

Distribution and transmission of *M. tuberculosis* in a high-HIV prevalence city in Malawi: a genomic and spatial analysis

Melanie H. Chitwood¹, Elizabeth L. Corbett², Victor Ndhlovu³, Benjamin Sobkowiak¹, Caroline Colijn⁴, Jason R. Andrews⁵, Rachael M. Burke^{2,6}, Patrick G.T. Cudahy⁷, Peter J. Dodd⁸, Jeffrey W. Imai-Eaton^{9,10}, David M. Engelthaler¹¹, Megan Folkerts¹¹, Helena Feasey^{2,6}, Yu Lan¹, Jen Lewis⁸, Nicolas A Menzies¹², Geoffrey Chipungu¹³, Marriott Nliwasa^{6,13}, Daniel M. Weinberger¹, Joshua L. Warren¹⁴, Joshua A. Salomon¹⁵, Peter MacPherson^{2,6,16,17*}, Ted Cohen^{1*}
*co-senior authors

1. Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven CT USA
2. Clinical Research Department, Faculty of Infectious and Tropical Disease, London School of Hygiene and Tropical Medicine, London, United Kingdom
3. School of Life Sciences and Health Professions, Kamuzu University of Health Sciences, Blantyre, Malawi
4. Department of Mathematics, Simon Fraser University, Burnaby BC, Canada
5. Division of Infectious Diseases and Geographic Medicine, Stanford University School of Medicine, Stanford CA, USA
6. Public Health Group, Malawi-Liverpool-Wellcome Programme, Blantyre, Malawi
7. Division of Infectious Diseases, Department of Internal Medicine, Yale School of Medicine, New Haven, CT, USA
8. Sheffield Centre for Health and Related Research, School of Medicine and Population Health, Sheffield, United Kingdom
9. Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA
10. MRC Centre for Global Infectious Disease Analysis, School of Public Health, Imperial College London, London, United Kingdom
11. Translational Genomics Research Institute, Flagstaff, Arizona USA
12. Center for Health Decision Science and Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Boston, Massachusetts USA
13. Helse Nord Tuberculosis Initiative, Department of Pathology, Kamuzu University of Health Sciences, Blantyre, Malawi
14. Department of Biostatistics, Yale School of Public Health, New Haven CT USA
15. Department of Health Policy, Stanford University School of Medicine, Stanford CA, USA
16. School of Health & Wellbeing, University of Glasgow, Glasgow, United Kingdom
17. Bacterial and Respiratory Pathogens Department, Public Health Scotland, Glasgow, United Kingdom

Abstract

Background

Delays in identifying and treating individuals with infectious tuberculosis (TB) contribute to poor health outcomes and allow ongoing community transmission of *M. tuberculosis* (*Mtb*). Current recommendations for screening for tuberculosis specify community characteristics (e.g., areas with high local tuberculosis prevalence) that can be used to target screening within the general population. However, areas of higher tuberculosis burden are not necessarily areas with higher rates of transmission. We investigated the genomic diversity and transmission of *Mtb* using high-resolution surveillance data in Blantyre, Malawi.

Methods and Findings

We extracted and performed whole genome sequencing on mycobacterial DNA from cultured *M. tuberculosis* isolates obtained from culture-positive tuberculosis cases at the time of tuberculosis (TB) notification in Blantyre, Malawi between 2015-2019. We constructed putative transmission networks identified using TransPhylo and investigated individual and pair-wise demographic, clinical, and spatial factors associated with person-to-person transmission. We found that 56% of individuals with sequenced isolates had a probable direct transmission link to at least one other individual in the study. We identified thirteen putative transmission networks that included five or more individuals. Five of these networks had a single spatial focus of transmission in the city, and each focus centered in a distinct neighborhood in the city. We also found that approximately two-thirds of inferred transmission links occurred between individuals residing in different geographic zones of the city.

Conclusion

While the majority of detected tuberculosis transmission events in Blantyre occurred between people living in different zones, there was evidence of distinct geographical concentration for five transmission networks. These findings suggest that targeted interventions in areas with evidence of localized transmission may be an effective local tactic, but will likely need to be augmented by city-wide interventions to improve case finding and to address social determinants of tuberculosis to have sustained impact.

Author Summary

Why was this study done?

- Tuberculosis (TB) is a major global health threat and a leading cause of death due to infectious disease. Rapid diagnosis and treatment of individuals with TB is vital to reduce the spread of disease.
- If public health programs can identify areas with ongoing TB transmission, resources might be directed toward intervening in those areas to interrupt transmission chains. However, in settings where many people have TB, it is often difficult to differentiate areas with high rates of disease from areas with high rates of local transmission.

What did the researchers do and find?

- We used whole genome sequencing data to infer networks of TB transmission in Blantyre, Malawi. We used individual residence data to identify whether transmission networks were concentrated in specific parts of the city and to describe the amount of transmission that occurred between vs. within distinct parts of the city.
- We found that most TB transmission in Blantyre occurred between individuals who did not live near each other. We also identified five transmission networks which had strong local foci of transmission.

What do these findings mean?

- Because most TB transmission in Blantyre does not occur in concentrated areas, city-wide interventions, such as improving access to TB care services and addressing social determinants of TB, may be needed to improve TB control.
- For areas where there is evidence of local concentrated transmission, additional resources and strategies, such as targeted active case finding, may help to more rapidly reduce transmission and TB incidence.

Introduction

Tuberculosis (TB) is a major global health threat and a leading infectious cause of death. The World Health Organization's (WHO) End TB Strategy aims to reduce global tuberculosis incidence by 80% by 2030 from 2015 levels⁽¹⁾. Rapid diagnosis and treatment, key pillars of the End TB Strategy, can reduce tuberculosis transmission by limiting the time an individual is infectious and potentially transmitting *M. tuberculosis* (*Mtb*). Passive case detection, which depends on individuals with tuberculosis seeking care, is insufficient to rapidly reduce *Mtb* transmission in most settings⁽²⁾. In high-burden settings, WHO recommends systematic screening for tuberculosis disease in communities⁽³⁾. However, there is inconsistent evidence to indicate whether screening decreases tuberculosis prevalence^(4, 5).

Targeting screening in areas where most transmission occurs may decrease TB prevalence^(6, 7), but identifying these areas using routinely collected data is challenging. Among newly infected individuals, the incubation period is variable and the risk of progressing to symptomatic disease is generally low⁽⁸⁾. Areas of high disease burden may therefore reflect higher risk of progression to active disease rather than higher risk of transmission⁽⁹⁾. This phenomenon may be more pronounced in high human immunodeficiency virus (HIV) prevalence settings because people living with HIV are more likely to progress to active tuberculosis disease⁽¹⁰⁾ and less likely to transmit *Mtb* to others⁽¹¹⁾. Consequently, methods are needed that can identify areas of active transmission, which may not necessarily align with areas of high notification rates.

The increasing availability of whole genome sequencing (WGS) data, paired with methodological advances in transmission inference, has improved the ability to understand pathogen transmission dynamics^(8, 12, 13), information that is critical for the design of targeted active case finding efforts. Several studies have leveraged these types of data to characterize spatial patterns in tuberculosis burden and *Mtb* transmission⁽¹⁴⁻¹⁹⁾, but few have been conducted in cities with high rates of tuberculosis/HIV co-infection⁽²⁰⁾.

In this study, we collected and sequenced mycobacterial specimens from individuals diagnosed with culture-positive tuberculosis in Blantyre, Malawi between 2015 and 2019. We used WGS data to infer networks of transmission and paired those findings with geographical coordinates (GPS) of patient home locations to identify local transmission of specific strains and describe patterns of transmission between administrative areas. We hypothesized that whole genome sequencing would provide novel insights into *Mtb* transmission dynamics in a city with a high prevalence of both tuberculosis and HIV.

Methods

Study setting and population

Blantyre is a city in southern Malawi, with a population of approximately 800,000⁽²¹⁾. It is the second largest city in Malawi and is the nation's industrial and commercial capital. Over half of the population live in areas without access to basic municipal services and 43% of city land is considered unplanned or rural⁽²²⁾. Blantyre is a hilly city, and its varied topography creates distinct neighborhoods separated by ridges and valleys.

The study protocols were reviewed and approved by the University of Malawi College of Medicine Research and Ethics Committee (#P.12/18/2556), the London School of Hygiene and

Tropical Medicine (#16228-4), and Yale (#2000028431). Oral consent was provided by people registering for TB treatment for electronic data capture, including recording of household coordinates. Oral consent and assent were used for the latter since the electronic register data capture was conducted as part of normal clinical practice by District TB Officers. Approval was sought and gained from the District Health Office and local chiefs for the study to be conducted in their areas.

