysis

2.5.1 Aim 1

Sensitivity and specificity of WGS-based resistance prediction were calculated using culture-based pDST results as the reference standard. Confidence intervals (95% CI) were estimated using the Score method. Subsequently, we compared the sensitivity and specificity estimates using Fisher's exact test in the following groups: 1) HIV-positive vs. HIV-negative participants; 2) HIV-positive participants with CD4+ T-cells ≤200 cells/mm^3 vs. HIV-negative participants; and 3) HIV-positive participants with CD4+ T-cells >200 cells/mm^3 vs. HIV-negative participants. Moreover, we investigated variations in sequencing quality among the

different groups, specifically the percentage of reads mapped to the reference Mtbc genome, to examine if any differences observed could be attributed to disparities in sequencing quality.

2.5.2 Aim 2

We described the frequencies of detected mutations associated with each first-line anti-tuberculous drug, stratified by HIV infection status. We used bivariate logistic regression analysis to assess the potential relationships between HIV-immunosuppression and the presence of low fitness among mutational variants phenotypically resistant to INH or RIF. We specifically focused on INH and RIF due to the previously demonstrated variations in the reproductive fitness of select mutations commonly associated with resistance to these drugs.[18-20] In particular, *kat*G S315T (associated with INH resistance) and *rpo*B S450L (associated with RIF resistance) were defined as mutations with little or no fitness cost (*i.e.,* high-fitness mutants).[19, 26, 27] Of note, *kat*G S315T and *rpo*B S450L mutations were found to be the most prevalent mutations among persons with INH and RIF resistance, respectively.[42, 43]

The primary dependent variable was the presence of low-fitness Mtbc variant isolates (resistance-conferring mutations other than *kat*G S315T and *rpo*B S450L). The primary independent variables included HIV-positivity and baseline CD4+ T-cell count. Participants were categorized as HIV-negative, HIV-positive with ≤200 CD4+ cells/mm^3, HIV-positive with >200 CD4+ cells/mm^3, and HIV-positive with unknown CD4+ status. Among HIV-positive participants with available CD4+ T-cell data, we also used cell counts as a continuous exposure variable in a separate analysis. Because of restricted sample size of drug-resistant isolates, we did not employ multivariable logistic regression. We repeated all analyses using multinomial logistic regression among all participants who had concordant WGS and pDST results to explore whether

HIV-associated immunosuppression was associated with the risk of INH or RIF low fitness variant infection.

Sensitivity analysis

Sensitivity analyses were conducted to assess the robustness of findings from the main analysis described above. We modified the definition of high or low fitness variants (outcome variable) and repeated the regression analysis. Specifically, isolates were categorized as 'high-fitness variants' if *kat*G S315T and/or *rpo*B S450L were present, otherwise as 'low-fitness variants'. In addition, we repeated the multinomial logistic regression analysis by including isolates with discordant genotypic and phenotypic DST results.

2.5.3 Aim 3

We reconstructed Mtbc population dynamics for L1, L2, and L4 sublineages that had at least 70 samples using BEAST v1.10.4.[44] The HKY nucleotide substitution model with gamma-distributed rate heterogeneity (discretized to 4 categories) was assumed. As previously described, we specified a strict molecular clock model with an initial substitution rate parameter set to 1E-7 per site per year, which translates to approximately 0.44 substitutions per genome per year.[45, 46] The prior for the substitution rate was assumed to be distributed log-normal with a mean of 1 and a sigma of 1.25. To estimate the mycobacterial population size change through time, we adopted the non-parametric, coalescent-based Gaussian Markov random fields (GMRF) Bayesian skyride tree prior.[47] This approach offers greater flexibility compared to the Bayesian skyline prior and does not require strong prior assumptions on the number of population size change points.[47, 48] Furthermore, to account for ascertainment bias, we made manual adjustments in the xml file to incorporate the number of invariant sites. This correction was applied individually for each xml

file corresponding to a specific lineage. The Markov chain Monte Carlo (MCMC) chain for each lineage ran for 500 million iterations, with parameters and trees sampled every 50,000th iteration. We verified adequate mixing and posterior convergence using Tracer v1.7.2 after discarding the first 10% as burn-in. For a parameter to be considered sufficiently sampled, we aimed for an effective sample size of at least 200. If any parameter fell short of this threshold, we ran a second chain of 500 million iterations and combined the chains using LogCombiner v1.10.4. We note that L4.3.4 – the largest sublineage in our dataset – was the only one that required a second MCMC chain of 500 million iterations to achieve sufficient mixing and sampling.

We explored differences in patterns of recent transmission between Mtbc lineages that were included in the coalescent analysis. Genetically similar Mtbc strains under a pre-specified SNP threshold (usually between 5 and 12-SNPs) are interpreted as a consequence of recent transmission events[49]. We assigned cases into clusters, which was defined as two or more individuals with Mtbc strains sharing ≤5-SNPs differences, as determined by pairwise SNP comparison.[49, 50] Next, we regressed clustered or non-clustered case on Mtbc lineage in a logistic model, adjusting for potential confounding by age, gender, enrollment district, and HIV status. To evaluate the robustness of our findings, we also conducted logistic regression analysis using a 12-SNPs cutoff for identifying clustered cases. Furthermore, to assess potential variation of HIV prevalence across Mtbc lineages, we used a multivariable logistic regression model controlling for age, gender, and district.

All statistical analysis and phylogenetic tree visualizations were conducted in R version 4.2.0.[51] WGS data from this study has been submitted to the European Nucleotide Archive (ENA) under study accession PRJEB 62480.

Chapter 3. Results and Discussion for Aim 1 and Aim 2

3.1 Aim 1 Results

Participants' characteristics

The Kopanyo Study enrolled 4,331 participants with presumed TB, of whom 2,162 participants had bacteriologically-confirmed TB. Of the 2,162 participants, 1,426 isolates were successfully sequenced by WGS. Overall, 1,350 participants had available WGS and pDST results. Of those, the median age was 33 years, 44.4% were female, and 19.0% had a previous episode of TB. About half the participants (52%) were HIV-positive, of whom 45.7% were ART-naive, and 32.5% had advanced immunosuppression (≤200 CD4+ cells/mm^3) at the time of enrollment. Participants' characteristics were similar across the groups with WGS data only, pDST data only, both WGS and pDST data, and neither WGS nor pDST data (Table 1).

Of the 1,350 participants, pDST identified 104 (7.7%), 72 (5.3%), 58 (4.3%), and 27 (2.0%) isolates to be resistant to INH, RIF, EMB, and PZA, respectively. MTBseq classified 103 (7.6%), 120 (8.9%), 36 (2.7%), and 105 (7.8%) isolates to be resistant to INH, RIF, EMB, and PZA, respectively. MDR-TB was detected among 49 (3.6%) and 68 (5.0%) participants based on pDST and MTBseq, respectively (Table 1).

Accuracy of WGS-based resistance prediction

We further excluded 17, 43, and 33 isolates for INH, RIF, and PZA, respectively, due to possible incorrect low-frequency mutations calls from non-Mtbc reads. Using pDST as the reference standard, MTBseq identified 70 of 103 INH-resistant isolates (INH sensitivity: 68.0%; 95% CI: 58.4–76.2%) and 1212 of 1228 INH-sensitive isolates (INH specificity: 98.7%; 95% CI:

97.9–99.2%); 65 of 71 RIF-resistant isolates (RIF sensitivity: 91.5%; 95% CI: 82.8–96.1%) and 1222 of 1234 RIF-sensitive isolates (RIF specificity: 99.0%; 95% CI: 98.3–99.4%); 24 of 58 EMB-resistant isolates (EMB sensitivity: 41.4%; 95% CI: 29.6–54.2%) and 1278 of 1290 EMB-sensitive isolates (EMB specificity: 99.1%; 95% CI: 98.4–99.5%); 9 of 26 PZA-resistant isolates (PZA sensitivity: 34.6%; 95% CI: 19.4–53.8%) and 666 of 689 PZA-sensitive isolates (PZA specificity: 96.7%; 95% CI: 95.0–97.8%). The specificity estimates remained consistently high regardless of participants' HIV-infection status and CD4+ cell counts for all drugs (Table 2). Similarly, we did not find a difference in MTBseq sensitivity stratified by HIV status for all drugs.

When stratified by CD4+ T-cell counts, we observed a significant reduction in INH sensitivity among HIV-positive participants who had ≤200 cells/mm^3 (sensitivity: 37.5%; 95% CI: 18.5–61.4%) compared to HIV-negative participants (sensitivity: 71.7%; 95% CI: 57.5–82.7%; *p*-value = 0.019). INH sensitivity estimates were similar between HIV-positive participants who had >200 CD4+ cells/mm^3 and HIV-negative participants. MTBseq sensitivity for RIF, EMB, and PZA remained statistically consistent across CD4+ T-cell groups (Table 2). We did not observe variations in sequencing quality among the comparison groups, as indicated by the percentage of reads mapped to the reference Mtbc genome (Figure 1). Additionally, no difference in the percentage of reads mapped to Mtbc was observed between isolates that belong to the false negative group (*i.e.,* phenotypically resistant and genotypically susceptible) and true positive group (phenotypically resistant and genotypically resistant) for INH (Figure 2).

3.2 Aim 2 Results

Distribution of resistance-conferring mutations

The frequencies for all resistance-associated mutations detected by MTBseq were described by HIV status in both phenotypically resistant and susceptible isolates (Table 3). Among isolates with phenotypic INH resistance, approximately 75% and 17% harbored the *kat*G S315T mutation and *inh*A promoter (*fab*G1) -15c>t mutation, respectively. One isolate was identified to carry both mutations. Conversely, *fab*G1 -8t>c or -8t>a mutations only co-occurred with other mutations linked to INH resistance; *fab*G1 -8t>a exclusively co-occurred with *kat*G S315T at 11.3%, while *fab*G1 -8t>c co-occurred with *inh*A I21V mutation at 2.8%. The overall distribution of *kat*G S315T and *fab*G1 -15c>t mutations appeared to be similar between HIV-positive and HIV-negative participants. However, the *fab*G1 -8t>a and *kat*G S315T combination appeared more frequently in persons with HIV (16% vs. 6% in HIV-negative group). Additionally, approximately 6% of INH-resistant isolates carried mutations outside of the target regions by conventional molecular diagnostics, including *fab*G1 L203L, *inh*A I21V, and/or *inh*A S94A (Table 3).

For isolates displaying phenotypic resistance to RIF, 43% were identified to harbor *rpo*B S450L mutation, and none of these isolates carried additional mutations linked to RIF-resistance. The second most prevalent mutation associated with RIF resistance in our sample was *rpo*B H445L at approximately 17%. While the occurrence of the *rpo*B S450L mutation was similar between participants with and without HIV infection, the *rpo*B H445L mutation appeared more frequently in persons with HIV as compared to HIV-negative person (22% vs. 11%). Approximately 6% of RIF-resistant isolates carried mutations outside of the target regions by

conventional molecular diagnostics, including *rpo*B V170F, *rpo*B 838_del_c, *rpo*B 1292_ins_cca, and *rpo*B 1348_ins_cgg (Table 3).

Association between HIV infection and presence of low-fitness Mtbc variants

Of 69 phenotypically INH-resistant isolates linked to HIV status, 42 (61%) exclusively harbored the *kat*G S315T mutation (defined as high-fitness INH-resistant variant). Similarly, among 65 phenotypically RIF-resistant isolates linked to HIV status, 29 (45%) displayed *rpo*B S450L mutation alone (defined as high-fitness RIF-resistant variant).

No statistical associations were observed between HIV-associated immunosuppression and the presence of low-fitness INH- or RIF-resistant isolates (Table 4). Similarly, among the entire cohort and irrespective of pDST resistance status, we did not find an association between HIV-associated immunosuppression and presence of low-fitness variants. However, among HIV-positive participants, there was an increased risk of having INH- and RIF-resistant isolate, both with high and low fitness (vs. INH- and RIF-sensitive strains, respectively) as the number of CD4+ cells increased (Table 5).

Sensitivity analysis

Despite modifications in the definition of high or low fitness variants in the regression analysis, the overall findings from the primary analyses remained unchanged (Tables 6 and 7). Furthermore, incorporating isolates with discordant genotypic and phenotypic DST results in the multinomial logistic regression analysis also yielded consistent results (Table 8).

3.3 Discussion

This study aimed to understand the population-level impact of HIV-associated immunosuppression and the molecular detection of drug resistance-conferring mutations. Our overall findings indicate that, in settings where HIV is prevalent, molecular detection may exhibit lower sensitivity in identifying INH-resistant TB. However, our study does not support an association between HIV-associated immunosuppression and the presence of less fit INH- or RIF-resistant isolates.

In our prior study of the utility of next-generation deep sequencing for detecting drug-resistant TB, we found lower sensitivity for detecting INH resistance among people with HIV compared to people without HIV.[31] That earlier study had a smaller sample size and did not account for CD4+ T-cells. In the present analysis, a marked reduction in INH sensitivity was observed among those with ≤200 CD4+ cells/mm^3. Further exploration of sequencing quality did not reveal a lower percentage of mapped reads against the reference Mtbc genome among the false negative group. This suggests that unaccounted novel mutations associated with INH resistance in the current bioinformatics pipeline could potentially drive phenotypic resistance among people with profound immunodeficiency. Indeed, the molecular mechanisms behind INH resistance remain unknown for a notable proportion of clinical isolates.[42, 52] Thus, it is possible that diminished immune pressure in the context of advanced immunodeficiency allows Mtbc strains with high-fitness cost mutations to survive, a scenario less likely in individuals with intact immune systems.

Consistent with existing studies, we found high concordance between WGS and pDST among the phenotypically susceptible isolates across all evaluated drugs.[53-57] However, distinct from prior research, our analysis revealed lower sensitivity of WGS for detecting INH resistance

(68%), EMB resistance (41%), and PZA resistance (37%). Possible explanations for these discrepancies include the existence of novel mutations associated with phenotypic resistance that have yet to be discovered and integrated into the existing MTBseq pipeline,[58, 59] the presence of undetected minority resistant variant,[60] and the susceptibility of culture-based DST to errors, particularly in cases involving EMB and PZA – drugs known to pose challenges in accurate phenotyping.[39, 61, 62] While WGS-based approaches have been shown to accurately detect first-line anti-tuberculous-drug resistance in other settings,[53] our findings suggest that the performance of WGS-based algorithms for identifying first-line anti-tuberculous-drug resistance (*e.g*. INH, EMB, and PZA) may depend on geography and epidemiological context, including the prevalence of HIV-associated immunosuppression.

Approximately 95% of RIF resistance could be attributed to mutations in the *rpo*B gene that encodes the binding site of RIF.[63] In addition, the majority of mutations are found within the 81-bp RIF-resistance determining region (RRDR; *rpo*B codons 426–452) of the rpoB gene,[63, 64] which is the target region of conventional molecular tests, including GeneXpert MTB/RIF® (Cephid, Sunnyvale, CA, USA) or other commercially available line-probe assays.[65, 66] However, the prevalence of *rpo*B mutations driving RIF resistance varies geographically, and variations in phenotypic resistance levels exist across different *rpo*B mutations.[67] For instance, *rpo*B S450L (associated with high-level RIF resistance) varies in frequency (38–90%) across geographical regions.[68-70] In our study in Botswana, 43% of isolates displaying phenotypic RIF resistance had the *rpo*B S450L mutation. The second most prevalent mutation in our population was *rpo*B H445L (approximately 17%). This mutation has been linked to low-level RIF-resistance,[67, 71] emphasizing the value of genotypic methods in identifying RIF resistance that could be missed by culture-based pDST (MGIT) using standard critical concentration for RIF.

In contrast to RIF resistance, the molecular mechanism underlying INH resistance are less well-understood and encompass a more extensive set of loci.[42] Collectively, *kat*G S315T and *fab*G1 -15c>t mutations account for approximately 80% of INH resistance globally.[42, 72] The frequencies of these mutations observed in Botswana (75% *kat*G S315T and 17% *fab*G1 -15c>t) are consistent with both quantified global estimates and those specific to the African region.[42] The single isolate we identified that carried both mutations supports earlier reports of their infrequency and suggests a potential disadvantage for Mtbc strains with mutations in both genes within a host.[20, 52, 73] A notable proportion (23%) of INH-resistant isolates in our study had independent *fab*G1 mutation or *fab*G1 mutation without concurrent *kat*G 315 mutation. While *kat*G S315T is known to confer high-level INH resistance, mutations in the *inh*A promoter region (*fab*G1) have been linked to low-level drug resistance to INH.[74] Consequently, the incorporation of high-dose INH in curative therapy might be beneficial for this subgroup of patients.[75, 76]

A small proportion of INH- or RIF-resistant isolates did not carry mutations in the target regions of conventional molecular diagnostics. These include INH-resistance-associated *fab*G1 L203L, *inh*A S94A, *inh*A I21V; and RIF-resistance-associated *rpo*B V170F, and multiple *rpo*B indels between positions 838 to 1348, which could be valuable additions to next-generation molecular diagnostics. The sensitivity of any molecular diagnostic to rapidly identify drug resistance is connected closely to the regional frequencies of mutations and the comprehensiveness of the resistance-conferring mutations targeted by the test. By understanding the specific landscape of mutations in Botswana, our study contributes to the potential for developing enhanced next-generation diagnostics and therapy recommendations specifically tailored to the context of Botswana.

Strengths of this study include providing a comprehensive description of mutations linked to first-line anti-tuberculous drugs, stratified by HIV status, using a population-representative cohort. Among the phenotypically resistant isolates, our findings did not support the hypothesis that low-fitness INH- or RIF-resistant variants were overrepresented in people with HIV or among those with severe immunosuppression as compared to persons without HIV. Our finding is consistent with another study conducted in Uganda,[27] a country with similar MDR-TB and HIV burden. However, Louiseau *et al.* found that HIV-associated immunosuppressed persons were more likely to harbor low-fitness RIF-resistant Mtbc variants (*i.e.* non-*rpo*B S450L variants) using a global sample of drug-resistant TB isolates from nine countries, including 39 isolates from the Kopanyo Study.[26] Geographic heterogeneity, local transmission dynamics, HIV prevalence, and MDR-TB prevalence might contribute to Mtbc genetic diversity, which may influence the proliferation of certain resistant mutations over others.[18, 20, 26, 77, 78] Strain genetics may also play an important role in the adaptation to specific antibiotics and driving distinct resistance-conferring mutations.[77, 78] Our findings remained when we extended the model to include the entire sample irrespective of phenotypic resistance status. While our study also revealed a higher prevalence of INH- and RIF-resistant Mtbc variants among HIV-infected TB patients with increasing CD4+ cell counts, the biological mechanisms underlying this observation remain unclear. Further investigations are warranted to elucidate the underlying mechanisms driving this phenomenon.

3.4 Limitations

Our study had several limitations. First, pDST is an imperfect reference standard that is prone to a variety of errors. Second, the Mtbc DNA extraction method used in the Kopanyo study

was optimized for tNGS (not WGS), and sequencing quality may be affected as a result. However, we mitigated this by excluding low-quality isolates during the initial bioinformatics stage, and the percentage of reads mapped to the Mtbc genome was comparable between the analyzed comparison groups. Third, the low prevalence of MDR-TB in Botswana limited the number of drug-resistant isolates in our study. This may have impeded our ability to measure associations with sufficient statistical power. Consequently, a study with a larger sample size that includes *rpo*B H445L and the combination of *fab*G1 -8t>a and *kat*G S315T mutations would likely improve statistical fidelity needed to accurately measure these associations. Fourth, the associations and diagnostic accuracy observed in our study may not be generalizable to other settings. Furthermore, our study was not designed to assess the clinical utility of MTBseq as a diagnostic tool. MTBseq is continually updated with emerging mutations in its resistance-associated mutation catalogue. Further studies should be conducted to evaluate the clinical utility of MTBseq to detect anti-TB drug resistance, and clinical outcome data may be linked to drug-resistant phenotypes to infer the potential clinical benefit of using WGS to identify drug resistance. Lastly, our study was observational in nature, and the observed associations (or lack thereof) between HIV-associated immunosuppression and low-fitness Mtbc variants must be interpreted with caution due to the potential presence of uncontrolled confounding.

3.5 Conclusion

Despite these limitations, our study highlights the possibility of the role of novel mutations in driving INH-resistance among individuals with HIV-associated immunosuppression. These novel mutations should be considered for the molecular detection of drug resistant TB, especially among populations where HIV is common. Ongoing molecular

surveillance for detecting emerging drug resistance is recommended and may inform local guidance in HIV endemic settings.

Chapter 4. Results and Discussion for Aim 3

4.1 Results

Distribution of Mtbc lineages in Botswana

During the study period, 1,426 Mtbc isolates underwent successful WGS and were assigned into known lineages. We excluded 72 isolates with evidence of mixed Mtbc strain infection for subsequent phylogenetic and phylodynamic analysis. Among the remaining 1,354 isolates, 1,184 belonged to lineage 4 (L4; 87.4%), 86 (6.4%) were classified as lineage 1 (L1), 72 (5.3%) were lineage 2 (L2), and 12 (0.9%) were lineage 3 (L3). L4 was found to encompass a number of sublineages, including L4.1 (0.1%), L4.1.1 (10.4%), L4.1.2 (11.0%), L4.2.2 (0.1%), L4.3.2 (8.6%), L4.3.3 (2.0%), L4.3.4 (29.5%), L4.4 (14.0%), L4.6 (0.1%), L4.7 (1.0%), L4.8 (6.8%), and L4.9 (3.7%). Within L1, 22 (1.6%) were L1.1 and 64 (4.7%) were L1.2. All of the L2 isolates were classified as L2.2 (Figure 3). Furthermore, Gaborone harbored more Mtbc strain diversity compared to Ghanzi. L4.3.4, the most prevalent sublineage, was found in higher proportion in Ghanzi than Gaborone (67% vs. 21%; Table 9). Host characteristics appear to be similar among the major Mtbc lineages (Table 10).

Historical origins and population dynamics of Mtbc lineages

Coalescent analysis inferred the most recent common ancestor (MRCA) to have emerged around year 1727 (95% highest posterior density [HPD], 1607–1828) and 1900 (95% HPD 1854–1938) for L1 and L2, respectively. Time to MRCA varied among L4 sublineages, ranging

from 1695 for L4.1.2 (95% HPD 1546–1817) to 1941 for L4.3.2 (95% HPD 1911–1966; Table 11). Our analysis revealed that L4.3.4 experienced a steady expansion from late 1800s to early 2000s, and remained the most prevalent sublineage in recent years. Similar population trajectory was observed for L1 and L4.4, with a substantial expansion observed from the late 1970s for two decades, followed by a continuous decline from the late 1990s. For L4.3.2, we observed significant growth between 1980 to early 2000s, followed by a sharp reduction afterward. L2 experienced a steady increase in population growth between 1940 and early 1990s, followed by a steep decline until 2000, after which the population plateaued followed by a second decline around 2012. Other L4 sublineages experienced several waves of expansion and contraction. Except for L4.1.1, all other Mtbc lineages underwent contraction after early 2000s (Figures 4 and 5).

Genomic clustering analysis

Using a 5-SNPs cutoff as an indicator of recent transmission, 48% (652/1354) were clustered isolates. The cluster sizes varied, ranging from 2 isolates per cluster to 23 isolates in the largest cluster. The median cluster size and range for each lineage was as follows: L1 and L2: 2 (2-7), L4.1.1: 3 (2-23), L4.1.2: 3 (2-12), L4.3.2 and L4.4: 2 (2-8), L4.3.4: 2 (2-19), and L4.8: 3.5 (2-9). Compared to L1, we found that isolates belonging to any other lineages were more likely to be part of a cluster (Table 12). Isolates belonging to L2, L4.1.1, and L4.8 had the highest cluster proportion. Additionally, the cluster proportion for the predominant L4.3.4 was significantly higher in Ghanzi than Gaborone (57% vs. 40%; p-value < 0.001; data not shown) These findings remained robust when we repeated the analysis using a 12-SNPs cutoff (Table

13). In addition, L4.1.1 and L4.8 were found to be significantly less common among HIV-infected individuals (Table 14).

4.2 Discussion

By reconstructing the evolutionary history and spread of Mtbc using sequences collected over a four-year period, we observed varying transmission dynamics and demographic trajectories for extant lineages circulating in Botswana. The expansion of many lineages coincided with the onset of the HIV/AIDS epidemic, while their contraction aligned with the implementation of antituberculosis and antiretroviral treatment programs.

With the exception of L4.1.1, L4.1.2 and L4.8, most lineages experienced an expansion of various magnitude in the second half of the 20th century, correlating closely with the region's high TB burden.[79] Notably, this period coincides with the emergence and rapid escalation of the HIV epidemic. Botswana experienced one of the world's most severe HIV epidemics, with initial cases emerging in the 1980s.[80] A phylodynamic study of HIV dated the virus's emergence in Botswana around 1960, followed by rapid growth until stabilization post-1990.[81] HIV increases the susceptibility of TB disease given infection and is the strongest risk factor for reactivating latent TB infections due to compromised immunity.[82] Further, the HIV crisis was exacerbated by rapid urbanization driven by economic growth in the mining industry. A network analysis of micro-census data from Botswana between 1981 and 2011 suggests a highly mobile population with considerable movement between urban and rural areas.[83] Mining towns, where HIV cases initially emerged, acted as major migration hubs for both inflow and outflow of populations. In Selebi Phikwe, a mining town for copper and nickel in Botswana, HIV prevalence had surged to 50% in 2000.[84] These towns also became hubs of TB transmission due to the high-risk conditions

associated with mining and the growing HIV prevalence.[85, 86] As the HIV epidemic escalated, the number of individuals susceptible to Mtbc infection and reactivation increased, driving the expansion of Mtbc populations. Thus, the intersection of the HIV epidemic, mining-driven urbanization, and high population mobility likely played a critical role in the significant expansion of Mtbc lineages in the latter part of the 20th century.

We observed a marked decline in Mtbc populations at the turn of the 21st century, which aligned with a consistent decrease in TB incidence from 2000 to 2016.[87] Several nationwide programs likely contributed to the reduction. The most significant decrease in L1, L4.3.2, L4.3.4, L4.4, and L4.8 occurred beginning the early 2000s, coinciding with the launch of the Masa Program in 2002.[80] This program, officially known as the Masa Antiretroviral Therapy (ART) Programme, initially aimed to provide free ART to individuals with advanced immune suppression but eligibility expanded over the years. "Masa" means "new dawn" in Setswana, symbolizing hope and a fresh start for those affected by the epidemic. Shortly after, the President's Emergency Plan for AIDS Relief (PEPFAR) was launched, focusing on combating HIV/AIDS globally, with a significant emphasis on sub-Saharan Africa.[88] Botswana, with its high HIV prevalence rates, was a key focus country for PEPFAR. Both initiatives aimed to support those affected by HIV/AIDS through comprehensive treatment, prevention, and care programs, thereby reducing HIV transmission.[88, 89] As a result, Botswana has seen significantly increased number of people receiving ART, reduced the rate of new HIV infections, and improved the overall healthcare system’s capacity to manage HIV/AIDS.[90, 91] While PEPFAR and Masa initiatives primarily targeted the HIV/AIDS epidemic, their impact on reducing HIV transmission indirectly led to a significant reduction in the burden of TB in Botswana.

Our study highlighted significant heterogeneity in transmission dynamics among the Mtbc lineages analysed. We found that L1 maintained a relatively stable effective population size until experiencing a significant expansion following the emergence of HIV. L1 is most prevalent in regions bordering the Indian Ocean, and several large-scale phylogeographic studies have shown that South Asia and eastern Africa played critical roles in dispersing this lineage via maritime trade.[92, 93] It is possible that MRCA of L1 strains was introduced into Botswana through subsequent migrations between eastern and southern regions of the continent. Post establishment, evidence suggests that L1 likely expanded due to changing environmental conditions (e.g. increased population densities, malnutrition, immunosuppression), rather than from large migration or transmission events.[92] Consistent with previous findings, our study found that L1 had the lowest proportion of cases that belonged to clusters and was more common among HIV-infected individuals compared to other lineages.[7, 9, 94] A phylodynamic study from Vietnam, an L1-endemic region, concluded that compared to L2 and L4, the disease burden due to L1 in the region was associated with reactivation of long-term remote infections, evidenced by larger SNP differences and limited inferred migration events among L1 isolates.[95] Similarly, in Botswana, reactivation of latent TB infections likely contributes to new TB cases and may have driven L1's expansion as population immunity waned due to HIV. L4.4 followed similar demographic trends but showed a slower decline in population size after nationwide introduction of ART interventions, which may be linked to its higher rate of recent transmission compared to L1.

Notably, L2 underwent rapid expansion in effective population size shortly after their estimated emergence in the early 20th century. Phylogenetic studies suggest that L2.2, often associated with the Beijing genotype, likely originated in China.[96, 97] The 19th and 20th centuries were characterized by intense interaction between Europe and China, driven by trade, imperial

ambitions, and cultural exchanges. Following the Opium Wars, numerous Chinese cities became treaty ports where European powers had extraterritorial rights and conducted trade with relative freedom. Prominent treaty ports included Shanghai, Guangzhou, and Tianjin, which became hubs of international trade and cultural exchange. Additionally, during the late 19th and early 20th centuries, Chinese laborers were often recruited for mining and railway construction projects worldwide, including in Africa.[98] South Africa, which borders Botswana, had a significant number of Chinese laborers working in its mines. This movement could have facilitated the introduction and spread of L2.2 to southern Africa. L2 has been associated with hypervirulence (ability to cause disease shortly after infection) and high transmissibility, supported by both clinical evidence and animal experiments.[99, 100] In our study, we also observed that L2 had one of the highest proportion of genomic clustering due to recent transmission. The enhanced ability to transmit and cause disease coupled with population mobility likely explain L2's expansion in Botswana.[7, 9, 83] However, L2 experienced a decline in the early 1990s, earlier than many other lineages, including L1. This decline might reflect a combination of factors, including implementation of facility-based directly observed therapy, local pathogen-host dynamics and its challenges in competing with other well-adapted indigenous lineages.[8] Our finding contrasts sharply with the genomic study in Vietnam, where L2 showed clear patterns of overtaking endemic L1 lineages over time.[95] This implies that the competitive dynamics between different lineages can vary significantly depending on the region and the specific lineages involved, and that studying lineage features without accounting for geographic or host genetic background can obscure important host-pathogen interactions.[8, 101]

Our analysis indicates that most TB cases in Botswana were caused by L4 strains, specifically L4.3 lineages. This finding aligns with previous reports of L4-belonging Latin

American and Mediterranean sublineage's dominant presence in southern Africa region.[102-105] Based on the sampled genetic diversity, we estimated that most L4 sublineages trace back to MRCAs that emerged between the late 17th and 19th centuries. Recent genomic studies of Mtbc suggests that the introduction and spread of L4 (Euro-American lineage) coincide with historical events involving European colonization and subsequent socioeconomic changes.[102, 106] European colonization of South Africa, Botswana's neighboring country, began in earnest with the establishment of a Dutch settlement at the Cape of Good Hope in 1652.[107] Later, British colonization intensified in early 19th century, and Botswana (then Bechuanaland) became a British protectorate in 1885. The establishment of colonial administration and subsequent movements of European settlers between Europe, South Africa and Botswana during this period may have facilitated the establishment and spread of L4 Mtbc strains still observed today.

The most common Mtbc lineage found in Botswana was L4.3.4, and it was overwhelmingly more prevalent in Ghanzi than Gaborone (67% vs. 21%). Another study using a different genotyping technique noted this, but was unable to differentiate whether it was due to recent transmission or reactivation of latent infections.[108] The effective population size of L4.3.4 grew steadily since its introduction around 1880 and only began to decline slightly after country-wide ART programs began. Unlike other lineages, L4.3.4's expansion during the HIV epidemic was modest. This contrast may be explained by the higher prevalence of this lineage in Ghanzi, which has lower HIV prevalence compared to Gaborone. Prior to the study, approximately 36% of TB patients in Ghanzi were coinfected with HIV, compared to 70% in Gaborone,[109] suggesting a less pronounced impact of the HIV epidemic on TB transmission dynamics in Ghanzi. In our analysis, L4.3.4 displayed moderate clustering due to recent transmission, falling between the lowest (L1) and the highest groups (e.g., L2, L4.1.1, L4.8). However, it exhibited significantly

higher cluster proportions in Ghanzi compared to Gaborone. Taken collectively, we propose that the success of L4.3.4 is driven primarily by sustained ongoing transmission, rather than reactivation of latent infections.[108] This highlights the need for public health efforts to prioritize activities that interrupt transmission, such as enhancing contact tracing, improving infection control measures, and ensuring timely diagnosis and treatment to reduce disease burden, especially in a rural setting like Ghanzi.

L4.3.4 displayed clear competitive advantage as evidenced by its dominant presence in both districts despite not being the earliest lineage to emerge in Botswana. It's also interesting to observe that L4.3.2, a strain closely related to L4.3.4, emerged around the time of estimated HIV introduction in Botswana.[81] However, unlike L4.3.4's steady growth, L4.3.2 experienced a rapid expansion followed by a sharp decline, mirroring remarkably well with the timing of the HIV epidemic. The striking difference in population dynamics between two closely related strains warrant further investigation to determine whether strain-specific mutations contribute to distinct phenotypes, and/or if potential host-pathogen interactions influence their respective trajectories and the success of L4.3.4.