This study included people diagnosed with tuberculosis in Blantyre, Malawi between 1 January 2015 and 31 December 2019. All people with notified tuberculosis in Blantyre were registered in the ePAL (electronic Participant Locator) system⁽²³⁻²⁵⁾. ePAL is an app-based data entry platform for the collection of patient information combined with an electronic case report form with high resolution satellite maps and community-identified points of interest⁽²⁵⁾. Data available in ePAL include age, sex, diagnosing clinic, microbiology results (acid-fast bacillus [(AFB)] smear and Xpert MTB/RIF), symptom history and duration, tuberculosis classification (pulmonary, extra-pulmonary), HIV and antiretroviral therapy (ART) status, HIV clinic (if applicable), presence of known TB exposures (e.g. household contacts), locations of the three most recent clinics attended, number of hospital admissions within the year preceding diagnosis, and a range of poverty indicators. The patient's home location, selected via touch-screen and converted in-app to GPS coordinates⁽²⁵⁾, is also available.

We estimated tuberculosis notification rates at a resolution of 500m² grid cells. Population denominators were calculated from WorldPop⁽²¹⁾ 2020 data and aggregated from the original resolution of 100m². We also present grid-specific HIV prevalence estimates, which were derived from two national HIV prevalence studies, one Blantyre-specific prevalence study, and antenatal prevalence data. These data were combined in a Bayesian model to derived highly spatially resolved HIV prevalence estimates, as described previously⁽²⁶⁾.

Laboratory regrowth, DNA isolation, and whole genome sequencing

We set out to re-culture and sequence samples from all culture-positive cases notified over the five-year study period. Study isolates were thawed from -80°C and cultured in liquid Middlebrook 7H9 media using the BD BACTECT™ MGIT™ 960 system, then subcultured on Löwenstein–Jensen medium at 37°C to obtain pure colonies. A minimum of 1 µg *Mtb DNA* (either 100µl of 10000 ng/ml, or 40µl of 25000 ng/ml, etc) was manually extracted following standard operating procedures as previously described⁽²⁷⁾ and validated in-country, with DNA stored at -20°C before shipping for sequencing at TGen in Flagstaff, Arizona.

Sequencing libraries were be constructed using either Illumina's DNA Prep kit or Watchmaker's DNA Library Prep Kit with Fragmentation. Whole genome sequencing (WGS) was performed on an Illumina NextSeq550 or NextSeq1000 to produce paired-end, 150bp reads. A phiX (Illumina) sequencing control was spiked into each run at 1% of the total library to be sequenced to facilitate run performance monitoring. Raw sequencing FASTQ files were checked for non-*Mycobacterium* DNA using Kraken;⁽²⁸⁾ isolates containing > 80% *Mtb* reads were retained and non-*Mtb* reads filtered out using a custom script. Sequence reads were mapped to the H37Rv reference strain (GenBank accession number NC_000962.3) with BWA v.0.7.17 'mem'⁽²⁹⁾, removing sequences with < 80% mapping to the reference strain and < 50x average coverage.

Variant calling was carried out using GATK v.4.4.0.0 'HaplotypeCaller' and 'GenotypeGVCFs'⁽³⁰⁾. Single nucleotide polymorphisms (SNPs) were filtered to remove sites with low quality (Q < 20), low read depth (DP < 5), or high proportion missingness (missing call

in $\geq 10\%$ of isolates). Sites showing more than one allele (mixed sites) were assigned the majority allele where $\geq 90\%$ of reads agreed, otherwise these were assigned an ambiguous character 'N'. Finally, sequences with a high likelihood of mixed infection identified using MixInfect⁽³¹⁾ were removed. *In silico* lineage prediction and drug resistance profiling was carried out on the remaining isolates using TB-Profiler v.5.0.1⁽³²⁾.

Phylogenetic reconstruction

A multi-sequence alignment of concatenated SNPs was used to produce phylogenetic trees. SNPs in repetitive regions and known microbial resistance-associated and PE/PPE genes were removed to account for potential homoplasmy that may confound phylogenetic reconstruction (Supplemental Table 1). IQ-tree v.2.2.2.6⁽³³⁾ was used to construct a maximum-likelihood phylogeny of all isolates, with the '-m TEST' parameter used to determine the optimal nucleotide substitution model of Kimura's model with unequal base frequencies (K3Pu+F).

Transmission inference

Transmission networks and the probability of person-to-person transmission among sequenced cases was inferred using a two-step process. First, we identified preliminary, broad clusters of sequences using a 50 SNP threshold and constructed timed phylogenies with BEAST2 v.2.7.5⁽³⁴⁾. We used a relaxed lognormal substitution rate and ran the model for 2×10^8 Markov chain Monte Carlo (MCMC) iterations or until convergence was achieved and an adequate number of posterior samples were collected, demonstrated by the collected samples from all parameters reaching an effective sample size ESS of greater than 200 after a 20% burn-in was discarded.

Second, we ran TransPhylo⁽³⁵⁾ on each broad cluster to identify transmission networks. We used the implementation of TransPhylo with simultaneous inference of multiple trees⁽¹⁴⁾ to account for phylogenetic uncertainty by taking a random sample of 50 posterior trees from the BEAST2 output, discarding the first 20% as burn-in. We assumed a prior gamma generation time distribution ($\alpha = 1.3$, $\beta = 0.3$) and a prior gamma sampling time distribution ($\alpha = 1.1$, $\beta = 0.4$), as has been previously applied for *Mtb* transmission reconstruction⁽³⁵⁻³⁷⁾. We ran the model for 1×10^5 MCMC iterations using a fixed within-host coalescent parameter of 100/365 and a beta sampling proportion distribution ($\alpha = 2$, $\beta = 20$) that updated through the runs. This produced a predicted pairwise probability of direct transmission between isolates in clusters, and all isolates that were not present in the same broad cluster were *de facto* assigned a pairwise transmission probability of 0.

Transmission networks predicted by TransPhylo for each broad cluster with the highest probability were further refined. Where inferred networks were linked by more than three non-sampled hosts, we considered these as separate networks. This allowed us to identify putative transmission networks while accounting for missing cases.

Identification of spatial foci of transmission

For each putative transmission network with at least five cases, we investigated spatial areas where individuals with tuberculosis had relatively higher likelihoods of transmission to or from others in the area. We call these areas spatial foci of transmission, and we identify them using a non-parametric distance-based mapping (DBM)⁽³⁸⁾ approach implemented in the R package hotspots⁽³⁹⁾. Using hotspots, we divided the city into a 100 by 100 grid of cells (each cell is 194 m by 144 m). The analysis then proceeds for each transmission network. First, we select a transmission network to analyze. Second, we use the home location of tuberculosis cases that

were not part of that transmission network to calculate the expected number of cases within each grid cell. Third, we calculate the risk that there are more cases belonging to the transmission network in a given grid cell than expected, assigning a score between 0 (no spatial aggregation of individuals from the same transmission network) and 1 (highest risk of spatial aggregation). We repeated these steps for each transmission network with five or more individuals. Any grid cell with a score ≥ 0.95 was considered a spatial focus of transmission, and we used a narrow window (width = 0.01) to smooth results across grid cells⁽³⁹⁾. Note that this method allows that there may be multiple foci of transmission for a single transmission network.

Pairwise regression analysis of transmission networks

We used the R package GenePair⁽⁴⁰⁾ to quantify the association between the genetic relatedness of pairs of cases and several pair- and individual-level factors, including spatial proximity and shared healthcare clinics. Regression models with dyadic (paired) outcomes will produce overly optimistic effect estimates if the correlation between dyadic outcomes is not properly accounted for^(40, 41); GenePair solves this problem by including spatially-structured individual-level random effect parameters within a regression framework that induce correlation between the dependent variables⁽⁴⁰⁾.

We fit a single regression model with every individual in a transmission network, using a binary outcome indicator of whether two individuals were in the same network. We included predictors (age, sex, HIV status, HIV clinic, tuberculosis clinic, home address) if there were no missing values of that predictor among individuals in the transmission network, and if there were more than four pairs of individuals in each level of the predictor. We allowed the impact proximity on the relationship between cases to differ for cases that were close together ($< 6\text{km}$) and cases that were further apart ($\geq 6\text{km}$). We chose to model distance in this way because we believe the association of distance and cluster membership may be stronger at shorter distances. We performed model comparisons using AIC, and found that a 6km threshold outperformed a single linear effect and also other choices of threshold.