L4.1.1, L4.1.2, and L4.8 share similar demographic trends with large overlapping CIs. Notably, L4.1.1 and L4.8 exhibited the highest cluster proportions due to recent transmission, almost doubling the estimate compared to L4.3.4. This variation in genotypic clustering within L4 sublineages was also observed in San Francisco, USA,[110] underscoring sublineage-level heterogeneity and the necessity of investigating them individually. Furthermore, L4.1.1 was the only lineage that expanded after the implementation of country-wide ART programs. In addition to possible sublineage-specific characteristics (e.g. virulence), other socioecological drivers, for example, overcrowding or malnutrition, may be responsible for the expansion and contraction of

this sublineage. However, the relatively wide CIs associated with the effective population size of these sublineages limit our ability to analyze local features of their curves.[111]

Our study had limitations. First, the available genetic data may not fully represent the diversity of Mtbc circulating in Botswana. However, the greater Gaborone and Ghanzi district encompass a considerable segment of the total population, and sampling in Gaborone likely captured a substantial portion of Mtbc diversity due to its role as a central urban hub for inflow and outflow of population. Second, coalescent models (e.g. skyride model applied in our analysis) use pathogen genomic data to estimate the timescale of population dynamics.[47] The relatively low mutation rate of Mtbc can result in broad uncertainty intervals for estimated effective population sizes and divergence times, making it challenging to pinpoint precise timings for events such as introductions or expansions of lineages. Inclusion of Mtbc sequences as they become available in the future may improve the reconstruction of Mtbc population dynamics and reduce uncertainty associated with population parameters. Lastly, we acknowledge the limitation of using a fixed SNP threshold to determine transmission clusters.[112] The chosen threshold is often arbitrary and might not accurately reflect recent transmission. However, a SNP cutoff provides a standardized approach, making it easier to compare results across different studies and useful for public health investigations. The chosen 5-SNP cutoff for determining TB transmission clusters is common in molecular epidemiology and has been shown to provide an informative overview of local transmission patterns.[49, 50]

In summary, our study provides insight into the recent evolutionary origins and population dynamics of Mtbc lineages in Botswana. We highlight the heterogeneous transmission dynamics of key lineages circulating in the region and emphasize that increasing awareness of these heterogeneities may be essential to further reducing TB burden. Additionally,

we recommend routine molecular surveillance and phylodynamic analysis to effectively monitor and control local TB transmission in high burden settings.

4.4 Conclusion

In summary, our study provides insight into the recent evolutionary origins and population dynamics of Mtbc lineages in Botswana. We highlight the heterogeneous transmission dynamics of key lineages circulating in the region and emphasize that increasing awareness of these heterogeneities may be essential to further reducing TB burden. Additionally, we recommend routine molecular surveillance and phylodynamic analysis to effectively monitor and control local TB transmission in high burden settings.

Tables

Table 1. Participants characteristics in the Kopanyo Study, 2012 – 2016

Characteristics	pDST available[1]		pDST not available[2]		Overall n = 2162
	Sequenced n = 1350	Not sequenced n = 608	Sequenced n = 76	Not sequenced n = 128	
Age in years, median (IQR)	33.0 (26.0, 41.0)	33.0 (26.0, 40.0)	36.0 (28.8, 42.3)	34.5 (27.0, 46.0)	33.0 (26.0, 41.0)
Sex					
Female	600 (44.4%)	269 (44.2%)	26 (34.2%)	68 (53.1%)	963 (44.5%)
Male	750 (55.6%)	339 (55.8%)	50 (65.8%)	60 (46.9%)	1199 (55.5%)
Income					
No income	529 (39.2%)	231 (38.0%)	33 (43.4%)	57 (44.5%)	850 (39.3%)
<3000 BWP	627 (46.4%)	293 (48.2%)	34 (44.7%)	55 (43.0%)	1009 (46.7%)
3000 – 4999 BWP	89 (6.6%)	40 (6.6%)	6 (7.9%)	6 (4.7%)	141 (6.5%)
5000 – 10,000 BWP	71 (5.3%)	32 (5.3%)	1 (1.3%)	6 (4.7%)	110 (5.1%)
>10,000 BWP	25 (1.9%)	10 (1.6%)	2 (2.6%)	4 (3.1%)	41 (1.9%)
Unknown	9 (0.7%)	2 (0.3%)	0	0	11 (0.5%)
Education					
None	175 (13.0%)	88 (14.5%)	13 (17.1%)	27 (21.1%)	303 (14.0%)
Primary or less	268 (19.9%)	127 (20.9%)	19 (25.0%)	30 (23.4%)	444 (20.5%)
Some secondary	428 (31.7%)	164 (27.0%)	27 (35.5%)	38 (29.7%)	657 (30.4%)
Completed secondary	244 (18.1%)	125 (20.6%)	8 (10.5%)	16 (12.5%)	393 (18.2%)
Some tertiary or Certificate	117 (8.7%)	58 (9.5%)	3 (3.9%)	6 (4.7%)	184 (8.5%)
Diploma or Degree	114 (8.4%)	44 (7.2%)	6 (7.9%)	11 (8.6%)	175 (8.1%)
Masters	3 (0.2%)	2 (0.3%)	0	0	5 (0.2%)
Unknown	1 (0.1%)	0	0	0	1 (0.0%)
Current smoking					
Yes	347 (25.7%)	155 (25.5%)	23 (30.3%)	39 (30.5%)	564 (26.1%)
No	1003 (74.3%)	453 (74.5%)	53 (69.7%)	89 (69.5%)	1598 (73.9%)
Current drinking					
Yes	425 (31.5%)	218 (35.9%)	30 (39.5%)	48 (37.5%)	721 (33.3%)
No	921 (68.2%)	390 (64.1%)	46 (60.5%)	80 (62.5%)	1437 (66.5%)
Unknown	4 (0.3%)	0	0	0	4 (0.2%)
History of incarceration					
Yes	74 (5.5%)	36 (5.9%)	5 (6.6%)	11 (8.6%)	126 (5.8%)

No	1274 (94.4%)	572 (94.1%)	71 (93.4%)	117 (91.4%)	2034 (94.1%)
Unknown	2 (0.1%)	0	0	0	2 (0.1%)
History of TB					
Yes	257 (19.0%)	88 (14.5%)	17 (22.4%)	29 (22.7%)	391 (18.1%)
No	1093 (81.0%)	520 (85.5%)	59 (77.6%)	99 (77.3%)	1771 (81.9%)
History of IPT					
Yes	47 (3.5%)	19 (3.1%)	4 (5.3%)	2 (1.6%)	72 (3.3%)
No	1284 (95.1%)	583 (95.9%)	70 (92.1%)	123 (96.1%)	2060 (95.3%)
Unknown	19 (1.4%)	6 (1.0%)	2 (2.6%)	3 (2.3%)	30 (1.4%)
Treatment group					
New	1096 (81.2%)	525 (86.3%)	59 (77.6%)	100 (78.1%)	1780 (82.3%)
Relapse	229 (17.0%)	74 (12.2%)	16 (21.1%)	26 (20.3%)	345 (16.0%)
Default	10 (0.7%)	4 (0.7%)	0	0	14 (0.6%)
Failure	8 (0.6%)	2 (0.3%)	1 (1.3%)	1 (0.8%)	12 (0.6%)
Other	4 (0.3%)	3 (0.5%)	0	1 (0.8%)	8 (0.4%)
Unknown	3 (0.2%)	0	0	0	3 (0.1%)
AFB smear quantification					
+	143 (10.6%)	55 (9.0%)	13 (17.1%)	10 (7.8%)	221 (10.2%)
++	169 (12.5%)	96 (15.8%)	15 (19.7%)	20 (15.6%)	300 (13.9%)
+++	484 (35.9%)	196 (32.2%)	23 (30.3%)	38 (29.7%)	741 (34.3%)
Scanty	71 (5.3%)	24 (3.9%)	2 (2.6%)	3 (2.3%)	100 (4.6%)
Smear negative	87 (6.4%)	43 (7.1%)	6 (7.9%)	6 (4.7%)	142 (6.6%)
Smear not done	395 (29.3%)	192 (31.6%)	16 (21.1%)	50 (39.1%)	653 (30.2%)
Unknown	1 (0.1%)	2 (0.3%)	1 (1.3%)	1 (0.8%)	5 (0.2%)
TB classification					
Pulmonary	1230 (91.1%)	560 (92.1%)	66 (86.8%)	116 (90.6%)	1972 (91.2%)
Extrapulmonary	113 (8.4%)	47 (7.7%)	9 (11.8%)	12 (9.4%)	181 (8.4%)
Unknown	7 (0.5%)	1 (0.2%)	1 (1.3%)	0	9 (0.4%)
TB symptom length					
<1 week	99 (7.3%)	35 (5.8%)	3 (3.9%)	13 (10.2%)	150 (6.9%)
1 – 2 weeks	239 (17.7%)	99 (16.3%)	17 (22.4%)	32 (25.0%)	387 (17.9%)
3 – 4 weeks	285 (21.1%)	133 (21.9%)	19 (25.0%)	25 (19.5%)	462 (21.4%)
>4 weeks	674 (49.9%)	314 (51.6%)	36 (47.4%)	56 (43.8%)	1080 (50.0%)
No symptom	49 (3.6%)	26 (4.3%)	1 (1.3%)	2 (1.6%)	78 (3.6%)
Unknown	4 (0.3%)	1 (0.2%)	0	0	5 (0.2%)
HIV co-infection					
Yes	702 (52.0%)	322 (53.0%)	47 (61.8%)	75 (58.6%)	1146 (53.0%)
No	603 (44.7%)	273 (44.9%)	29 (38.2%)	49 (38.3%)	954 (44.1%)

Unknown	45 (3.3%)	13 (2.1%)	0	4 (3.1%)	62 (2.9%)
ART status[3]					
Never taken ART	321 (45.7%)	155 (48.1%)	15 (31.9%)	37 (49.3%)	528 (46.1%)
Taking ART	307 (43.7%)	130 (40.4%)	27 (57.4%)	31 (41.3%)	495 (43.2%)
Took ART but stopped	14 (2.0%)	11 (3.4%)	3 (6.4%)	2 (2.7%)	30 (2.6%)
Unknown	60 (8.5%)	26 (8.1%)	2 (4.3%)	5 (6.7%)	93 (8.1%)
CD4+ T cells, cells/mm[3]					
Less than or equal to 200	228 (32.5%)	130 (40.4%)	14 (29.8%)	26 (34.7%)	398 (34.7%)
201 – 350	129 (18.4%)	50 (15.5%)	7 (14.9%)	10 (13.3%)	196 (17.1%)
351 – 500	102 (14.5%)	37 (11.5%)	6 (12.8%)	9 (12.0%)	154 (13.4%)
More than 500	77 (11.0%)	40 (12.4%)	7 (14.9%)	6 (8.0%)	130 (11.3%)
Unknown	166 (23.6%)	65 (20.2%)	13 (27.7%)	24 (32.0%)	268 (23.4%)
Mtbc lineage					
Lineage 1	79 (5.9%)	NA	7 (9.2%)	NA	NA
Lineage 2	72 (5.3%)	NA	4 (5.3%)	NA	NA
Lineage 3	11 (0.8%)	NA	1 (1.3%)	NA	NA
Lineage 4	1188 (88.0%)	NA	64 (84.2%)	NA	NA
pDST results					
INH-resistant	104 (7.7%)	43 (7.1%)	NA	NA	NA
RIF-resistant	72 (5.3%)	28 (4.6%)	NA	NA	NA
MDR-TB[4]	49 (3.6%)	19 (3.1%)	NA	NA	NA
EMB-resistant	58 (4.3%)	18 (3.0%)	NA	NA	NA
PZA-resistant	27 (2.0%)	17 (2.8%)	NA	NA	NA
WGS drug resistance prediction					
INH-resistant	103 (7.6%)	NA	16 (21.1%)	NA	NA
RIF-resistant	120 (8.9%)	NA	21 (27.6%)	NA	NA
MDR-TB[4]	68 (5.0%)	NA	11 (14.5%)	NA	NA
EMB-resistant	36 (2.7%)	NA	6 (7.9%)	NA	NA
PZA-resistant	102 (7.6%)	NA	17 (22.4%)	NA	NA

Abbreviations: pDST: phenotypic drug-susceptibility testing. IPT: isoniazid preventive therapy. AFB: acid-fast bacilli. ART: antiretroviral therapy. Mtbc: mycobacterium tuberculosis complex. WGS: whole-genome sequencing. INH: isoniazid. RIF: rifampin. EMB: ethambutol. PZA: pyrazinamide. MDR-TB: multidrug-resistant TB.
[1]pDST available for at least one of the following anti-TB drugs: INH, RIF, EMB, PZA.
[2]pDST not available for all first-line anti-TB drugs.
[3]Among participants with HIV-infection only.
[4]TB resistant to at least INH and RIF

Table 2. Sensitivity and specificity of whole genome sequencing to predict culture-based phenotypic first-line anti-tuberculous drug resistance

Drug	Phenotypically resistant		Phenotypically susceptible		Sensitivity (95% CI)	Specificity (95% CI)
	WGS-based resistance	WGS-based susceptible	WGS-based resistance	WGS-based susceptible		
All participants						
INH[1]	70	33	16	1212	68.0 (58.4, 76.2)	98.7 (97.9, 99.2)
RIF[1]	65	6	12	1222	91.5 (82.8, 96.1)	99.0 (98.3, 99.4)
EMB[1]	24	34	12	1278	41.4 (29.6, 54.2)	99.1 (98.4, 99.5)
PZA[2]	9	17	23	666	34.6 (19.4, 53.8)	96.7 (95.0, 97.8)
Participants without HIV						
INH	33	13	9	543	71.7 (57.5, 82.7)	98.4 (96.9, 99.1)
RIF	28	5	5	546	84.8 (69.1, 93.3)	99.1 (97.9, 99.6)
EMB	13	16	8	566	44.8 (28.4, 62.5)	98.6 (97.3, 99.3)
PZA[3]	6	7	12	317	46.2 (23.2, 70.9)	96.4 (93.7, 97.9)
Participants with HIV						
INH[1]	36	20	6	627	64.3 (51.2, 75.5)	99.1 (97.9, 99.6)
RIF[1]	37	1	7	635	97.4 (86.5, 99.5)	98.9 (97.8, 99.5)
EMB[1]	11	16	4	669	40.7 (24.5, 59.3)	99.4 (98.5, 99.8)
PZA[4]	3	10	11	329	23.1 (8.2, 50.3)	96.8 (94.3, 98.2)

Participants with HIV-associated immunosuppression (CD4+ >200 cells/mm^3)

INH	22	7	4	269	75.9 (57.9, 87.8)	98.5 (96.3, 99.4)
RIF	23	1	4	268	95.8 (79.8, 99.3)	98.5 (96.3, 99.4)
EMB	8	7	2	291	53.3 (30.1, 75.2)	99.3 (97.5, 99.8)
PZA[5]	3	2	7	148	60.0 (23.1, 88.2)	95.5 (91.0, 97.8)
Participants with severe HIV-associated immunosuppression (CD4+ ≤200 cells/mm^3)						
INH[6]	6	10	1	206	37.5 (18.5, 61.4)	99.5 (97.3, 99.9)
RIF[6]	10	0	1	211	100.0 (72.2, 100.0)	99.5 (97.4, 99.9)
EMB[6]	3	5	0	219	37.5 (13.7, 69.4)	100.0 (98.3, 100.0)
PZA[7]	0	6	4	100	0 (0, 39.0)	96.2 (90.5, 98.5)

Abbreviations: WGS: whole-genome sequencing; INH: isoniazid; RIF: rifampin; EMB: ethambutol; PZA: pyrazinamide.
[1]Excluded 2 participants with unknown pDST results for INH, RIF, and EMB
[2]Excluded 602 participants with unknown pDST results for PZA
[3]Excluded 249 participants with unknown pDST results for PZA
[4]Excluded 333 participants with unknown pDST results for PZA
[5]Excluded 143 participants with unknown pDST results for PZA
[6]Excluded 1 participant with unknown pDST results for INH, RIF, and EMB
[7]Excluded 111 participants with unknown pDST results for PZA