In a separate analysis, we fitted regression models for each transmission network with at least 10 cases, using the number of SNPs by which pairs differ as the outcome. This allows us to investigate the risk of specific predictors on SNP distance conditional on belonging to the same transmission network. We use the same set of covariates as above, and also include the zone in which the individuals resides. For all analyses, we based inference on samples from the joint posterior distribution, removing 10,000 iterations of burn-in and thinning the remaining 25,000 by a factor of 5 to reduce correlation in the posterior samples.

Transmission flow analysis

We divided the city into seven zones based on existing administrative boundaries and expert opinion about intracity mobility (Supplemental Figure 1). We used the R package PhyloFlows⁽⁴²⁾ to estimate transmission within and between these zones, accounting for the population size and sequencing coverage of each zone (Supplemental Table 2). Zone-level population sizes were calculated by aggregating across grid cells (see ‘Study setting and population’). We fitted PhyloFlows using a transmission probability threshold of 10% and assuming only one infector per individual. We ran the model for 20,000 MCMC iterations.

Results

Between January 2015 and December 2019, 11,137 tuberculosis disease episodes among individuals aged ≥ 15 were notified and recorded in ePAL. The mean tuberculosis notification rate over the study period was 279 per 100,000 population per year, which declined from 299 per 100,000 in 2015 to 224 per 100,000 in 2019. The notification rate varied across city zones, and grid cells with high notification rates abutted areas of apparent lower burden (Figure 1A). Among individuals notified in ePAL, most were new cases (12,046; 87%), most were male (7,438; 61%), the median age was 35 years (IQR: 26, 43), and most were living with HIV (7,860; 64%). Areas of higher HIV prevalence in the general populations in Zone 2 and Zone 6 overlapped with areas with high tuberculosis notification rates (Figure 1B). Population coverage of ART is estimated to be greater than 50%⁽⁴³⁾.

Whole genome sequencing analysis

Over the study period, 3,856 individuals tested positive by tuberculosis culture, 1,333 (34.6%) specimens were available for whole genome sequencing, and 1,009 (26.2%) samples could be matched to patient clinical data (Supplemental Figure 2). Among sequenced isolates, 861 (85%) passed quality control checks and 717 (83%) were found to be pure (non-mixed) samples, which were included in the final sample dataset. Home location data were available for 707 (99%) of these cases (Figure 1C).

A freezer failure, during which alarms were not acted on (due to COVID-19 lockdowns) occurred in 2020, resulting in lower-than-expected sequencing yield for affected frozen *Mtb* isolates. The fraction of individuals with a successfully sequenced diagnostic specimen did not differ meaningfully by sex, age, HIV status (Supplemental Table 3A), or city zone of residence (Supplemental Table 3B). However, the fraction of successfully sequenced culture positive cases varied by year, ranging from 12% to 33% (Supplemental Table 3C).

Most isolates included in the final sample dataset belonged to lineage 4 (Euro-American, 72%) followed by lineage 1 (Indo-Oceanic, 14%) (Figure 2), similar to previously reported studies in Blantyre⁽⁴⁴⁾. The proportion of sequences collected from each lineage remained consistent throughout the study period (Supplemental Figure 3). There was a low prevalence of drug resistance, with 94% (676/717) of samples susceptible to all antimicrobials (Figure 2). We identified 2 rifampin mono-resistant, 22 isoniazid mono-resistant strains, and 4 multidrug resistant strains ($< 1\%$).

Transmission inference

There were 393 isolates (56%) that belonged to one of 130 transmission networks inferred with TransPhylo and had an associated GPS coordinate. Most transmission networks contained only one pair of individuals (87/130; 67%); 13 transmission networks comprised five or more individuals (10%) and three transmission networks comprised 10 more individuals (2%) (Figure 3). The largest transmission network contained 25 individuals. There were 50 pairs of individuals with a high probability of direct transmission (probability ≥ 0.5) and a further 97 pairs with a moderate probability of direct transmission (probability ≥ 0.25).

Identification of spatial foci of transmission

In distance-based mapping (DBM) of 13 transmission networks with ≥ 5 individuals (Supplemental Table 4), we detected spatial foci of recent transmission in five transmission networks (Figure 4). For the largest transmission network (network 2, $n = 25$), we did not detect any spatial aggregation of cases, suggesting that it was widespread throughout the city.

Transmission network #47 was the largest network ($n = 17$) for which we identified a spatial focus. Most individuals in this network were people living with HIV ($n = 10$, 67%) and the single spatial focus of transmission was in a high HIV prevalence area. Network #71 was the second largest transmission network, and had a spatial focus on the periphery of the city where there was a high case notification rate. Only 20% ($n = 3$) of individuals within network #71 were people living with HIV.

Not all members of a transmission cluster with a spatial focus of transmission lived in or near the focus we identified (Supplemental Table 4). In network #80, 75% of network members ($n = 6$) lived inside the focus, and, in network #1, 60% of network members ($n = 3$) lived within 1km of the focus boundary (Figure 3A). Overall, of the 393 individuals with TB belonging to a transmission network, 252 (64%) lived within 1 km of a transmission focus.

Pairwise regression analysis of transmission networks

We analyzed all individuals belonging to a putative transmission network (excluding individuals with missing home location data; $n = 389$), and found that two individuals had a higher odds of belonging to the same putative transmission network if they shared a tuberculosis diagnostic clinic (adjusted odds ratio (aOR) 1.61 95% highest posterior density interval (HPD) 1.14, 2.13), were both HIV negative (aOR 1.51, 95% HPD 1.09, 2.02), and were both male (aOR 1.45, 95% HPD 1.09, 1.91) (Figure 5A). A 1 km increase in the distance between home location was associated with a 0.84 (0.79, 0.88) aOR of belonging to the same cluster for distances up to 6km. Each 1 km increase in distance above the 6 km threshold was associated with a 0.91 (0.89, 0.94) aOR. We did not find a significant association between shared HIV clinic and belonging to the same putative transmission cluster (aOR 0.86, 95% HPD 0.58, 1.21). Because there was very high coverage of antiretroviral therapy among HIV positive individuals in this study (91%), we did not include ART as a covariate in our model.

We also analyzed the three largest transmission networks (#2, #43, and #71) in separate analyses (Figure 5B). In network #2, increasing distance between participant homes was associated with a small but significant increase in the number of SNPs by which the sequences differed (adjusted Relative Risk (aRR) per 2 km increase of 1.13, 95% HPD 1.03, 1.23). In addition, individuals in transmission network #2 who were diagnosed at the same tuberculosis clinic had sequences with smaller SNP differences on average (aRR 0.57, 95% HPD 0.28, 0.92). We did not identify statistically significant effects of any covariates on SNP differences in networks #43 and #71.

Transmission flows analysis

We estimated between and within zone rates of transmission based on transmission pairs identified using TransPhylo (Figure 6, Supplemental Table 5). We found that 68.2% (95% credible interval (CrI): 62.7%, 73.1%) of transmission within the city occurred between, rather than within, zones. Zone 7 in the south-west of the city had the highest percentage of within-zone transmission, followed by Zone 4 in the center of the city. The highest rates of between zone transmission were between Zone 1 and Zone 7 and between Zone 4 and Zone 7. We estimate that 64.5% (95% CrI: 58.9%, 69.9%) of all tuberculosis transmission in Blantyre could be attributed to infectious individuals from Zone 1, Zone 4, or Zone 7. Finally, there was effectively no transmission flow through Zone 3 (posterior mean 1.7×10^{-5}), a mostly industrial zone in the center of the city with a smaller residential population relative to other zones.

Discussion

This analysis presents novel insights into transmission patterns of *M. tuberculosis* in Blantyre, Malawi between 2015 and 2019, a period when treatment coverage for HIV was being rapidly scaled up and TB notifications were falling. In phylogenetic analyses, 67% of sequenced isolates could be mapped to a putative transmission network. We identified two main characteristics of transmission with implications for the design of public health interventions. First, despite evidence that Blantyre's tuberculosis epidemic is receding, most detected transmission events occurred between zones, suggesting that disease control efforts must include improved access to tuberculosis services through primary clinics and action to address the social determinants of tuberculosis. Second, we identified a sizeable minority of transmission events that occurred in local outbreaks in distinct areas of the city. Therefore, we anticipate that adding targeted active case finding in these areas is likely to improve tuberculosis control, at least in these neighborhoods.