Table 3. Mutations associated with first-line anti-TB drugs

Gene	Mutations	pDST resistant (TP)			pDST sensitive (FP)		
		HIV-infected	HIV-uninfected	HIV unknown	HIV-infected	HIV-uninfected	HIV unknown
Isoniazid							
Number of individuals		36	33	1	6	9	1
katG	S315T*	27 (75.0)	25 (75.8)	1 (100.0)	2 (33.3)	5 (55.6)	
katG	R418_*				1 (16.7)		
fabG1	-15c>t*	6 (16.7)	6 (18.2)		2 (33.3)		
fabG1	-8t>c*	1 (2.8)	1 (3.0)			1 (11.1)	
fabG1	-8t>a*	6 (16.7)	2 (6.1)			1 (11.1)	
fabG1	L203L*	2 (5.6)	1 (3.0)		1 (16.7)	3 (33.3)	
inhA	I21V	1 (2.8)	2 (6.1)			2 (22.2)	
inhA	S94A	1 (2.8)	1 (3.0)				1 (100.0)
Rifampin							
Number of individuals		37	28	0	7	5	0
rpoB	S450L*	17 (45.9)	12 (42.9)		1 (14.3)	2 (40.0)	
rpoB	S450W*	2 (5.4)	4 (14.3)		1 (14.3)		
rpoB	L452P*	1 (2.7)					
rpoB	H445L*	8 (21.6)	3 (10.7)		3 (42.9)	1 (20.0)	
rpoB	H445D*	5 (13.5)	3 (10.7)		1 (14.3)	1 (20.0)	
rpoB	H445Y*	2 (5.4)	1 (3.6)				
rpoB	D435Y*	1 (2.7)					
rpoB	D435G*	1 (2.7)					
rpoB	D435V*	1 (2.7)	2 (7.1)				
rpoB	L430P*		1 (3.6)		1 (14.3)		
rpoB	V170F*		1 (3.6)			1 (20.0)	
rpoB	838_del_c					1 (20.0)	
rpoB	1292_ins_cca*		1 (3.6)				
rpoB	1348_ins_cgg	1 (2.7)					
Ethambutol							
Number of individuals		11	13	0	4	8	0
embA	-12c>t*		1 (7.7)				
embA	-16c>t	2 (18.2)					

embA	-16c>g					1 (12.5)	
embB	Q497R*	4 (36.4)	7 (53.8)		4 (100.0)	4 (50.0)	
embB	Q497K*		1 (7.7)				
embB	G406A*	2 (18.2)	1 (7.7)				
embB	G406S*	1 (9.1)					
embB	D354A*		1 (7.7)				
embB	M306I*		3 (23.1)			1 (12.5)	
embB	M306V*	3 (27.3)	1 (7.7)			3 (37.5)	
			Pyrazinamide				
Number of individuals		3	6	0	11	12	0
pncA	L159P					1 (8.3)	
pncA	R154G*		1 (16.7)				
pncA	T135P*		1 (16.7)				
pncA	A134V*	1 (33.3)					
pncA	G132S*	1 (33.3)					
pncA	G108E					1 (8.3)	
pncA	F106S*					1 (8.3)	
pncA	G97N		1 (16.7)				
pncA	G97C*					1 (8.3)	
pncA	G97S*					1 (8.3)	
pncA	T76P*	1 (33.3)					
pncA	L35P		1 (16.7)		1 (9.1)		
pncA	L35R		2 (33.3)		10 (90.9)	9 (75.0)	
pncA	A28D					1 (8.3)	
pncA	Q10P*					1 (8.3)	

*Mutations categorized as “Associated with Resistance” in the Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance by the World Health Organization (2021): https://www.who.int/publications/i/item/9789240028173

Table 4. Association between HIV-associated immunosuppression and low-fitness Mtbc variants among persons with INH and RIF resistant Mtbc isolates

	Low fitness	High fitness	OR (95% CI)
Isoniazid			
HIV positive			
No	11	22	-
Yes	16	20	1.60 (0.61, 4.33)
CD4+ T-cells			
HIV negative	11	22	-
≤ 200	2	4	1.00 (0.12, 6.00)
> 200	10	12	1.67 (0.55, 5.12)
Unknown	4	4	2.00 (0.40, 10.01)
CD4+ T-cells[1] (+50 cells/mm^3)			0.99 (0.85, 1.14)
Rifampin			
HIV positive			
No	16	12	-
Yes	20	17	0.88 (0.32, 2.37)
CD4+ T-cells[2]			
HIV negative	16	12	-
≤ 200	7	3	1.75 (0.39, 9.45)
> 200	12	11	0.82 (0.27, 2.49)
CD4+ T-cells[1] (+50 cells/mm^3)			0.95 (0.82, 1.09)

Abbreviations: Mtbc: *Mycobacterium tuberculosis*-complex; OR: odds ratio; CI: confidence interval.
[1]Treated as a continuous variable.
[2]Excluded 4 participants with unknown CD4+ T-cell counts due to small cell size (≤1)

Table 5. Association between HIV-associated immunosuppression and low-fitness TB variants among all Mtbc isolates

	Sensitive	High fitness	Low fitness	High fitness vs. sensitive OR (95% CI)	Low fitness vs. sensitive OR (95% CI)
Isoniazid					
HIV positive					
No	543	22	11	-	-
Yes	627	20	16	0.79 (0.43, 1.46)	1.26 (0.58, 2.74)
CD4+ T-cells					
HIV negative	543	22	11	-	-
≤ 200	206	4	2	0.48 (0.16, 1.41)	0.48 (0.11, 2.18)
> 200	269	12	10	1.10 (0.54, 2.26)	1.84 (0.77, 4.38)
Unknown	152	4	4	0.65 (0.22, 1.91)	1.30 (0.41, 4.14)
CD4+ T-cells[1] (+50 cells/mm^3)				1.11 (1.01, 1.22)	1.10 (0.98, 1.23)
Rifampin					
HIV positive					
No	546	12	16	-	-
Yes	635	17	20	1.22 (0.58, 2.57)	1.08 (0.55, 2.10)
CD4+ T-cells[2]					
HIV negative	546	12	16	-	-
≤ 200	211	3	7	0.65 (0.18, 2.32)	1.13 (0.46, 2.79)
> 200	268	11	12	1.87 (0.81, 4.29)	1.53 (0.71, 3.28)
CD4+ T-cells[1] (+50 cells/mm^3)				1.11 (1.00, 1.22)	1.05 (0.95, 1.16)

Abbreviations: Mtbc: *Mycobacterium tuberculosis*-complex; OR: odds ratio; CI: confidence interval.
[1]Treated as a continuous variable.
[2]Excluded 4 participants with unknown CD4+ T-cell counts due to small cell size (≤1)

Table 6. Association between HIV infection and low fitness TB variants among cases of INH and RIF resistance (modified definition of high or low fitness Mtbc variants)

	High fitness	Low fitness	OR (95% CI)
Isoniazid			
HIV infection			
No	25	8	-
Yes	27	9	1.04 (0.35, 3.18)
CD4+ T cells			
HIV negative	25	8	-
≤ 200	5	1	0.62 (0.03, 4.70)
> 200	17	5	0.92 (0.24, 3.25)
Unknown	5	3	1.87 (0.33, 9.55)
CD4+ T cells (+50 cells)			0.98 (0.81, 1.16)
Rifampin			
HIV infection			
No	12	16	-
Yes	17	20	0.88 (0.32, 2.37)
CD4+ T-cells			
HIV negative	12	16	-
≤ 200	3	7	1.75 (0.39, 9.45)
> 200	11	12	0.82 (0.27, 2.49)
Unknown	3	1	0.25 (0.01, 2.23)
CD4+ T-cells[2] (+50 cells)			0.95 (0.82, 1.09)

Abbreviations: OR: odds ratio. CI: confidence interval.

Table 7. Association between HIV infection and low fitness TB variants among all Mtbc isolates (modified definition of high or low fitness Mtbc variants)

	Sensitive	High fitness	Low fitness	High fitness vs. sensitive OR (95% CI)	Low fitness vs. sensitive OR (95% CI)
Isoniazid					
HIV infection					
No	543	25	8	-	-
Yes	627	27	9	0.94 (0.54, 1.63)	0.97 (0.37, 2.54)
CD4+ T cells					
HIV negative	543	25	8	-	-
≤ 200	206	5	1	0.53 (0.20, 1.40)	0.33 (0.04, 2.65)
> 200	269	17	5	1.37 (0.73, 2.59)	1.26 (0.41, 3.89)
Unknown	152	5	3	0.71 (0.27, 1.90)	1.34 (0.35, 5.11)
CD4+ T cells (+50 cells)				1.11 (1.02, 1.21)	1.09 (0.92, 1.28)
Rifampin					
HIV infection					
No	546	12	16	-	-
Yes	635	17	20	1.22 (0.58, 2.57)	1.08 (0.55, 2.10)
CD4+ T cells					
HIV negative	546	12	16	-	-
≤ 200	211	3	7	0.65 (0.18, 2.32)	1.13 (0.46, 2.79)
> 200	268	11	12	1.87 (0.81, 4.29)	1.53 (0.71, 3.28)
Unknown	156	3	1	0.87 (0.24, 3.14)	0.22 (0.03, 1.66)
CD4+ T cells (+50 cells)				1.11 (1.00, 1.22)	1.05 (0.95, 1.16)

Abbreviations: OR: odds ratio. CI: confidence interval.

Table 8. Association between HIV infection and low fitness TB variants among all Mtbc isolates (including isolates with discordant genotypic and phenotypic drug susceptibility testing results)

	Sensitive	High fitness	Low fitness	High fitness vs. sensitive OR (95% CI)	Low fitness vs. sensitive OR (95% CI)
Isoniazid					
HIV infection					
No	556	26	16	-	-
Yes	649	22	20	0.72 (0.41, 1.29)	1.07 (0.55, 2.09)
CD4+ T cells					
HIV negative	556	26	16	-	-
≤ 200	217	4	3	0.39 (0.14, 1.14)	0.48 (0.14, 1.67)
> 200	276	13	13	1.01 (0.51, 1.99)	1.64 (0.78, 3.45)
Unknown	156	5	4	0.69 (0.26, 1.81)	0.89 (0.29, 2.70)
CD4+ T cells (+50 cells)				1.11 (1.01, 1.22)	1.09 (0.98, 1.20)
Rifampin					
HIV infection					
No	551	14	19	-	-
Yes	638	18	26	1.11 (0.55, 2.25)	1.18 (0.65, 2.16)
CD4+ T cells					
HIV negative	551	14	19	-	-
≤ 200	212	4	7	0.74 (0.24, 2.28)	0.96 (0.40, 2.31)
> 200	269	11	16	1.61 (0.72, 3.59)	1.72 (0.87, 3.41)
Unknown	157	3	3	0.75 (0.21, 2.65)	0.55 (0.16, 1.90)
CD4+ T cells (+50 cells)				1.09 (0.98, 1.21)	1.09 (1.00, 1.18)

Abbreviations: OR: odds ratio. CI: confidence interval.

Table 9. Distribution of Mtbc lineages by enrollment district in Botswana, 2012-2016

	District		
	Gaborone n (%)	**Ghanzi n (%)**	**Total n (%)**
Lineage 1 Indo-Oceanic	83 (7.6)	3 (1.2)	86 (6.4)
L1.1	22 (2.0)	0	22 (1.6)
L1.2	61 (5.6)	3 (1.2)	64 (4.7)
Lineage 2 East Asian			
L2.2	65 (5.9)	7 (2.7)	72 (5.3)
Lineage 3 East African-Indian	12 (1.1)	0	12 (0.9)
Lineage 4 Euro-American	938 (85.4)	246 (96.1)	1184 (87.4)
L4.1	2 (0.2)	0	2 (0.1)
L4.1.1	136 (12.4)	5 (2.0)	141 (10.4)
L4.1.2	129 (11.7)	20 (7.8)	149 (11.0)
L4.2.2	1 (0.1)	0	1 (0.1)
L4.3.2	107 (9.7)	10 (3.9)	117 (8.6)
L4.3.3	24 (2.2)	3 (1.2)	27 (2.0)
L4.3.4	229 (20.9)	171 (66.8)	400 (29.5)
L4.4	164 (14.9)	25 (9.8)	189 (14.0)
L4.6	2 (0.2)	0	2 (0.1)
L4.7	13 (1.2)	1 (0.4)	14 (1.0)
L4.8	84 (7.7)	8 (3.1)	92 (6.8)
L4.9	47 (4.3)	3 (1.2)	50 (3.7)
Total	1098 (100)	256 (100)	1354 (100)

Table 10. Participant characteristics by *Mycobacterium tuberculosis* complex (Mtbc) lineages in Botswana

	Mtbc Lineage				
	L1	**L2**	**L3**	**L4**	**Total**
	n (%)	**n (%)**	**n (%)**	**n (%)**	**n (%)**
Gender					
Female	35 (40.7)	38 (52.8)	6 (50.0)	522 (44.1)	601 (44.4)
Male	51 (59.3)	34 (47.2)	6 (50.0)	662 (55.9)	753 (55.6)
Age					
<=15	2 (2.3)	3 (4.2)	1 (8.3)	21 (1.8)	27 (2.0)
16-24	10 (11.6)	9 (12.5)	1 (8.3)	226 (19.1)	246 (18.2)
25-40	47 (54.7)	46 (63.9)	7 (58.3)	632 (53.4)	732 (54.1)
41-64	24 (27.9)	14 (19.4)	3 (25.0)	280 (23.6)	321 (23.7)
>=65	3 (3.5)	0	0	25 (2.1)	28 (2.1)
Previous TB					
No	63 (73.3)	57 (79.2)	10 (83.3)	967 (81.7)	1097 (81.0)
Yes	23 (26.7)	15 (20.8)	2 (16.7)	217 (18.3)	257 (19.0)
Isoniazid preventive therapy					
No	83 (96.5)	64 (88.9)	12 (100.0)	1127 (95.2)	1286 (95.0)
Yes	2 (2.3)	5 (6.9)	0	42 (3.5)	49 (3.6)
Unknown	1 (1.2)	3 (4.2)	0	15 (1.3)	19 (1.4)
AFB smear results					
Negative	5 (5.8)	7 (9.7)	3 (25.0)	76 (6.4)	91 (6.7)
Positive	61 (70.9)	46 (63.9)	8 (66.7)	752 (63.5)	867 (64.0)
Not done	20 (23.3)	19 (26.4)	1 (8.3)	356 (30.1)	396 (29.2)
TB type					
Pulmonary	79 (91.9)	65 (90.3)	11 (91.7)	1075 (90.8)	1230 (90.8)
Extrapulmonary	7 (8.1)	6 (8.3)	0	105 (8.9)	118 (8.7)
Unknown	0	1 (1.4)	1 (8.3)	4 (0.3)	6 (0.4)
HIV status					
Negative	27 (31.4)	27 (37.5)	5 (41.7)	533 (45.0)	592 (43.7)
Positive	56 (65.1)	42 (58.3)	7 (58.3)	616 (52.0)	721 (53.2)
Unknown	3 (3.5)	3 (4.2)	0	35 (3.0)	41 (3.0)
ART status[1]					
Never taken ART	26 (46.4)	17 (40.5)	1 (14.3)	278 (45.1)	322 (44.7)
Taking ART	25 (44.6)	20 (47.6)	3 (42.9)	277 (45.0)	325 (45.1)
Took ART but stopped	1 (1.8)	1 (2.4)	2 (28.6)	11 (1.8)	15 (2.1)
Unknown	4 (7.1)	4 (9.5)	1 (14.3)	50 (8.1)	59 (8.2)
History of incarceration					
No	83 (96.5)	69 (95.8)	10 (83.3)	1116 (94.3)	1278 (94.4)
Yes	3 (3.5)	3 (4.2)	2 (16.7)	66 (5.6)	74 (5.5)
Unknown	0	0	0	2 (0.1)	2 (0.1)
Currently smoking					
No	66 (76.7)	58 (80.6)	10 (83.3)	880 (74.3)	1014 (74.9)
Yes	20 (23.3)	14 (19.4)	2 (16.7)	304 (25.7)	340 (25.1)

Abbreviations: AFB: acid-fast bacilli. ART: antiretroviral therapy.

[1] Among HIV-positive participants only.

Table 11. Time to most recent common anscestor (tMRCA) of the *Mycobacterium tuberculosis* complex (Mtbc) lineages in Botswana

Lineage	**tMRCA (Posterior mean)**	**tMRCA (Posterior median)**	**95% HPD interval**
L1	1727	1736	1607 – 1828
L2	1900	1903	1854 – 1938
L4.1.1	1729	1739	1587 – 1842
L4.1.2	1695	1705	1546 – 1817
L4.3.2	1941	1943	1911 – 1966
L4.3.4	1808	1814	1723 – 1881
L4.4	1739	1748	1621 – 1847
L4.8	1853	1858	1782 – 1912

Abbreviations: HPD: highest posterior density.

Table 12. Genomic cluster rate (based on a 5-SNPs cutoff) of the *Mycobacterium tuberculosis*-complex (Mtbc) lineages in Botswana

Mtbc lineage	**Cluster rate**	**Crude OR (95% CI)**	**Adjusted[1] OR (95% CI)**
L1	25/86	-	-
L2	42/72	3.42 (1.78, 6.69)	3.26 (1.68, 6.47)
L4.1.1	83/141	3.49 (1.99, 6.28)	3.44 (1.94, 6.25)
L4.1.2	74/149	2.41 (1.38, 4.29)	2.26 (1.28, 4.07)
L4.3.2	52/117	1.95 (1.09, 3.56)	1.91 (1.05, 3.52)
L4.3.4	190/400	2.21 (1.35, 3.71)	1.68 (1.00, 2.89)
L4.4	82/189	1.87 (1.09, 3.27)	1.78 (1.03, 3.16)
L4.8	55/92	3.63 (1.96, 6.86)	3.27 (1.74, 6.26)

Abbreviations: SNP: single nucleotide polymorphism. OR: odds ratio. CI: confidence interval.
[1] Adjusted for age, gender, HIV status, and enrollment district.

Table 13. Genomic cluster rate (based on a 12-SNPs cutoff) of the *Mycobacterium tuberculosis*-complex (Mtbc) lineages in Botswana

Mtbc lineage	**Cluster rate**	**Crude OR (95% CI)**	**Adjusted[1] OR (95% CI)**
L1	38/86	-	-
L2	57/72	4.80 (2.40, 10.01)	4.44 (2.19, 9.37)
L4.1.1	103/141	3.42 (1.96, 6.07)	3.48 (1.97, 6.25)
L4.1.2	96/149	2.29 (1.34, 3.95)	2.08 (1.20, 3.64)
L4.3.2	79/117	2.63 (1.48, 4.70)	2.52 (1.41, 4.55)
L4.3.4	260/400	2.35 (1.47, 3.78)	1.49 (0.91, 2.47)
L4.4	124/189	2.41 (1.44, 4.08)	2.21 (1.30, 3.79)
L4.8	73/92	4.85 (2.54, 9.57)	4.32 (2.24, 8.62)

Abbreviations: SNP: single nucleotide polymorphism. OR: odds ratio. CI: confidence interval.
[1] Adjusted for age, gender, HIV status, and enrollment district.

Table 14. HIV co-infection prevalence by *Mycobacterium tuberculosis*-complex (Mtbc) lineages in Botswana

Mtbc lineage	**HIV prevalence**	**Crude OR (95% CI)**	**Adjusted[1] OR (95% CI)**
L1	56/86	-	-
L2	42/72	0.75 (0.38, 1.46)	0.69 (0.33, 1.44)
L4.1.1	71/141	0.50 (0.28, 0.87)	0.53 (0.28, 0.99)
L4.1.2	95/149	0.88 (0.49, 1.55)	1.01 (0.54, 1.90)
L4.3.2	72/117	0.81 (0.44, 1.46)	0.89 (0.46, 1.71)
L4.3.4	182/400	0.44 (0.26, 0.72)	0.67 (0.38, 1.17)
L4.4	104/189	0.63 (0.36, 1.07)	0.64 (0.35, 1.16)
L4.8	42/92	0.43 (0.23, 0.80)	0.45 (0.22, 0.88)

Abbreviations: OR: odds ratio. CI: confidence interval.
[1] Adjusted for age, gender, and enrollment district.

Figures

Figure 1. Percentage of sequence reads mapped to reference *Mycobacterium tuberculosis*-complex (Mtbc) genome by HIV status and CD4+ T-cell count

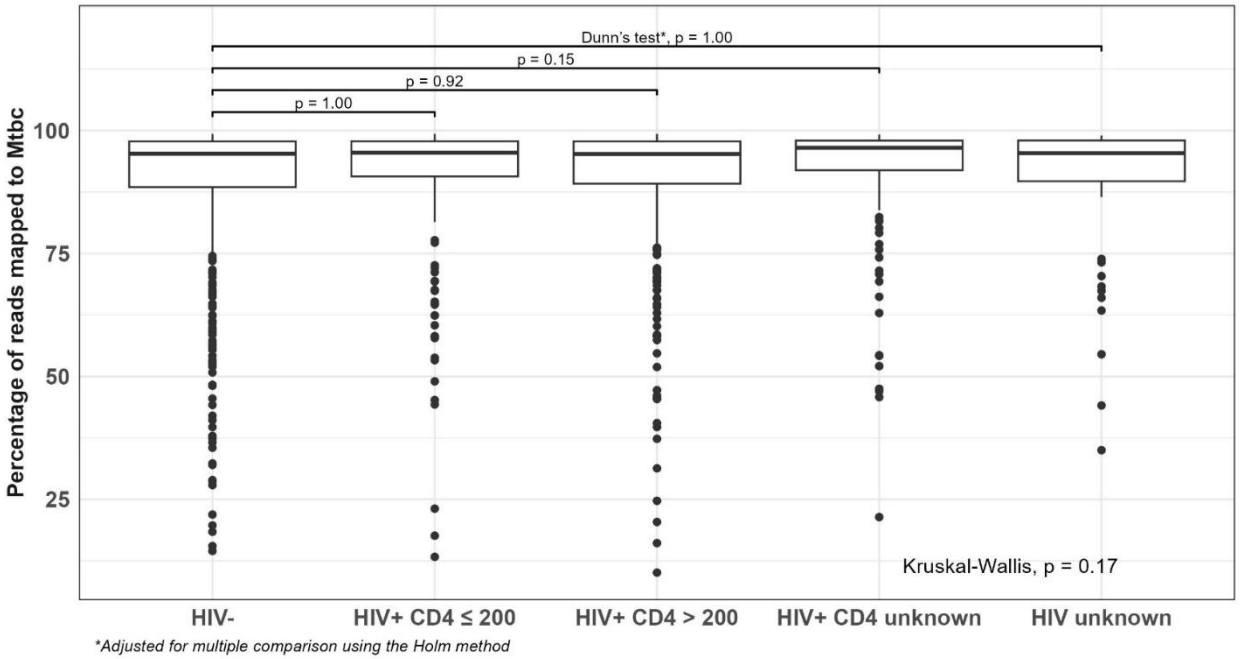


Figure 2. Percentage of sequence reads mapped to reference *Mycobacterium tuberculosis*-complex (Mtbc) genome among phenotypically isoniazid-resistant isolates

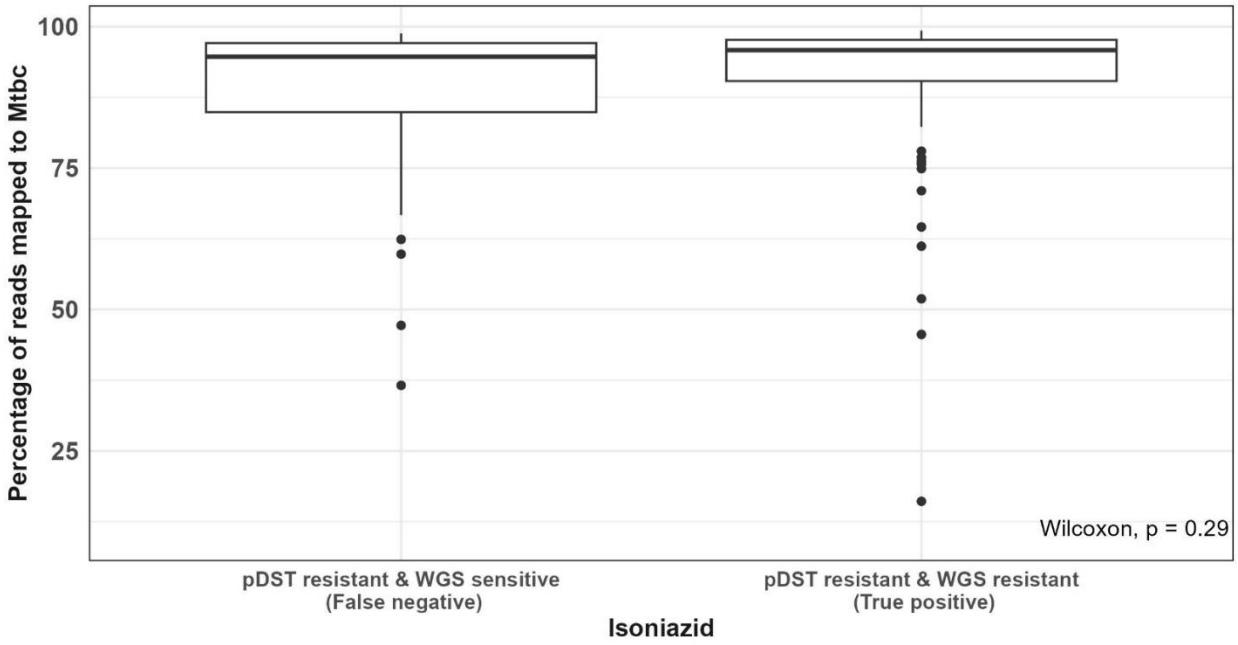


Figure 3. Phylogeny of *Mycobacterium tuberculosis*-complex (Mtbc) isolates collected in Botswana, 2012-2016 (n = 1,354): A. Phylogeny of Lineage 1 Indo-Oceanic Mtbc isolates (n = 86). B. Phylogeny of Lineage 2 East Asian Mtbc isolates (n = 72). C. Phylogeny of Lineage 3 East African-Indian Mtbc isolates (n = 12). D. Phylogeny of Lineage 4 Euro-American Mtbc isolates (n = 1184)

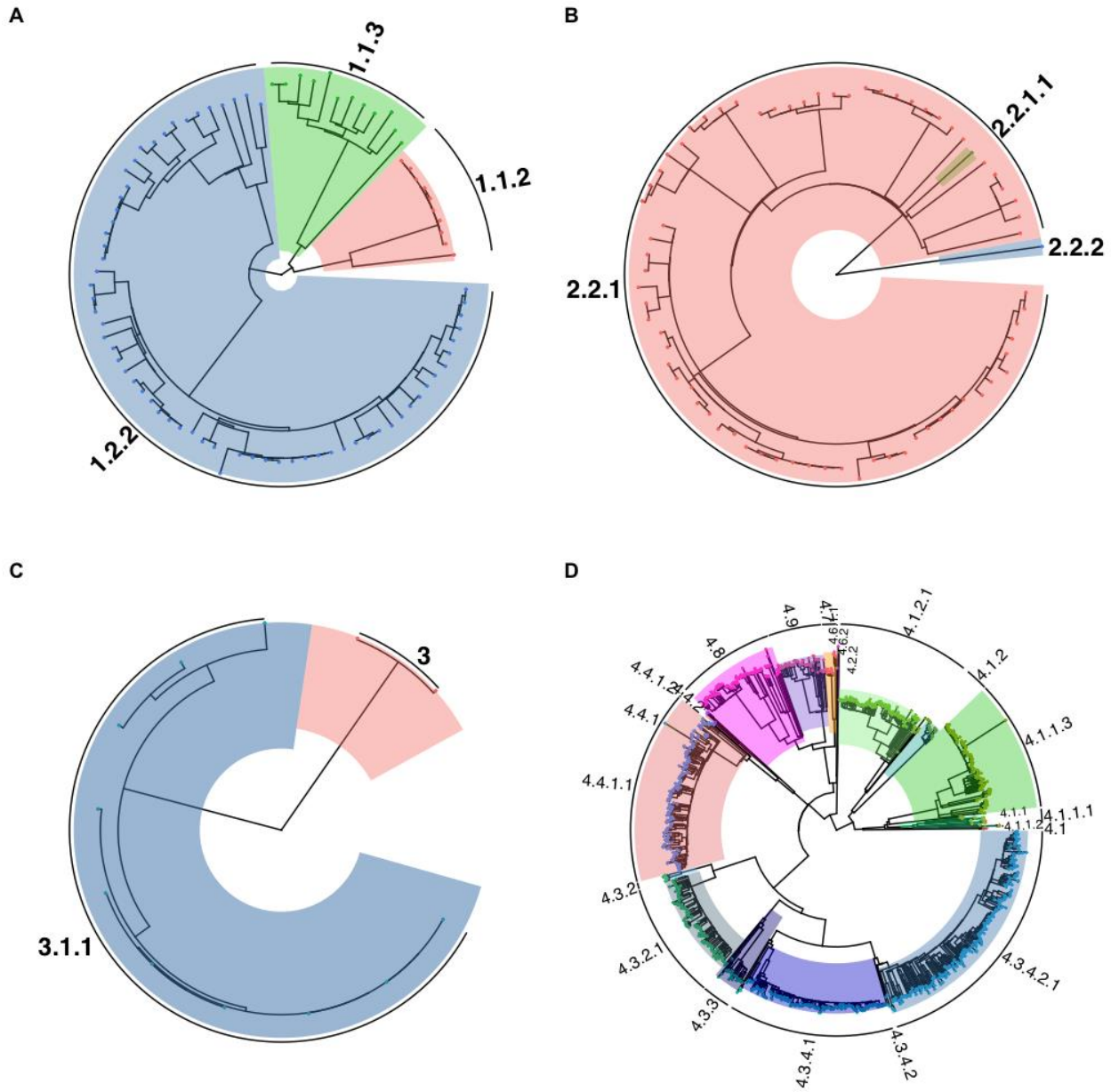


Figure 4. Inferred effective population size over time for *Mycobacterium tuberculosis*-complex (Mtbc) lineages circulating in Botswana

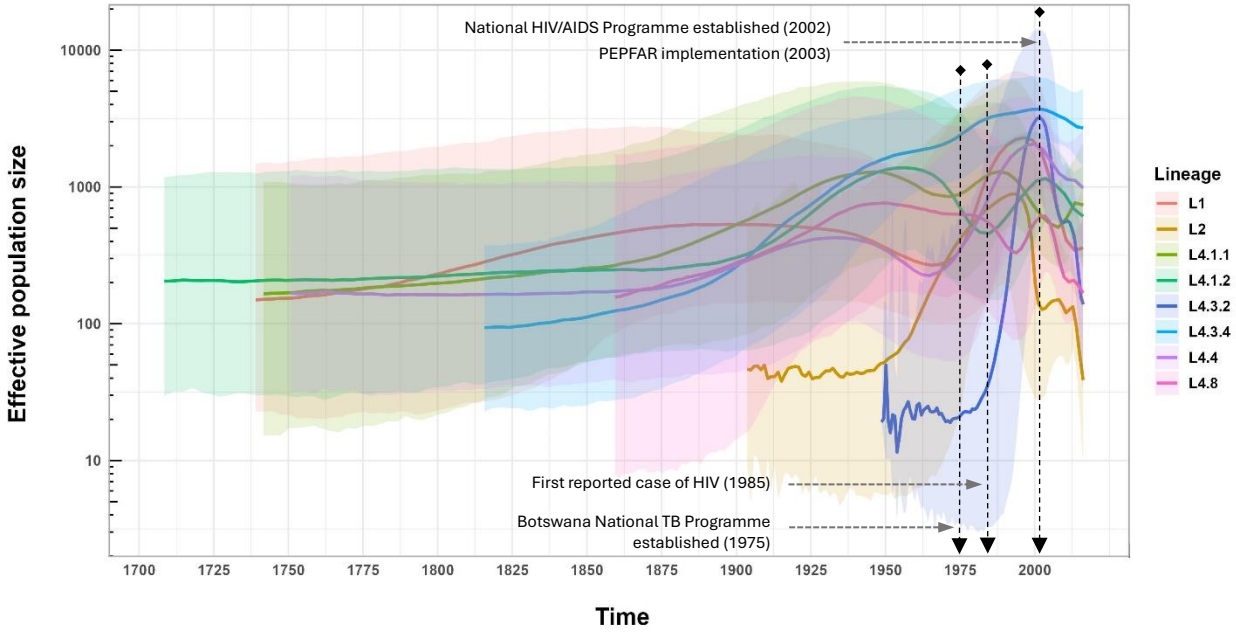
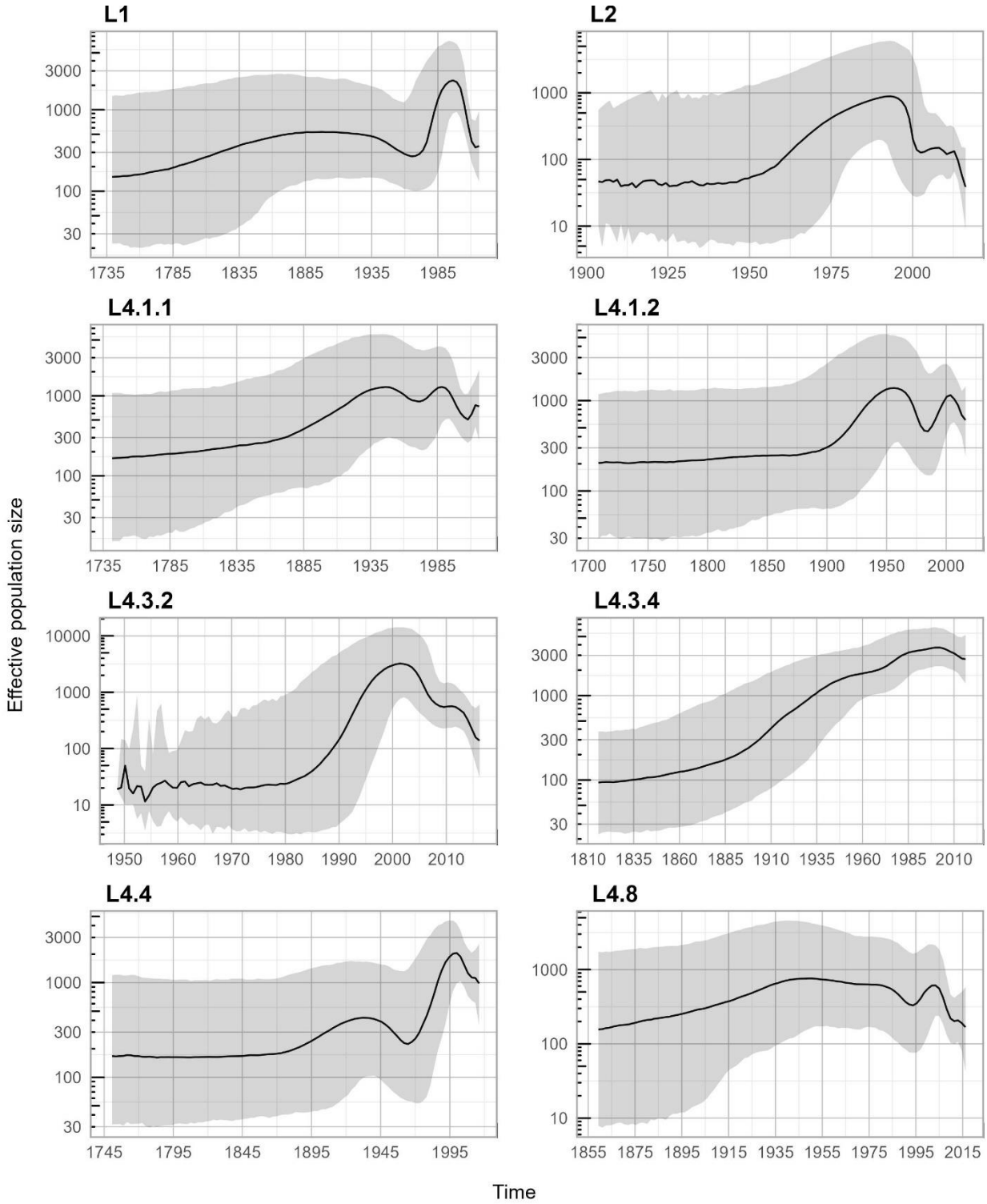