Geographical location of residence has long been known to be linked to tuberculosis epidemiology, with poor-quality housing, air pollution, undernutrition, crowding, and suboptimal access to healthcare as key drivers of incidence. While the residential proximity of two individuals is an important predictor of membership in the same transmission network (Figure 5A), we found that the majority of transmission events occurred between individuals residing in different zones of the city (Figure 6). High rates of inter-zone transmission are consistent with findings in other high TB/HIV burden settings⁽¹⁸⁾, and indicate that high levels of population mixing play a role in sustaining Blantyre's tuberculosis epidemic⁽⁴⁵⁾. These findings illustrate the continuing need for broad based accessible TB services for all individuals.

These findings also highlight the role that increasingly-available whole genome sequencing data can play in identifying areas of ongoing transmission. Reactive interventions to interrupt transmission may be effective at reducing incidence in these areas, and previous models have suggested that interventions targeted to a single transmission focus could have much broader benefits to reducing incidence in entire city⁽⁶⁾. However, the impact of targeted interventions on city-wide tuberculosis incidence depends on the level of connectedness of localized epidemics. In this study, five of the thirteen largest transmission networks were associated with a spatial focus of transmission, though none of the foci overlapped or abutted each other (Figure 4). This finding is distinct from patterns reported from other genomic analyses of tuberculosis in urban settings^(20, 39), and raises questions about the impact of targeted interventions in locations with complex multifocal epidemics.

An important limitation of this study is the relatively low fraction of sequencing (18% of culture-positive isolates over the study period). Diagnostic isolates were available for approximately one-third of eligible cases, and lab contamination further reduced the number of sequenced isolates. This poses a challenge for inferring transmission networks, as most existing methods have been designed for more densely sampled outbreaks. We accounted for this by allowing for a higher number of unsampled intermediate cases in transmission networks, and this introduces additional uncertainty which may have impacted the results of our transmission inference.

In this study, we used genomic and spatial analyses to describe the transmission dynamics of *Mtb* in a high tuberculosis and HIV prevalence city. Our findings reveal the importance of both local and longer range transmission as drivers of tuberculosis transmission in this city. Further modeling is needed to estimate whether targeted screening in areas of focal transmission can be an impactful addition to broader case finding efforts in this city.

Data availability

The genomic data used in this study will be made available in GenBank upon publication. Additional data used in the analysis (with the exception of patient location and clinic data), will be provided as a file in Supporting Information.

Ethical Approval

The study protocols were reviewed and approved by the University of Malawi College of Medicine Research and Ethics Committee (#P.12/18/2556), the London School of Hygiene and Tropical Medicine (#16228-4), and Yale (#2000028431).

Citations

1. Implementing the end TB strategy: the essentials, 2022 update. Geneva: World Health Organization; 2022. License: CC BY-NC-SA 3.0 IGO.
2. Yuen CM, Amanullah F, Dharmadhikari A, Nardell EA, Seddon JA, Vasilyeva I, et al. Turning off the tap: stopping tuberculosis transmission through active case-finding and prompt effective treatment. *The Lancet*. 2015;386(10010):2334-43.
3. WHO operational handbook on tuberculosis. Module 2: screening – systematic screening for tuberculosis disease. Geneva: World Health Organization; 2021. License: CC BY-NC-SA 3.0 IGO.
4. Burke RM, Nliwasa M, Feasey HRA, Chaisson LH, Golub JE, Naufal F, et al. Community-based active case-finding interventions for tuberculosis: a systematic review. *The Lancet Public Health*. 2021;6(5):e283-e99.
5. Feasey HRA, Khundi M, Soko RN, Bottomley C, Chiume L, Burchett HED, et al. Impact of active case-finding for tuberculosis on case-notifications in Blantyre, Malawi: A community-based cluster-randomised trial (SCALE). *PLOS Global Public Health*. 2023;3(12):e0002683.
6. Dowdy DW, Golub JE, Chaisson RE, Saraceni V. Heterogeneity in tuberculosis transmission and the role of geographic hotspots in propagating epidemics. *Proceedings of the National Academy of Sciences*. 2012;109(24):9557-62.
7. Khundi M, Carpenter JR, Nliwasa M, Cohen T, Corbett EL, MacPherson P. Effectiveness of spatially targeted interventions for control of HIV, tuberculosis, leprosy and malaria: a systematic review. *BMJ Open*. 2021;11(7):e044715.
8. Mathema B, Andrews JR, Cohen T, Borgdorff MW, Behr M, Glynn JR, et al. Drivers of Tuberculosis Transmission. *The Journal of Infectious Diseases*. 2017;216(suppl_6):S644-S53.
9. Cudahy PGT, Andrews JR, Bilinski A, Dowdy DW, Mathema B, Menzies NA, et al. Spatially targeted screening to reduce tuberculosis transmission in high-incidence settings. *Lancet Infect Dis*. 2019;19(3):e89-e95.
10. Pawlowski A, Jansson M, Sköld M, Rottenberg ME, Källenius G. Tuberculosis and HIV co-infection. *PLoS Pathog*. 2012;8(2):e1002464.
11. Martinez L, Woldu H, Chen C, Hallowell BD, Castellanos ME, Lu P, et al. Transmission Dynamics in Tuberculosis Patients With Human Immunodeficiency Virus: A Systematic Review and Meta-analysis of 32 Observational Studies. *Clin Infect Dis*. 2021;73(9):e3446-e55.
12. Lemey P, Rambaut A, Welch JJ, Suchard MA. Phylogeography takes a relaxed random walk in continuous space and time. *Mol Biol Evol*. 2010;27(8):1877-85.
13. Pybus OG, Suchard MA, Lemey P, Bernardin FJ, Rambaut A, Crawford FW, et al. Unifying the spatial epidemiology and molecular evolution of emerging epidemics. *Proc Natl Acad Sci U S A*. 2012;109(37):15066-71.
14. Xu Y, Cancino-Muñoz I, Torres-Puente M, Villamayor LM, Borrás R, Borrás-Máñez M, et al. High-resolution mapping of tuberculosis transmission: Whole genome sequencing and phylogenetic modelling of a cohort from Valencia Region, Spain. *PLOS Medicine*. 2019;16(10):e1002961.

15. Yang C, Sobkowiak B, Naidu V, Codreanu A, Ciobanu N, Gunasekera KS, et al. Phylogeography and transmission of *M. tuberculosis* in Moldova: A prospective genomic analysis. *PLOS Medicine*. 2022;19(2):e1003933.
16. Yang C, Lu L, Warren JL, Wu J, Jiang Q, Zuo T, et al. Internal migration and transmission dynamics of tuberculosis in Shanghai, China: an epidemiological, spatial, genomic analysis. *The Lancet Infectious Diseases*. 2018;18(7):788-95.
17. Huang C-C, Trevisi L, Becerra MC, Calderón RI, Contreras CC, Jimenez J, et al. Spatial scale of tuberculosis transmission in Lima, Peru. *Proceedings of the National Academy of Sciences*. 2022;119(45):e2207022119.
18. Nelson KN, Shah NS, Mathema B, Ismail N, Brust JCM, Brown TS, et al. Spatial Patterns of Extensively Drug-Resistant Tuberculosis Transmission in KwaZulu-Natal, South Africa. *The Journal of Infectious Diseases*. 2018;218(12):1964-73.
19. Sobkowiak B, Banda L, Mzembe T, Crampin AC, Glynn JR, Clark TG. Bayesian reconstruction of *Mycobacterium tuberculosis* transmission networks in a high incidence area over two decades in Malawi reveals associated risk factors and genomic variants. *Microb Genom*. 2020;6(4).
20. Baker CR, Barilar I, de Araujo LS, Rimoin AW, Parker DM, Boyd R, et al. Use of High-Resolution Geospatial and Genomic Data to Characterize Recent Tuberculosis Transmission, Botswana. *Emerg Infect Dis*. 2023;29(5):977-87.
21. WorldPop. Malawi 100m Population, Version 2.: University of Southampton; 2017 [Available from: <https://hub.worldpop.org/doi/10.5258/SOTON/WP00538>].
22. Malawi: Blantyre Urban Profile. UN-HABITAT. [Available from: <https://unhabitat.org/sites/default/files/download-manager-files/Malawi%20Blantyre%20Urban%20Profile.pdf>]; 2011.
23. MacPherson P, Choko AT, Webb EL, Thindwa D, Squire SB, Sambakunsi R, et al. Development and Validation of a Global Positioning System–based “Map Book” System for Categorizing Cluster Residency Status of Community Members Living in High-Density Urban Slums in Blantyre, Malawi. *American Journal of Epidemiology*. 2013;177(10):1143-7.
24. MacPherson P, Khundi M, Nliwasa M, Choko AT, Phiri VK, Webb EL, et al. Disparities in access to diagnosis and care in Blantyre, Malawi, identified through enhanced tuberculosis surveillance and spatial analysis. *BMC Medicine*. 2019;17(1):21.
25. Harris R. Informing development strategies for new tuberculosis vaccines: mathematical modelling and novel epidemiological tools: London School of Hygiene & Tropical Medicine; 2018.
26. Burke R, Khundi M, Lam K, Feasey H, Kawalazira G, Choko A, et al. Fine-resolution estimates of HIV prevalence in Blantyre, Malawi: a Bayesian modelling analysis of survey, health facility, and household testing data. *AIDS* 2022; Montreal, Canada. <https://programme.aids2022.org/Abstract/Abstract/?abstractid=16662022>.
27. Somerville W, Thibert L, Schwartzman K, Behr MA. Extraction of *Mycobacterium tuberculosis* DNA: a question of containment. *J Clin Microbiol*. 2005;43(6):2996-7.
28. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*. 2014;15(3):R46.
29. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14):1754-60.

30. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20(9):1297-303.
31. Sobkowiak B, Glynn JR, Houben RMGJ, Mallard K, Phelan JE, Guerra-Assunção JA, et al. Identifying mixed *Mycobacterium tuberculosis* infections from whole genome sequence data. *BMC Genomics.* 2018;19(1):613.
32. Phelan JE, O'Sullivan DM, Machado D, Ramos J, Oppong YEA, Campino S, et al. Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. *Genome Medicine.* 2019;11(1):41.
33. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution.* 2020;37(5):1530-4.
34. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology.* 2019;15(4):e1006650.
35. Didelot X, Fraser C, Gardy J, Colijn C. Genomic Infectious Disease Epidemiology in Partially Sampled and Ongoing Outbreaks. *Molecular Biology and Evolution.* 2017;34(4):997-1007.
36. Sobkowiak B, Romanowski K, Sekirov I, Gardy JL, Johnston JC. Comparing *Mycobacterium tuberculosis* transmission reconstruction models from whole genome sequence data. *Epidemiol Infect.* 2023;151:e105.
37. Ayabina D, Ronning JO, Alfsnes K, Debech N, Brynildsrud OB, Arnesen T, et al. Genome-based transmission modelling separates imported tuberculosis from recent transmission within an immigrant population. *Microb Genom.* 2018;4(10).
38. Jeffery C, Ozonoff A, White LF, Pagano M. Distance-Based Mapping of Disease Risk. *The International Journal of Biostatistics.* 2013;9(2):265-90.
39. Zelner JL, Murray MB, Becerra MC, Galea J, Lecca L, Calderon R, et al. Identifying Hotspots of Multidrug-Resistant Tuberculosis Transmission Using Spatial and Molecular Genetic Data. *The Journal of Infectious Diseases.* 2015;213(2):287-94.
40. Warren JL, Chitwood MH, Sobkowiak B, Colijn C, Cohen T. Spatial modeling of *Mycobacterium tuberculosis* transmission with dyadic genetic relatedness data. *Biometrics.* 2023;n/a(n/a).
41. Council NR. Dynamic social network modeling and analysis: Workshop summary and papers: National Academies Press; 2003.
42. Xi X. PhyloFlows 2019 [updated 2022. Available from: <https://github.com/BDI-pathogens/phyloscanner/tree/master/phyloflows>.
43. Kanyerere H, Girma B, Mpunga J, Tayler-Smith K, Harries AD, Jahn A, et al. Scale-up of ART in Malawi has reduced case notification rates in HIV-positive and HIV-negative tuberculosis. *Public Health Action.* 2016;6(4):247-51.
44. Ndhlovu V, Kiran A, Sloan D, Mandala W, Kontogianni K, Kamdolozi M, et al. Genetic diversity of *Mycobacterium tuberculosis* clinical isolates in Blantyre, Malawi. *Heliyon.* 2019;5(10):e02638.
45. Brown TS, Robinson DA, Buckee CO, Mathema B. Connecting the dots: understanding how human mobility shapes TB epidemics. *Trends Microbiol.* 2022;30(11):1036-44.

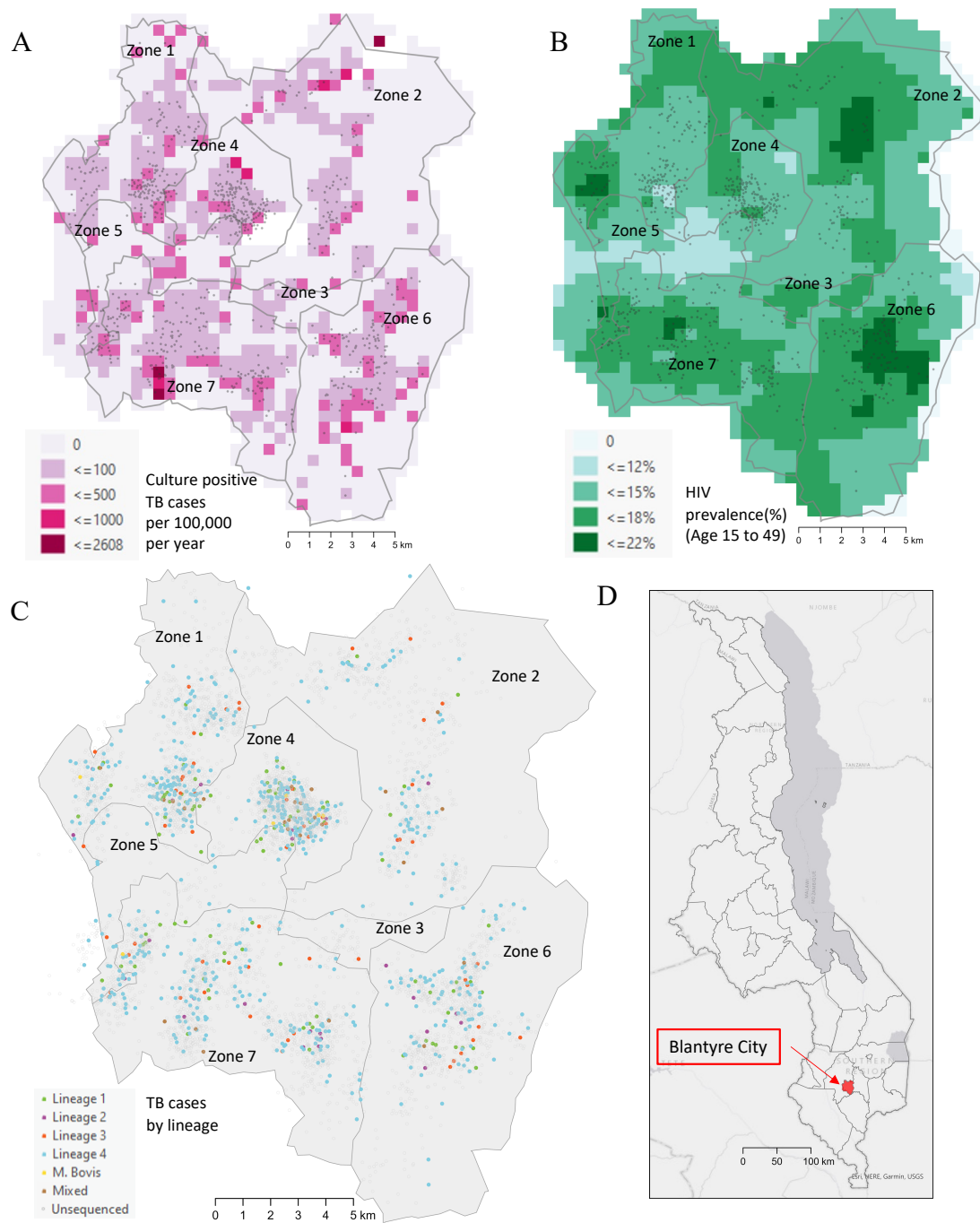


Figure 1 Map of Blantyre, Malawi with (A) culture positive tuberculosis case notifications per 100,000 population per year over the study period; (B) HIV prevalence estimates (population aged 15-49 years) by zone in the urban area of Blantyre; (C) all notified tuberculosis cases, colored by major lineage or shown in grey if no sequencing data were available; (D) inset map indicating the location of Blantyre in Malawi. Points have been jittered for privacy.