Figure 5. Lineage-specific effective population size over time for *Mycobacterium tuberculosis*-complex (Mtbc) lineages circulating in Botswana



References

1. World Health Organization. Global Tuberculosis Report 2021 Geneva: World Health Organization 2021. Available from: https://www.who.int/publications/i/item/9789240037021.

2. Bos KI, Harkins KM, Herbig A, Coscolla M, Weber N, Comas I, Forrest SA, Bryant JM, Harris SR, Schuenemann VJ, Campbell TJ, Majander K, Wilbur AK, Guichon RA, Wolfe Steadman DL, Cook DC, Niemann S, Behr MA, Zumarraga M, Bastida R, Huson D, Nieselt K, Young D, Parkhill J, Buikstra JE, Gagneux S, Stone AC, Krause J. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. Nature. 2014;514(7523):494-7. Epub 20140820. doi: 10.1038/nature13591. PubMed PMID: 25141181; PMCID: PMC4550673.

3. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B, Berg S, Thwaites G, Yeboah-Manu D, Bothamley G, Mei J, Wei L, Bentley S, Harris SR, Niemann S, Diel R, Aseffa A, Gao Q, Young D, Gagneux S. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. Nat Genet. 2013;45(10):1176-82. Epub 20130901. doi: 10.1038/ng.2744. PubMed PMID: 23995134; PMCID: PMC3800747.

4. Coscolla M, Gagneux S, Menardo F, Loiseau C, Ruiz-Rodriguez P, Borrell S, Otchere ID, Asante-Poku A, Asare P, Sanchez-Buso L, Gehre F, Sanoussi CN, Antonio M, Affolabi D, Fyfe J, Beckert P, Niemann S, Alabi AS, Grobusch MP, Kobbe R, Parkhill J, Beisel C, Fenner L, Bottger EC, Meehan CJ, Harris SR, de Jong BC, Yeboah-Manu D, Brites D. Phylogenomics of Mycobacterium africanum reveals a new lineage and a complex evolutionary history. Microb Genom. 2021;7(2). doi: 10.1099/mgen.0.000477. PubMed PMID: 33555243; PMCID: PMC8208692.

5. Ngabonziza JCS, Loiseau C, Marceau M, Jouet A, Menardo F, Tzfadia O, Antoine R, Niyigena EB, Mulders W, Fissette K, Diels M, Gaudin C, Duthoy S, Ssengooba W, Andre E, Kaswa MK, Habimana YM, Brites D, Affolabi D, Mazarati JB, de Jong BC, Rigouts L, Gagneux S, Meehan CJ, Supply P. A sister lineage of the Mycobacterium tuberculosis complex discovered in the African Great Lakes region. Nat Commun. 2020;11(1):2917. Epub 20200609. doi: 10.1038/s41467-020-16626-6. PubMed PMID: 32518235; PMCID: PMC7283319.

6. Coll F, McNerney R, Guerra-Assuncao JA, Glynn JR, Perdigao J, Viveiros M, Portugal I, Pain A, Martin N, Clark TG. A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. Nat Commun. 2014;5:4812. Epub 20140901. doi: 10.1038/ncomms5812. PubMed PMID: 25176035; PMCID: PMC4166679.

7. Freschi L, Vargas R, Jr., Husain A, Kamal SMM, Skrahina A, Tahseen S, Ismail N, Barbova A, Niemann S, Cirillo DM, Dean AS, Zignol M, Farhat MR. Population structure, biogeography and transmissibility of Mycobacterium tuberculosis. Nat Commun. 2021;12(1):6099. Epub 20211020. doi: 10.1038/s41467-021-26248-1. PubMed PMID: 34671035; PMCID: PMC8528816.

8. Gagneux S. Host-pathogen coevolution in human tuberculosis. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2012;367(1590):850-9. doi: 10.1098/rstb.2011.0316. PubMed PMID: 22312052; PMCID: PMC3267123.

9. Guerra-Assuncao JA, Crampin AC, Houben RM, Mzembe T, Mallard K, Coll F, Khan P, Banda L, Chiwaya A, Pereira RP, McNerney R, Fine PE, Parkhill J, Clark TG, Glynn JR. Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a high prevalence area. Elife. 2015;4. Epub 2015/03/04. doi: 10.7554/eLife.05166. PubMed PMID: 25732036; PMCID: PMC4384740.

10. Coscolla M, Gagneux S. Consequences of genomic diversity in Mycobacterium tuberculosis. Semin Immunol. 2014;26(6):431-44. Epub 20141022. doi: 10.1016/j.smim.2014.09.012. PubMed PMID: 25453224; PMCID: PMC4314449.

11. World Health Organization. Global Tuberculosis Report 2023 Geneva: World Health Organization; 2023. Available from: https://www.who.int/teams/global-tuberculosis-programme/tb-reports/global-tuberculosis-report-2023.

12. World Health Organization. The end TB strategy 2015. Available from: https://www.who.int/publications-detail-redirect/WHO-HTM-TB-2015.19.

13. World Health Organization. Global tuberculosis report 2022 2022. Available from: https://www.who.int/publications/i/item/9789240061729.

14. Menzies HJ, Moalosi G, Anisimova V, Gammino V, Sentle C, Bachhuber MA, Bile E, Radisowa K, Kachuwaire O, Basotli J, Maribe T, Makombe R, Shepherd J, Kim B, Samandari T, El-Halabi S, Chirenda J, Cain KP. Increase in anti-tuberculosis drug resistance in Botswana: results from the fourth National Drug Resistance Survey. Int J Tuberc Lung Dis. 2014;18(9):1026-33. Epub 2014/09/06. doi: 10.5588/ijtld.13.0749. PubMed PMID: 25189548; PMCID: PMC5531285.

15. Dean AS, Tosas Auguet O, Glaziou P, Zignol M, Ismail N, Kasaeva T, Floyd K. 25 years of surveillance of drug-resistant tuberculosis: achievements, challenges, and way forward. The Lancet Infectious diseases. 2022;22(7):e191-e6. Epub 20220303. doi: 10.1016/S1473-3099(21)00808-2. PubMed PMID: 35248168; PMCID: PMC8893725.

16. Gillespie SH. Evolution of drug resistance in Mycobacterium tuberculosis: clinical and molecular perspective. Antimicrobial agents and chemotherapy. 2002;46(2):267-74. doi: 10.1128/AAC.46.2.267-274.2002. PubMed PMID: 11796329; PMCID: PMC127054.

17. Andersson DI, Levin BR. The biological cost of antibiotic resistance. Curr Opin Microbiol. 1999;2(5):489-93. doi: 10.1016/s1369-5274(99)00005-3. PubMed PMID: 10508723.

18. Gagneux S, Long CD, Small PM, Van T, Schoolnik GK, Bohannan BJ. The competitive cost of antibiotic resistance in Mycobacterium tuberculosis. Science (New York, NY). 2006;312(5782):1944-6. Epub 2006/07/01. doi: 10.1126/science.1124410. PubMed PMID: 16809538.

19. Gagneux S. Fitness cost of drug resistance in Mycobacterium tuberculosis. Clin Microbiol Infect. 2009;15 Suppl 1:66-8. Epub 2009/03/03. doi: 10.1111/j.1469-0691.2008.02685.x. PubMed PMID: 19220360.

20. Gagneux S, Burgos MV, DeRiemer K, Encisco A, Munoz S, Hopewell PC, Small PM, Pym AS. Impact of bacterial genetics on the transmission of isoniazid-resistant Mycobacterium tuberculosis. PLoS Pathog. 2006;2(6):e61. Epub 20060616. doi: 10.1371/journal.ppat.0020061. PubMed PMID: 16789833; PMCID: PMC1479046.

21. Mesfin YM, Hailemariam D, Biadgilign S, Kibret KT. Association between HIV/AIDS and multi-drug resistance tuberculosis: a systematic review and meta-analysis. PLoS One. 2014;9(1):e82235. Epub 2014/01/15. doi: 10.1371/journal.pone.0082235. PubMed PMID: 24416139; PMCID: PMC3885391.

22. Sultana ZZ, Hoque FU, Beyene J, Akhlak-Ul-Islam M, Khan MHR, Ahmed S, Hawlader DH, Hossain A. HIV infection and multidrug resistant tuberculosis: a systematic review and meta-analysis. BMC Infect Dis. 2021;21(1):51. Epub 2021/01/13. doi: 10.1186/s12879-020-05749-2. PubMed PMID: 33430786; PMCID: PMC7802168.

23. Shah NS, Auld SC, Brust JC, Mathema B, Ismail N, Moodley P, Mlisana K, Allana S, Campbell A, Mthiyane T, Morris N, Mpangase P, van der Meulen H, Omar SV, Brown TS,

Narechania A, Shaskina E, Kapwata T, Kreiswirth B, Gandhi NR. Transmission of Extensively Drug-Resistant Tuberculosis in South Africa. The New England journal of medicine. 2017;376(3):243-53. Epub 2017/01/19. doi: 10.1056/NEJMoa1604544. PubMed PMID: 28099825; PMCID: PMC5330208.

24. Gandhi NR, Weissman D, Moodley P, Ramathal M, Elson I, Kreiswirth BN, Mathema B, Shashkina E, Rothenberg R, Moll AP, Friedland G, Sturm AW, Shah NS. Nosocomial transmission of extensively drug-resistant tuberculosis in a rural hospital in South Africa. J Infect Dis. 2013;207(1):9-17. Epub 2012/11/21. doi: 10.1093/infdis/jis631. PubMed PMID: 23166374; PMCID: PMC3523793.

25. Khan PY, Yates TA, Osman M, Warren RM, van der Heijden Y, Padayatchi N, Nardell EA, Moore D, Mathema B, Gandhi N, Eldholm V, Dheda K, Hesseling AC, Mizrahi V, Rustomjee R, Pym A. Transmission of drug-resistant tuberculosis in HIV-endemic settings. The Lancet Infectious diseases. 2019;19(3):e77-e88. Epub 2018/12/18. doi: 10.1016/S1473-3099(18)30537-1. PubMed PMID: 30554996; PMCID: PMC6474238.

26. Loiseau C, Brites D, Reinhard M, Zurcher K, Borrell S, Ballif M, Fenner L, Cox H, Rutaihwa LK, Wilkinson RJ, Yotebieng M, Carter EJ, Abimiku A, Marcy O, Gotuzzo E, Avihingsanon A, Zetola N, Doulla B, Bottger EC, Egger M, Gagneux S. HIV Coinfection Is Associated with Low-Fitness rpoB Variants in Rifampicin-Resistant Mycobacterium tuberculosis. Antimicrobial agents and chemotherapy. 2020;64(10). Epub 2020/07/29. doi: 10.1128/AAC.00782-20. PubMed PMID: 32718966; PMCID: PMC7508592.

27. Ssengooba W, Lukoye D, Meehan CJ, Kateete DP, Joloba ML, de Jong BC, Cobelens FG, van Leth F. Tuberculosis resistance-conferring mutations with fitness cost among HIV-positive

individuals in Uganda. Int J Tuberc Lung Dis. 2017;21(5):531-6. Epub 2017/04/13. doi: 10.5588/ijtld.16.0544. PubMed PMID: 28399968.

28. World Health Organization. Use of targeted next-generation sequencing to detect drug-resistant tuberculosis 2023.

29. Jouet A, Gaudin C, Badalato N, Allix-Beguec C, Duthoy S, Ferre A, Diels M, Laurent Y, Contreras S, Feuerriegel S, Niemann S, Andre E, Kaswa MK, Tagliani E, Cabibbe A, Mathys V, Cirillo D, de Jong BC, Rigouts L, Supply P. Deep amplicon sequencing for culture-free prediction of susceptibility or resistance to 13 anti-tuberculous drugs. The European respiratory journal. 2021;57(3). Epub 20210318. doi: 10.1183/13993003.02338-2020. PubMed PMID: 32943401; PMCID: PMC8174722.

30. Cabibbe AM, Spitaleri A, Battaglia S, Colman RE, Suresh A, Uplekar S, Rodwell TC, Cirillo DM. Application of Targeted Next-Generation Sequencing Assay on a Portable Sequencing Platform for Culture-Free Detection of Drug-Resistant Tuberculosis from Clinical Samples. J Clin Microbiol. 2020;58(10). Epub 20200922. doi: 10.1128/JCM.00632-20. PubMed PMID: 32727827; PMCID: PMC7512157.

31. Wang Q, Modongo C, Allender C, Engelthaler DM, Warren RM, Zetola NM, Shin SS. Utility of Targeted, Amplicon-Based Deep Sequencing To Detect Resistance to First-Line Tuberculosis Drugs in Botswana. Antimicrobial agents and chemotherapy. 2019;63(11). Epub 2019/08/14. doi: 10.1128/AAC.00982-19. PubMed PMID: 31405858; PMCID: PMC6811446.

32. Zetola NM, Modongo C, Moonan PK, Click E, Oeltmann JE, Shepherd J, Finlay A. Protocol for a population-based molecular epidemiology study of tuberculosis transmission in a high HIV-burden setting: the Botswana Kopanyo study. BMJ open. 2016;6(5):e010046. Epub

2016/05/11. doi: 10.1136/bmjopen-2015-010046. PubMed PMID: 27160840; PMCID: PMC4874111.

33. Avert. HIV and AIDS in Botswana 2019. Available from: https://www.avert.org/professionals/hiv-around-world/sub-saharan-africa/botswana.

34. Mupfumi L, Moyo S, Shin SS, Wang Q, Zetola N, Molebatsi K, Nnawa J, Kgwaadira BT, Bewlay L, Chebani T, Iketleng T, Mogashoa T, Makhema J, Musonda RM, Essex M, Kasvosve I, Gaseitsiwe S. High incidence of tuberculosis in the first year of antiretroviral therapy in the Botswana National antiretroviral therapy programme between 2011 and 2015. AIDS. 2019;33(15):2415-22. doi: 10.1097/QAD.0000000000002363. PubMed PMID: 31764106; PMCID: PMC7903483.

35. Moonan PK, Zetola NM, Tobias JL, Basotli J, Boyd R, Click ES, Dima M, Fane O, Finlay AM, Ogopotse M, Wen XJ, Modongo C, Oeltmann JE. A Neighbor-Based Approach to Identify Tuberculosis Exposure, the Kopanyo Study. Emerging infectious diseases. 2020;26(5):1010-3. Epub 2020/04/21. doi: 10.3201/eid2605.191568. PubMed PMID: 32310058; PMCID: PMC7181937.