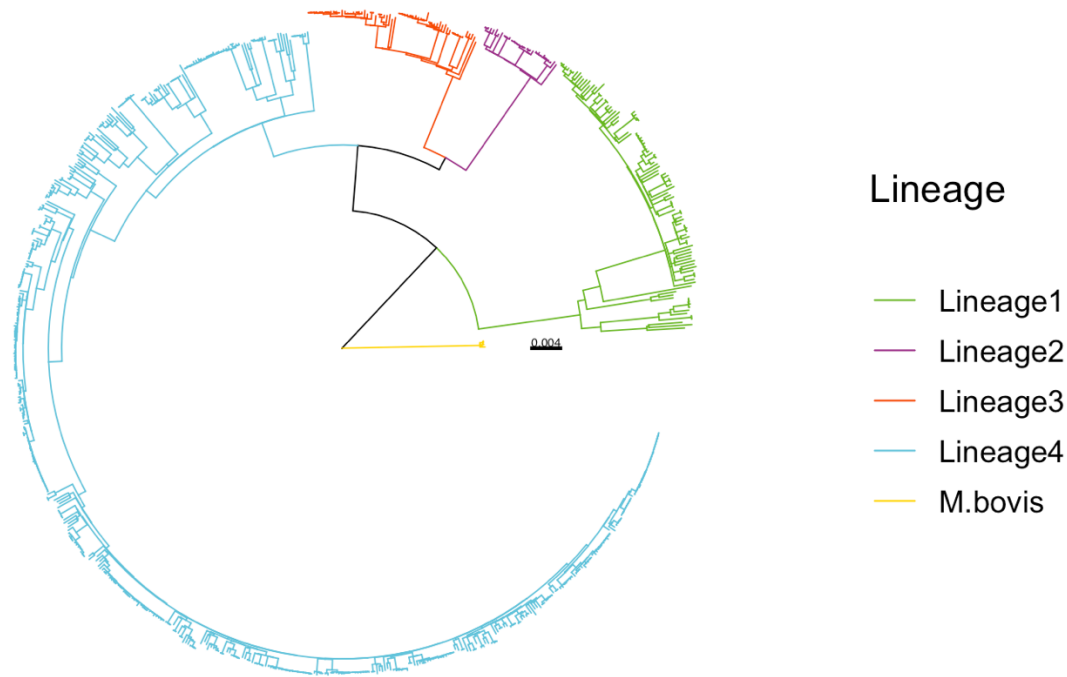


Figure 2 A maximum likelihood phylogeny illustrating the genetic relatedness of the 717 isolates included in the study, colored by major *Mycobacterium tuberculosis* complex lineage. Branch lengths are scaled by substitutions per site.

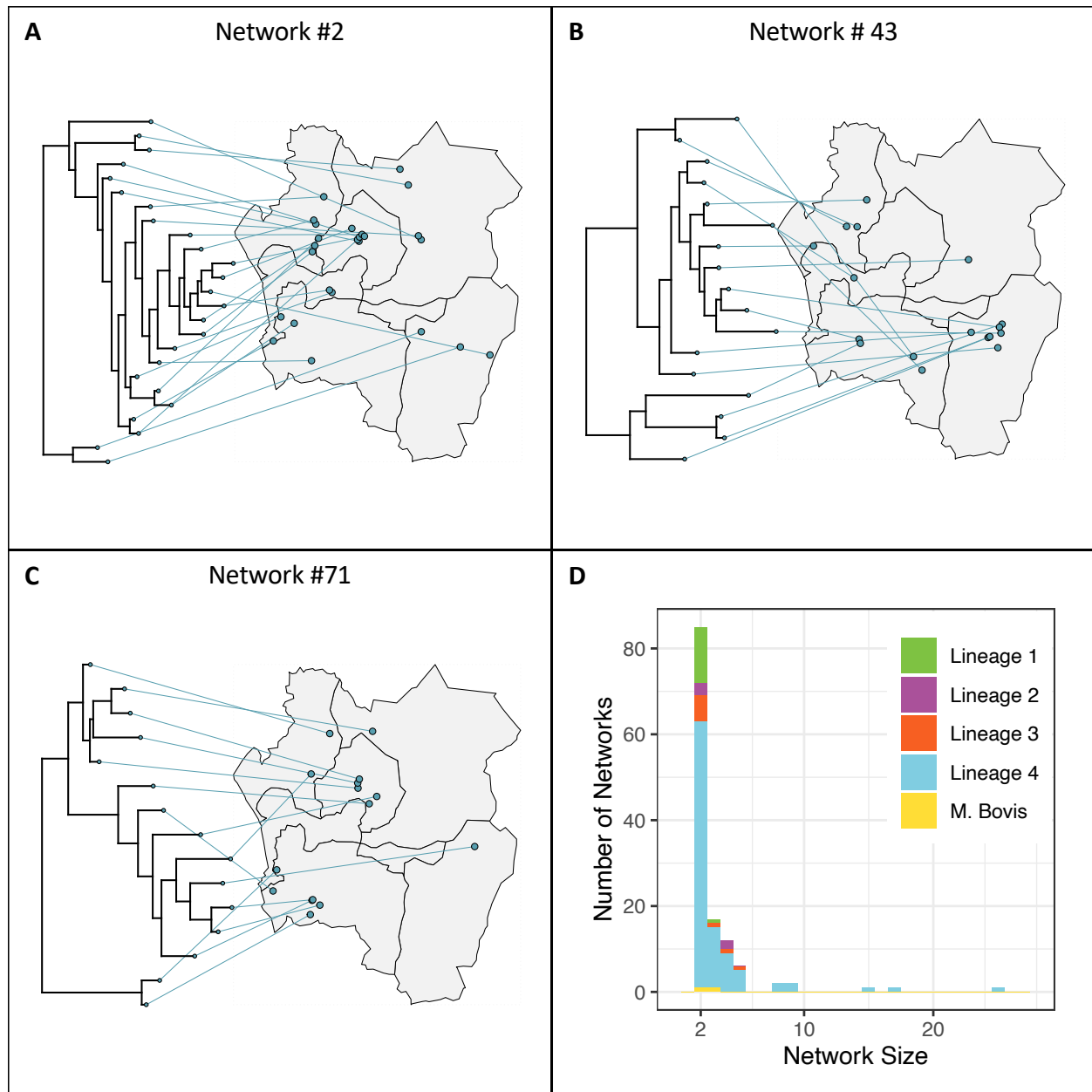


Figure 3 (A-C) Time-resolved phylogenies with taxa linked to case home location, colored by main lineage. Each panel contains a different transmission network. The three networks with 10 or more individuals are shown. Points have been jittered for privacy. **(D)** Distribution of transmission network sizes (number of sampled hosts within the network), colored by main lineage.

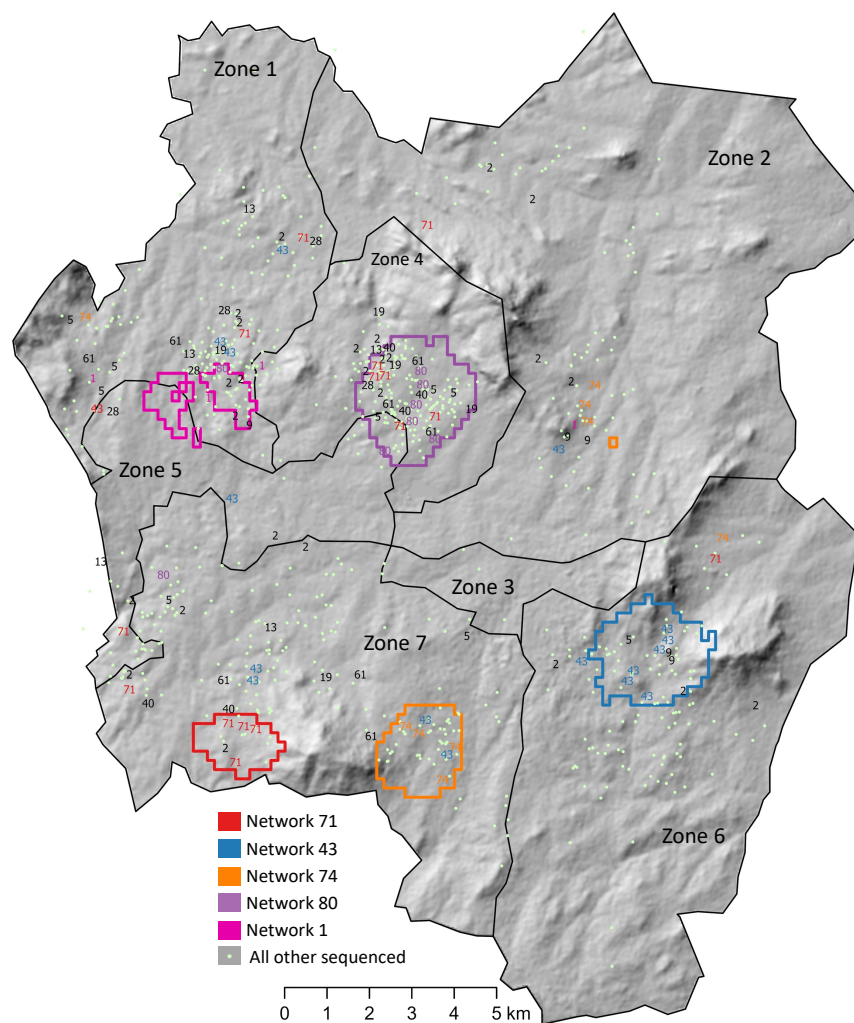


Figure 4 Foci of recent transmission identified using a distance based mapping (DBM) approach based on home address, applied to transmission networks with 5 or more individuals. Circled regions represent areas where the risk of spatial aggregation for a network is greater than 95% (spatial foci). A single transmission network may have multiple spatial foci. Individuals belonging to a transmission network with a spatial focus may reside outside the focus. Points have been jittered for privacy. Base map citation: “Malawi SRTM DEM 30meters.” n.d. Accessed April 24, 2024. <https://www.africageoportal.com/datasets/rcmr::malawi-srtm-dem-30meters/about>.

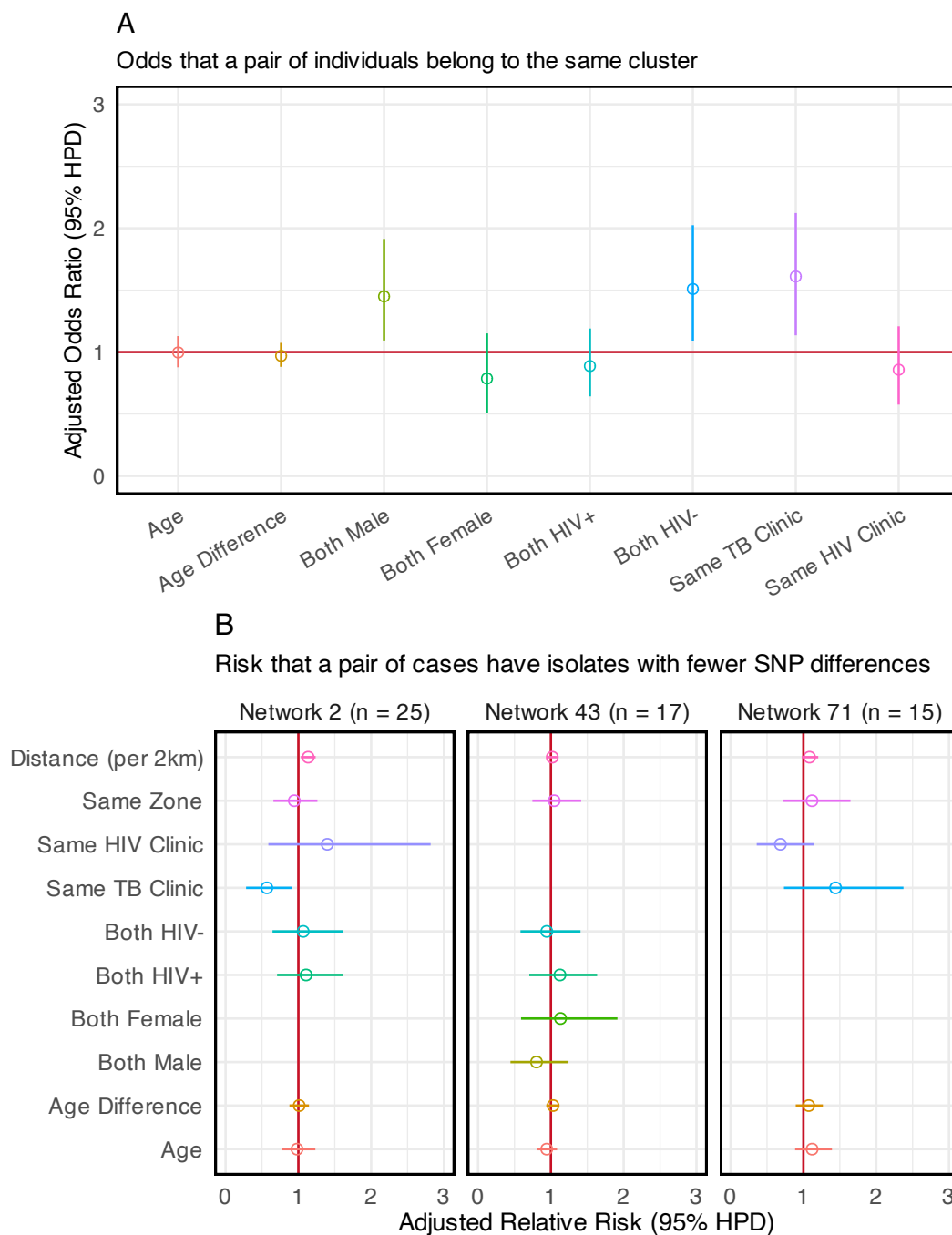
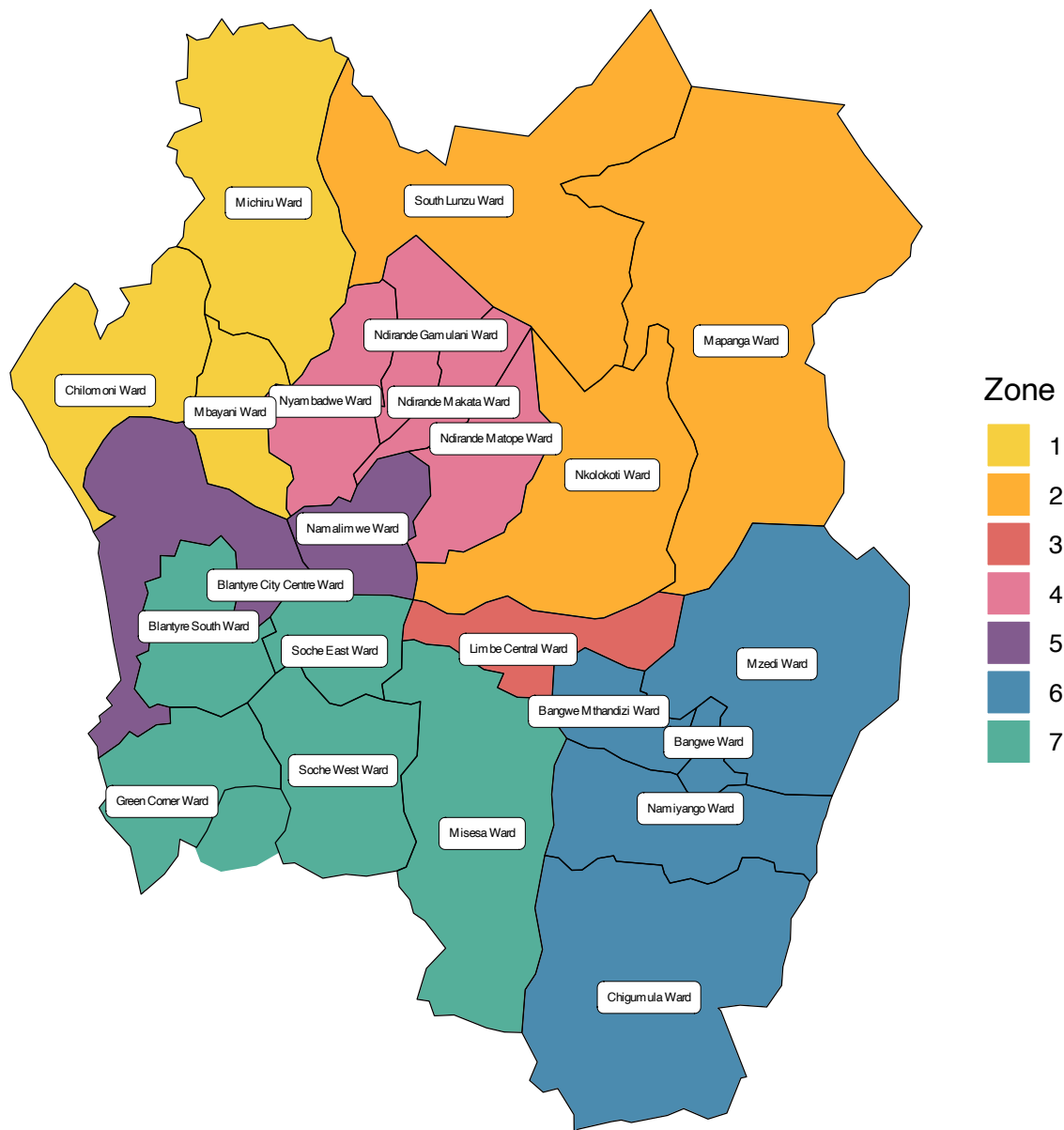