36. Botswana S. Botswana Population and Housing Census Gaborone, Botswana: Statistics Botswana; 2011. Available from: https://www.statsbots.org.bw/sites/default/files/2011%20Population%20and%20housing%20Census.pdf.

37. Zetola NM, Moonan PK, Click E, Oeltmann JE, Basotli J, Wen X-J, Boyd R, Tobias JL, Finlay A, Modongo C. Population-Based Geospatial and Molecular Epidemiologic Study of Tuberculosis Transmission Dynamics, Botswana, 2012–2016. Emerging infectious diseases. 2021;27(3):835-44. doi: 10.3201/eid2703.203840.

38. Ardito F, Posteraro B, Sanguinetti M, Zanetti S, Fadda G. Evaluation of BACTEC Mycobacteria Growth Indicator Tube (MGIT 960) automated system for drug susceptibility testing of Mycobacterium tuberculosis. J Clin Microbiol. 2001;39(12):4440-4. doi: 10.1128/JCM.39.12.4440-4444.2001. PubMed PMID: 11724858; PMCID: PMC88562.

39. World Health Organization. Technical Report on critical concentrations for drug susceptibility testing of medicines used in the treatment of drug-resistant tuberculosis 2018. Available from: https://www.who.int/publications/i/item/WHO-CDS-TB-2018.5.

40. Kohl TA, Utpatel C, Schleusener V, De Filippo MR, Beckert P, Cirillo DM, Niemann S. MTBseq: a comprehensive pipeline for whole genome sequence analysis of Mycobacterium tuberculosis complex isolates. PeerJ. 2018;6:e5895. Epub 2018/11/28. doi: 10.7717/peerj.5895. PubMed PMID: 30479891; PMCID: PMC6238766.

41. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268-74. Epub 20141103. doi: 10.1093/molbev/msu300. PubMed PMID: 25371430; PMCID: PMC4271533.

42. Seifert M, Catanzaro D, Catanzaro A, Rodwell TC. Genetic mutations associated with isoniazid resistance in Mycobacterium tuberculosis: a systematic review. PLoS One. 2015;10(3):e0119628. Epub 2015/03/24. doi: 10.1371/journal.pone.0119628. PubMed PMID: 25799046; PMCID: PMC4370653.

43. Rostamian M, Kooti S, Abiri R, Khazayel S, Kadivarian S, Borji S, Alvandi A. Prevalence of Mycobacterium tuberculosis mutations associated with isoniazid and rifampicin resistance: A systematic review and meta-analysis. J Clin Tuberc Other Mycobact Dis.

2023;32:100379. Epub 20230601. doi: 10.1016/j.jctube.2023.100379. PubMed PMID: 37389010; PMCID: PMC10302537.

44. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol. 2018;4(1):vey016. Epub 2018/06/27. doi: 10.1093/ve/vey016. PubMed PMID: 29942656; PMCID: PMC6007674.

45. Xu Y, Stockdale JE, Naidu V, Hatherell H, Stimson J, Stagg HR, Abubakar I, Colijn C. Transmission analysis of a large tuberculosis outbreak in London: a mathematical modelling study using genomic data. Microb Genom. 2020;6(11). Epub 2020/11/12. doi: 10.1099/mgen.0.000450. PubMed PMID: 33174832; PMCID: PMC7725332.

46. Roetzer A, Diel R, Kohl TA, Ruckert C, Nubel U, Blom J, Wirth T, Jaenicke S, Schuback S, Rusch-Gerdes S, Supply P, Kalinowski J, Niemann S. Whole genome sequencing versus traditional genotyping for investigation of a Mycobacterium tuberculosis outbreak: a longitudinal molecular epidemiological study. PLoS medicine. 2013;10(2):e1001387. Epub 20130212. doi: 10.1371/journal.pmed.1001387. PubMed PMID: 23424287; PMCID: PMC3570532.

47. Minin VN, Bloomquist EW, Suchard MA. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol Biol Evol. 2008;25(7):1459-71. Epub 20080411. doi: 10.1093/molbev/msn090. PubMed PMID: 18408232; PMCID: PMC3302198.

48. Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol. 2005;22(5):1185-92. Epub 20050209. doi: 10.1093/molbev/msi103. PubMed PMID: 15703244.

49. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P, Smith EG, Peto TE. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. The Lancet Infectious diseases. 2013;13(2):137-46. Epub 2012/11/20. doi: 10.1016/S1473-3099(12)70277-3. PubMed PMID: 23158499; PMCID: PMC3556524.

50. Zhang X, Martinez E, Lam C, Crighton T, Sim E, Gall M, Donnan EJ, Marais BJ, Sintchenko V. Exploring programmatic indicators of tuberculosis control that incorporate routine Mycobacterium tuberculosis sequencing in low incidence settings: a comprehensive (2017-2021) patient cohort analysis. Lancet Reg Health West Pac. 2023;41:100910. Epub 20230927. doi: 10.1016/j.lanwpc.2023.100910. PubMed PMID: 37808343; PMCID: PMC10550799.

51. R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available from: https://www.R-project.org/.

52. Hazbon MH, Brimacombe M, Bobadilla del Valle M, Cavatore M, Guerrero MI, Varma-Basil M, Billman-Jacobe H, Lavender C, Fyfe J, Garcia-Garcia L, Leon CI, Bose M, Chaves F, Murray M, Eisenach KD, Sifuentes-Osornio J, Cave MD, Ponce de Leon A, Alland D. Population genetics study of isoniazid resistance mutations and evolution of multidrug-resistant Mycobacterium tuberculosis. Antimicrobial agents and chemotherapy. 2006;50(8):2640-9. Epub 2006/07/28. doi: 10.1128/AAC.00112-06. PubMed PMID: 16870753; PMCID: PMC1538650.

53. The CRyPTIC Consortium and the 100,000 Genomes Project. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. The New England journal of medicine.

2018;379(15):1403-15. Epub 2018/10/04. doi: 10.1056/NEJMoa1800474. PubMed PMID: 30280646; PMCID: PMC6121966.

54. Xiao YX, Liu KH, Lin WH, Chan TH, Jou R. Whole-genome sequencing-based analyses of drug-resistant Mycobacterium tuberculosis from Taiwan. Sci Rep. 2023;13(1):2540. Epub 20230213. doi: 10.1038/s41598-023-29652-3. PubMed PMID: 36781938; PMCID: PMC9925824.

55. Madrazo-Moya CF, Cancino-Munoz I, Cuevas-Cordoba B, Gonzalez-Covarrubias V, Barbosa-Amezcua M, Soberon X, Muniz-Salazar R, Martinez-Guarneros A, Backer C, Zarrabal-Meza J, Sampieri-Ramirez C, Enciso-Moreno A, Lauzardo M, Comas I, Zenteno-Cuevas R. Whole genomic sequencing as a tool for diagnosis of drug and multidrug-resistance tuberculosis in an endemic region in Mexico. PLoS One. 2019;14(6):e0213046. Epub 20190605. doi: 10.1371/journal.pone.0213046. PubMed PMID: 31166945; PMCID: PMC6550372.

56. Feliciano CS, Namburete EI, Rodrigues Placa J, Peronni K, Dippenaar A, Warren RM, Silva WA, Jr., Bollela VR. Accuracy of whole genome sequencing versus phenotypic (MGIT) and commercial molecular tests for detection of drug-resistant Mycobacterium tuberculosis isolated from patients in Brazil and Mozambique. Tuberculosis (Edinb). 2018;110:59-67. Epub 20180405. doi: 10.1016/j.tube.2018.04.003. PubMed PMID: 29779775; PMCID: PMC9328935.

57. Chen X, He G, Wang S, Lin S, Chen J, Zhang W. Evaluation of Whole-Genome Sequence Method to Diagnose Resistance of 13 Anti-tuberculosis Drugs and Characterize Resistance Genes in Clinical Multi-Drug Resistance Mycobacterium tuberculosis Isolates From China. Front Microbiol. 2019;10:1741. Epub 2019/08/17. doi: 10.3389/fmicb.2019.01741. PubMed PMID: 31417530; PMCID: PMC6685394.

58. Nguyen TNA, Anton-Le Berre V, Banuls AL, Nguyen TVA. Molecular Diagnosis of Drug-Resistant Tuberculosis; A Literature Review. Front Microbiol. 2019;10:794. Epub 2019/05/07. doi: 10.3389/fmicb.2019.00794. PubMed PMID: 31057511; PMCID: PMC6477542.

59. Campbell PJ, Morlock GP, Sikes RD, Dalton TL, Metchock B, Starks AM, Hooks DP, Cowan LS, Plikaytis BB, Posey JE. Molecular detection of mutations associated with first- and second-line drug resistance compared with conventional drug susceptibility testing of Mycobacterium tuberculosis. Antimicrobial agents and chemotherapy. 2011;55(5):2032-41. Epub 2011/02/09. doi: 10.1128/AAC.01550-10. PubMed PMID: 21300839; PMCID: PMC3088277.

60. Metcalfe JZ, Streicher E, Theron G, Colman RE, Allender C, Lemmer D, Warren R, Engelthaler DM. Cryptic Microheteroresistance Explains Mycobacterium tuberculosis Phenotypic Resistance. American journal of respiratory and critical care medicine. 2017;196(9):1191-201. Epub 2017/06/15. doi: 10.1164/rccm.201703-0556OC. PubMed PMID: 28614668; PMCID: PMC5694839.

61. Chedore P, Bertucci L, Wolfe J, Sharma M, Jamieson F. Potential for erroneous results indicating resistance when using the Bactec MGIT 960 system for testing susceptibility of Mycobacterium tuberculosis to pyrazinamide. J Clin Microbiol. 2010;48(1):300-1. Epub 20091118. doi: 10.1128/JCM.01775-09. PubMed PMID: 19923479; PMCID: PMC2812260.

62. Kim SJ. Drug-susceptibility testing in tuberculosis: methods and reliability of results. The European respiratory journal. 2005;25(3):564-9. doi: 10.1183/09031936.05.00111304. PubMed PMID: 15738303.

63. Jagielski T, Bakula Z, Brzostek A, Minias A, Stachowiak R, Kalita J, Napiorkowska A, Augustynowicz-Kopec E, Zaczek A, Vasiliauskiene E, Bielecki J, Dziadek J. Characterization of Mutations Conferring Resistance to Rifampin in Mycobacterium tuberculosis Clinical Strains.

Antimicrobial agents and chemotherapy. 2018;62(10). Epub 20180924. doi: 10.1128/AAC.01093-18. PubMed PMID: 30061294; PMCID: PMC6153850.

64. Andre E, Goeminne L, Cabibbe A, Beckert P, Kabamba Mukadi B, Mathys V, Gagneux S, Niemann S, Van Ingen J, Cambau E. Consensus numbering system for the rifampicin resistance-associated rpoB gene mutations in pathogenic mycobacteria. Clin Microbiol Infect. 2017;23(3):167-72. Epub 20160921. doi: 10.1016/j.cmi.2016.09.006. PubMed PMID: 27664776.

65. Blakemore R, Story E, Helb D, Kop J, Banada P, Owens MR, Chakravorty S, Jones M, Alland D. Evaluation of the analytical performance of the Xpert MTB/RIF assay. J Clin Microbiol. 2010;48(7):2495-501. Epub 2010/05/28. doi: 10.1128/JCM.00128-10. PubMed PMID: 20504986; PMCID: PMC2897495.

66. Dorman SE, Chihota VN, Lewis JJ, van der Meulen M, Mathema B, Beylis N, Fielding KL, Grant AD, Churchyard GJ. Genotype MTBDRplus for direct detection of Mycobacterium tuberculosis and drug resistance in strains from gold miners in South Africa. J Clin Microbiol. 2012;50(4):1189-94. Epub 2012/01/13. doi: 10.1128/JCM.05723-11. PubMed PMID: 22238443; PMCID: PMC3318561.

67. Li MC, Lu J, Lu Y, Xiao TY, Liu HC, Lin SQ, Xu D, Li GL, Zhao XQ, Liu ZG, Zhao LL, Wan KL. rpoB Mutations and Effects on Rifampin Resistance in Mycobacterium tuberculosis. Infect Drug Resist. 2021;14:4119-28. Epub 20211005. doi: 10.2147/IDR.S333433. PubMed PMID: 34675557; PMCID: PMC8502021.

68. Khan AS, Phelan JE, Khan MT, Ali S, Qasim M, Napier G, Campino S, Ahmad S, Cabral-Marques O, Zhang S, Rahman H, Wei DQ, Clark TG, Khan TA. Characterization of rifampicin-resistant Mycobacterium tuberculosis in Khyber Pakhtunkhwa, Pakistan. Sci Rep.

2021;11(1):14194. Epub 20210709. doi: 10.1038/s41598-021-93501-4. PubMed PMID: 34244539; PMCID: PMC8270973.

69. Zeng MC, Jia QJ, Tang LM. rpoB gene mutations in rifampin-resistant Mycobacterium tuberculosis isolates from rural areas of Zhejiang, China. J Int Med Res. 2021;49(3):300060521997596. doi: 10.1177/0300060521997596. PubMed PMID: 33715498; PMCID: PMC7952843.

70. Claassens M, Dreyer V, Nepolo E, Mokomele Q, van Rooyen G, Ruswa N, Gunther G, Niemann S. Whole-Genome Sequencing for Resistance Prediction and Transmission Analysis of Mycobacterium tuberculosis Complex Strains from Namibia. Microbiol Spectr. 2022;10(5):e0158622. Epub 20220927. doi: 10.1128/spectrum.01586-22. PubMed PMID: 36165641; PMCID: PMC9603870.

71. Shea J, Halse TA, Kohlerschmidt D, Lapierre P, Modestil HA, Kearns CH, Dworkin FF, Rakeman JL, Escuyer V, Musser KA. Low-Level Rifampin Resistance and rpoB Mutations in Mycobacterium tuberculosis: an Analysis of Whole-Genome Sequencing and Drug Susceptibility Test Data in New York. J Clin Microbiol. 2021;59(4). Epub 20210319. doi: 10.1128/JCM.01885-20. PubMed PMID: 32999007; PMCID: PMC8579749.

72. Torres JN, Paul LV, Rodwell TC, Victor TC, Amallraja AM, Elghraoui A, Goodmanson AP, Ramirez-Busby SM, Chawla A, Zadorozhny V, Streicher EM, Sirgel FA, Catanzaro D, Rodrigues C, Gler MT, Crudu V, Catanzaro A, Valafar F. Novel katG mutations causing isoniazid resistance in clinical M. tuberculosis isolates. Emerg Microbes Infect. 2015;4(7):e42. Epub 2015/08/08. doi: 10.1038/emi.2015.42. PubMed PMID: 26251830; PMCID: PMC4522615.

73. Moga S, Bobosha K, Fikadu D, Zerihun B, Diriba G, Amare M, Kempker RR, Blumberg HM, Abebe T. Diagnostic performance of the GenoType MTBDRplus VER 2.0 line probe assay

for the detection of isoniazid resistant Mycobacterium tuberculosis in Ethiopia. PLoS One. 2023;18(4):e0284737. Epub 20230426. doi: 10.1371/journal.pone.0284737. PubMed PMID: 37099514; PMCID: PMC10132600.

74. Banerjee A, Dubnau E, Quemard A, Balasubramanian V, Um KS, Wilson T, Collins D, de Lisle G, Jacobs WR, Jr. inhA, a gene encoding a target for isoniazid and ethionamide in Mycobacterium tuberculosis. Science (New York, NY). 1994;263(5144):227-30. doi: 10.1126/science.8284673. PubMed PMID: 8284673.

75. Van Deun A, Maug AK, Salim MA, Das PK, Sarker MR, Daru P, Rieder HL. Short, highly effective, and inexpensive standardized treatment of multidrug-resistant tuberculosis. American journal of respiratory and critical care medicine. 2010;182(5):684-92. Epub 20100504. doi: 10.1164/rccm.201001-0077OC. PubMed PMID: 20442432.

76. Katiyar SK, Bihari S, Prakash S, Mamtani M, Kulkarni H. A randomised controlled trial of high-dose isoniazid adjuvant therapy for multidrug-resistant tuberculosis. Int J Tuberc Lung Dis. 2008;12(2):139-45. PubMed PMID: 18230245.

77. Borrell S, Gagneux S. Strain diversity, epistasis and the evolution of drug resistance in Mycobacterium tuberculosis. Clin Microbiol Infect. 2011;17(6):815-20. doi: 10.1111/j.1469-0691.2011.03556.x. PubMed PMID: 21682802; PMCID: PMC3122159.

78. Fenner L, Egger M, Bodmer T, Altpeter E, Zwahlen M, Jaton K, Pfyffer GE, Borrell S, Dubuis O, Bruderer T, Siegrist HH, Furrer H, Calmy A, Fehr J, Stalder JM, Ninet B, Bottger EC, Gagneux S, Swiss HIVCS, the Swiss Molecular Epidemiology of Tuberculosis Study G. Effect of mutation and genetic background on drug resistance in Mycobacterium tuberculosis. Antimicrobial agents and chemotherapy. 2012;56(6):3047-53. Epub 20120402. doi: 10.1128/AAC.06460-11. PubMed PMID: 22470121; PMCID: PMC3370767.

79. Dodd PJ, Shaweno D, Ku CC, Glaziou P, Pretorius C, Hayes RJ, MacPherson P, Cohen T, Ayles H. Transmission modeling to infer tuberculosis incidence prevalence and mortality in settings with generalized HIV epidemics. Nat Commun. 2023;14(1):1639. Epub 20230324. doi: 10.1038/s41467-023-37314-1. PubMed PMID: 36964130; PMCID: PMC10037365.

80. Ramogola-Masire D, Poku O, Mazhani L, Ndwapi N, Misra S, Arscott-Mills T, Blank L, Ho-Foster A, Becker TD, Yang L. Botswana's HIV response: Policies, context, and future directions. J Community Psychol. 2020;48(3):1066-70. Epub 20200117. doi: 10.1002/jcop.22316. PubMed PMID: 31951283; PMCID: PMC7103557.

81. Wilkinson E, Engelbrecht S, de Oliveira T. History and origin of the HIV-1 subtype C epidemic in South Africa and the greater southern African region. Sci Rep. 2015;5:16897. Epub 20151117. doi: 10.1038/srep16897. PubMed PMID: 26574165; PMCID: PMC4648088.

82. Corbett EL, Watt CJ, Walker N, Maher D, Williams BG, Raviglione MC, Dye C. The growing burden of tuberculosis: global trends and interactions with the HIV epidemic. Archives of internal medicine. 2003;163(9):1009-21. doi: 10.1001/archinte.163.9.1009. PubMed PMID: 12742798.

83. Song J, Okano JT, Ponce J, Busang L, Seipone K, Valdano E, Blower S. Population mobility and the development of Botswana's generalized HIV epidemic: a network analysis. medRxiv. 2023. Epub 20230202. doi: 10.1101/2023.02.01.23285339. PubMed PMID: 36778345; PMCID: PMC9915826.

84. UNAIDS. Botswana: country factsheets Geneva, Switzerland: UNAIDS; 2004. Available from: https://www.unaids.org/en/regionscountries/countries/botswana.

85. Stuckler D, Basu S, McKee M, Lurie M. Mining and risk of tuberculosis in sub-Saharan Africa. Am J Public Health. 2011;101(3):524-30. Epub 20100601. doi: 10.2105/AJPH.2009.175646. PubMed PMID: 20516372; PMCID: PMC3036676.

86. Basu S, Stuckler D, Gonsalves G, Lurie M. The production of consumption: addressing the impact of mineral mining on tuberculosis in southern Africa. Global Health. 2009;5:11. Epub 20090929. doi: 10.1186/1744-8603-5-11. PubMed PMID: 19785769; PMCID: PMC2770998.

87. The World Bank. Incidence of tuberculosis in Botswana: The World Bank Group; 2022. Available from: https://data.worldbank.org/indicator/SH.TBS.INCD?locations=BW.

88. Chin RJ, Sangmanee D, Piergallini L. PEPFAR Funding and Reduction in HIV Infection Rates in 12 Focus Sub-Saharan African Countries: A Quantitative Analysis. Int J MCH AIDS. 2015;3(2):150-8. PubMed PMID: 27621994; PMCID: PMC5005989.

89. Farahani M, Vable A, Lebelonyane R, Seipone K, Anderson M, Avalos A, Chadborn T, Tilahun H, Roumis D, Moeti T, Musuka G, Busang L, Gaolathe T, Malefho KC, Marlink R. Outcomes of the Botswana national HIV/AIDS treatment programme from 2002 to 2010: a longitudinal analysis. Lancet Glob Health. 2014;2(1):e44-50. Epub 20131211. doi: 10.1016/S2214-109X(13)70149-9. PubMed PMID: 25104635.

90. Mine M, Stafford KA, Laws RL, Marima R, Lekone P, Ramaabya D, Makhaola K, Patel HK, Mapondera P, Wray-Gordon F, Agbakwuru C, Okui L, Matroos S, Onyadile E, Ngidi J, Abimiku A, Bagapi K, Nkomo B, Bodika SM, Kim KJ, Moloney M, Mitchell A, Ehoche A, Ussery FL, Hong SY, Keipeile S, Matlhaga M, Mathumo R, Selato R, Charurat ME, Voetsch AC. Progress towards the UNAIDS 95-95-95 targets in the Fifth Botswana AIDS Impact Survey (BAIS V 2021): a nationally representative survey. Lancet HIV. 2024;11(4):e245-e54. Epub 20240308. doi: 10.1016/S2352-3018(24)00003-1. PubMed PMID: 38467135.

91. Statistics Botswana. Botswana AIDS Impact Survey V 2021 (BAIS V): Report. Gaborone, Botswana: National AIDS & Health Promotion Agency; 2021. Available from: https://www.statsbots.org.bw/fifth-botswana-aids-impact-survey-2021-bais-v-report.

92. O'Neill MB, Shockey A, Zarley A, Aylward W, Eldholm V, Kitchen A, Pepperell CS. Lineage specific histories of Mycobacterium tuberculosis dispersal in Africa and Eurasia. Mol Ecol. 2019;28(13):3241-56. Epub 20190709. doi: 10.1111/mec.15120. PubMed PMID: 31066139; PMCID: PMC6660993.

93. Menardo F, Rutaihwa LK, Zwyer M, Borrell S, Comas I, Conceicao EC, Coscolla M, Cox H, Joloba M, Dou HY, Feldmann J, Fenner L, Fyfe J, Gao Q, Garcia de Viedma D, Garcia-Basteiro AL, Gygli SM, Hella J, Hiza H, Jugheli L, Kamwela L, Kato-Maeda M, Liu Q, Ley SD, Loiseau C, Mahasirimongkol S, Malla B, Palittapongarnpim P, Rakotosamimanana N, Rasolofo V, Reinhard M, Reither K, Sasamalo M, Silva Duarte R, Sola C, Suffys P, Batista Lima KV, Yeboah-Manu D, Beisel C, Brites D, Gagneux S. Local adaptation in populations of Mycobacterium tuberculosis endemic to the Indian Ocean Rim. F1000Res. 2021;10:60. Epub 20210201. doi: 10.12688/f1000research.28318.2. PubMed PMID: 33732436; PMCID: PMC7921886.

94. Couvin D, Reynaud Y, Rastogi N. Two tales: Worldwide distribution of Central Asian (CAS) versus ancestral East-African Indian (EAI) lineages of Mycobacterium tuberculosis underlines a remarkable cleavage for phylogeographical, epidemiological and demographical characteristics. PLoS One. 2019;14(7):e0219706. Epub 20190712. doi: 10.1371/journal.pone.0219706. PubMed PMID: 31299060; PMCID: PMC6625721.

95. Holt KE, McAdam P, Thai PVK, Thuong NTT, Ha DTM, Lan NN, Lan NH, Nhu NTQ, Hai HT, Ha VTN, Thwaites G, Edwards DJ, Nath AP, Pham K, Ascher DB, Farrar J, Khor CC,

Teo YY, Inouye M, Caws M, Dunstan SJ. Frequent transmission of the Mycobacterium tuberculosis Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. Nat Genet. 2018;50(6):849-56. Epub 20180521. doi: 10.1038/s41588-018-0117-9. PubMed PMID: 29785015; PMCID: PMC6143168.

96. Liu Q, Ma A, Wei L, Pang Y, Wu B, Luo T, Zhou Y, Zheng HX, Jiang Q, Gan M, Zuo T, Liu M, Yang C, Jin L, Comas I, Gagneux S, Zhao Y, Pepperell CS, Gao Q. China's tuberculosis epidemic stems from historical expansion of four strains of Mycobacterium tuberculosis. Nat Ecol Evol. 2018;2(12):1982-92. Epub 20181105. doi: 10.1038/s41559-018-0680-6. PubMed PMID: 30397300; PMCID: PMC6295914.

97. Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, Blum MG, Rusch-Gerdes S, Mokrousov I, Aleksic E, Allix-Beguec C, Antierens A, Augustynowicz-Kopec E, Ballif M, Barletta F, Beck HP, Barry CE, 3rd, Bonnet M, Borroni E, Campos-Herrero I, Cirillo D, Cox H, Crowe S, Crudu V, Diel R, Drobniewski F, Fauville-Dufaux M, Gagneux S, Ghebremichael S, Hanekom M, Hoffner S, Jiao WW, Kalon S, Kohl TA, Kontsevaya I, Lillebaek T, Maeda S, Nikolayevskyy V, Rasmussen M, Rastogi N, Samper S, Sanchez-Padilla E, Savic B, Shamputa IC, Shen A, Sng LH, Stakenas P, Toit K, Varaine F, Vukovic D, Wahl C, Warren R, Supply P, Niemann S, Wirth T. Evolutionary history and global spread of the Mycobacterium tuberculosis Beijing lineage. Nat Genet. 2015;47(3):242-9. Epub 20150119. doi: 10.1038/ng.3195. PubMed PMID: 25599400; PMCID: PMC11044984.

98. Altan S. Chinese Workers of the World: Colonialism, Chinese Labor, and the Yunnan–Indochina Railway. Stanford: Stanford University Press; 2024.

99. Hanekom M, Gey van Pittius NC, McEvoy C, Victor TC, Van Helden PD, Warren RM. Mycobacterium tuberculosis Beijing genotype: a template for success. Tuberculosis (Edinb).

2011;91(6):510-23. Epub 20110810. doi: 10.1016/j.tube.2011.07.005. PubMed PMID: 21835699.

100. Parwati I, van Crevel R, van Soolingen D. Possible underlying mechanisms for successful emergence of the Mycobacterium tuberculosis Beijing genotype strains. The Lancet Infectious diseases. 2010;10(2):103-11. doi: 10.1016/S1473-3099(09)70330-5. PubMed PMID: 20113979.

101. Gagneux S. Ecology and evolution of Mycobacterium tuberculosis. Nat Rev Microbiol. 2018;16(4):202-13. Epub 20180219. doi: 10.1038/nrmicro.2018.8. PubMed PMID: 29456241.

102. Stucki D, Brites D, Jeljeli L, Coscolla M, Liu Q, Trauner A, Fenner L, Rutaihwa L, Borrell S, Luo T, Gao Q, Kato-Maeda M, Ballif M, Egger M, Macedo R, Mardassi H, Moreno M, Tudo Vilanova G, Fyfe J, Globan M, Thomas J, Jamieson F, Guthrie JL, Asante-Poku A, Yeboah-Manu D, Wampande E, Ssengooba W, Joloba M, Henry Boom W, Basu I, Bower J, Saraiva M, Vaconcellos SEG, Suffys P, Koch A, Wilkinson R, Gail-Bekker L, Malla B, Ley SD, Beck HP, de Jong BC, Toit K, Sanchez-Padilla E, Bonnet M, Gil-Brusola A, Frank M, Penlap Beng VN, Eisenach K, Alani I, Wangui Ndung'u P, Revathi G, Gehre F, Akter S, Ntoumi F, Stewart-Isherwood L, Ntinginya NE, Rachow A, Hoelscher M, Cirillo DM, Skenders G, Hoffner S, Bakonyte D, Stakenas P, Diel R, Crudu V, Moldovan O, Al-Hajoj S, Otero L, Barletta F, Jane Carter E, Diero L, Supply P, Comas I, Niemann S, Gagneux S. Mycobacterium tuberculosis lineage 4 comprises globally distributed and geographically restricted sublineages. Nat Genet. 2016;48(12):1535-43. Epub 20161031. doi: 10.1038/ng.3704. PubMed PMID: 27798628; PMCID: PMC5238942.

103. Mogashoa T, Melamu P, Ley SD, Streicher EM, Iketleng T, Kelentse N, Mupfumi L, Mokomane M, Kgwaadira B, Novitsky V, Kasvosve I, Moyo S, Warren RM, Gaseitsiwe S.

Genetic diversity of Mycobacterium tuberculosis strains circulating in Botswana. PLoS One. 2019;14(5):e0216306. Epub 20190507. doi: 10.1371/journal.pone.0216306. PubMed PMID: 31063472; PMCID: PMC6504092.

104. Chihota VN, Niehaus A, Streicher EM, Wang X, Sampson SL, Mason P, Kallenius G, Mfinanga SG, Pillay M, Klopper M, Kasongo W, Behr MA, Gey van Pittius NC, van Helden PD, Couvin D, Rastogi N, Warren RM. Geospatial distribution of Mycobacterium tuberculosis genotypes in Africa. PLoS One. 2018;13(8):e0200632. Epub 20180801. doi: 10.1371/journal.pone.0200632. PubMed PMID: 30067763; PMCID: PMC6070189.

105. Viegas SO, Machado A, Groenheit R, Ghebremichael S, Pennhag A, Gudo PS, Cuna Z, Miotto P, Hill V, Marrufo T, Cirillo DM, Rastogi N, Kallenius G, Koivula T. Molecular diversity of Mycobacterium tuberculosis isolates from patients with pulmonary tuberculosis in Mozambique. BMC Microbiol. 2010;10:195. Epub 20100721. doi: 10.1186/1471-2180-10-195. PubMed PMID: 20663126; PMCID: PMC2914001.

106. Brynildsrud OB, Pepperell CS, Suffys P, Grandjean L, Monteserin J, Debech N, Bohlin J, Alfsnes K, Pettersson JO, Kirkeleite I, Fandinho F, da Silva MA, Perdigao J, Portugal I, Viveiros M, Clark T, Caws M, Dunstan S, Thai PVK, Lopez B, Ritacco V, Kitchen A, Brown TS, van Soolingen D, O'Neill MB, Holt KE, Feil EJ, Mathema B, Balloux F, Eldholm V. Global expansion of Mycobacterium tuberculosis lineage 4 shaped by colonial migration and local adaptation. Sci Adv. 2018;4(10):eaat5869. Epub 20181017. doi: 10.1126/sciadv.aat5869. PubMed PMID: 30345355; PMCID: PMC6192687.

107. Oliver E, Oliver WH. The Colonisation of South Africa: A unique case. HTS Theological Studies. 2017;73(3):1-8. doi: https://dx.doi.org/10.4102/hts.v73i3.4498.

108. Click ES, Finlay A, Oeltmann JE, Basotli J, Modongo C, Boyd R, Wen XJ, Shepherd J, Moonan PK, Zetola NM. Phylogenetic diversity of Mycobacterium tuberculosis in two geographically distinct locations in Botswana - The Kopanyo Study. Infect Genet Evol. 2020;81:104232. Epub 20200203. doi: 10.1016/j.meegid.2020.104232. PubMed PMID: 32028055.

109. Statistics Botswana. Botswana AIDS Impact Survey IV 2013 (BAIS IV): Report. Gaborone, Botswana: National AIDS & Health Promotion Agency; 2013. Available from: https://www.statsbots.org.bw/sites/default/files/publications/BOTSWANA%20AIDS%20IMPACT%20SURVEY%20IV%202013.pdf.

110. Anderson J, Jarlsberg LG, Grindsdale J, Osmond D, Kawamura M, Hopewell PC, Kato-Maeda M. Sublineages of lineage 4 (Euro-American) Mycobacterium tuberculosis differ in genotypic clustering. Int J Tuberc Lung Dis. 2013;17(7):885-91. doi: 10.5588/ijtld.12.0960. PubMed PMID: 23743309.

111. Du DH, Geskus RB, Zhao Y, Codecasa LR, Cirillo DM, van Crevel R, Pascapurnama DN, Chaidir L, Niemann S, Diel R, Omar SV, Grandjean L, Rokadiya S, Ortitz AT, Lan NH, Ha DTM, Smith EG, Robinson E, Dedicoat M, Nhat LTH, Thwaites GE, Van LH, Thuong NTT, Walker TM. The effect of M. tuberculosis lineage on clinical phenotype. PLOS Glob Public Health. 2023;3(12):e0001788. Epub 20231220. doi: 10.1371/journal.pgph.0001788. PubMed PMID: 38117783; PMCID: PMC10732390.

112. Hatherell HA, Colijn C, Stagg HR, Jackson C, Winter JR, Abubakar I. Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. BMC medicine. 2016;14:21. Epub 20160323. doi: 10.1186/s12916-016-0566-x. PubMed PMID: 27005433; PMCID: PMC4804562.