Figure 5 (A) Effect of covariates on the odds that case pairs belong to the same putative transmission cluster. Adjusted Odds Ratio > 1 is associated with increased odds of belonging to the same putative transmission cluster. **(B)** Effect estimates of covariates on SNP difference in isolates from case pairs within the same transmission network. Adjusted Relative Risk < 1 is associated with smaller SNP differences on average; smaller SNP differences are associated with increased likelihood of direct transmission between case pairs. Gender is excluded in the analyses of network 2 and network 71 because only three individuals were female. HIV status is excluded in the analysis of network 71 because only three individuals were people living with HIV.



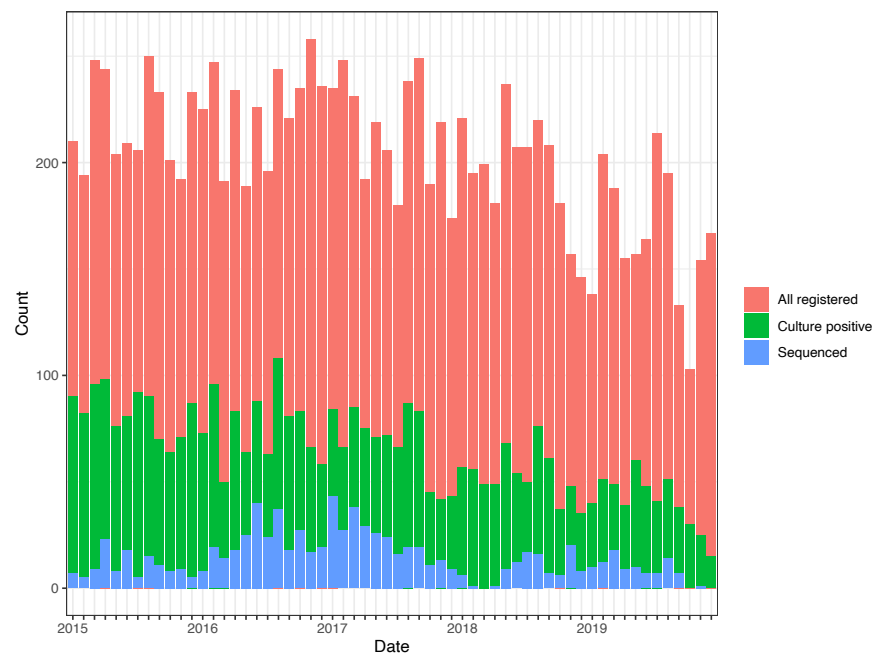
Figure 6 Estimates of the percent of overall transmission occurring within- and between-zones of Blantyre. Arrows point from infector zone to infected zone, and cyclic arrows represent within-zone transmission. Arrows are omitted if there was no evidence of within-zone transmission (e.g. Zone 3) or if there was no evidence of between-zone transmission (e.g. Zone 5 to Zone 3).

Supplemental Figure 1: Zones of Blantyre with ward boundaries. Zone 7 has been expanded to include a small area outside the city boundary with a high TB notification rate (unlabeled ward between Green Corner Ward and Soche West Ward).

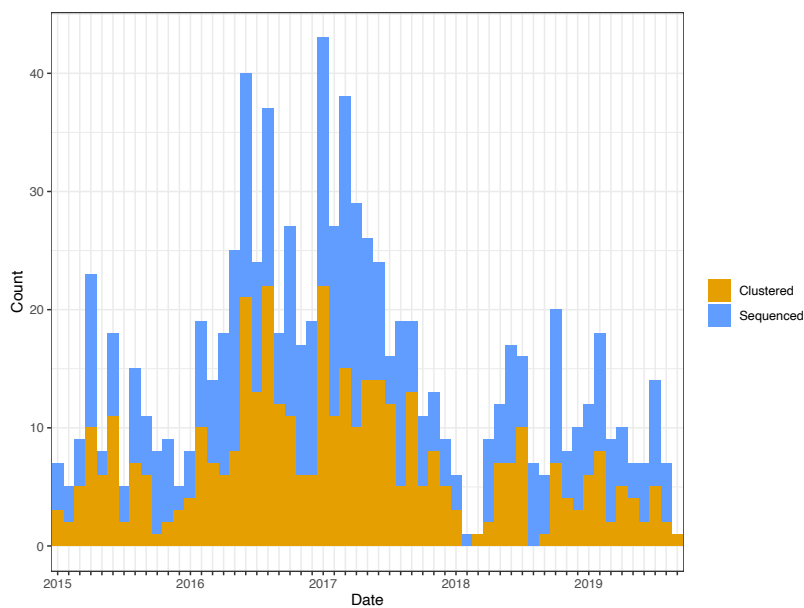


Supplemental Figure 2: (A) Fraction of cases that were culture-positive and sequenced over time and (B) fraction of sequenced cases belonging to a transmission network (“clustered”) over time.

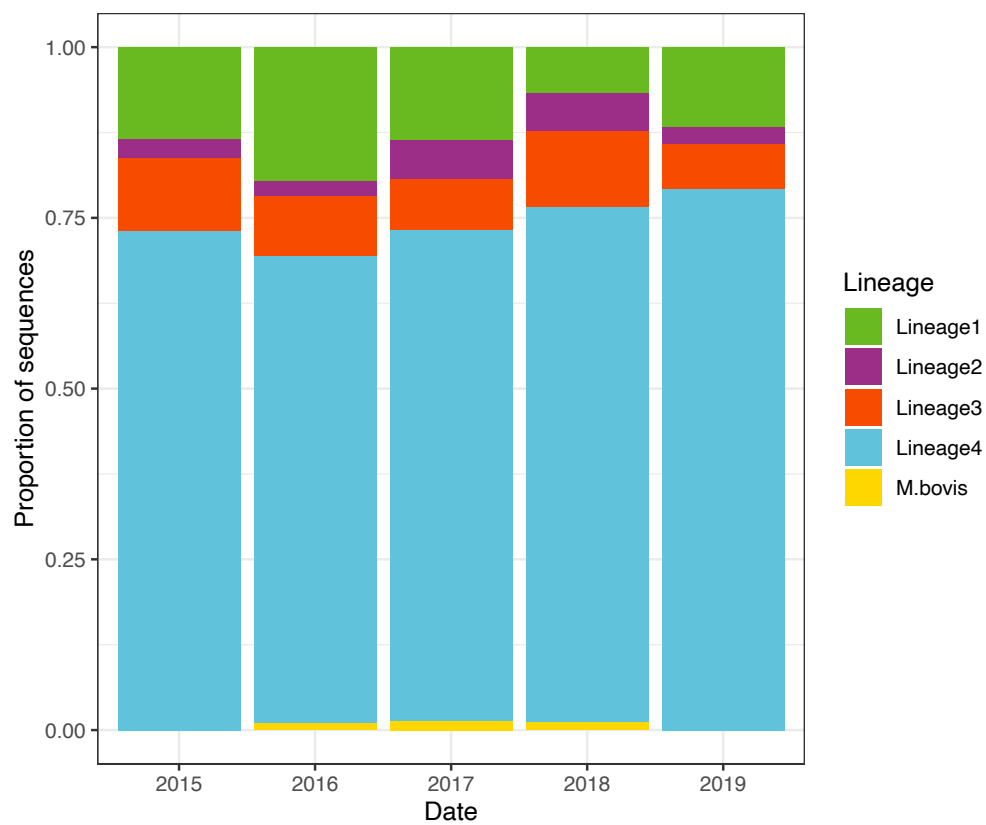
A



B



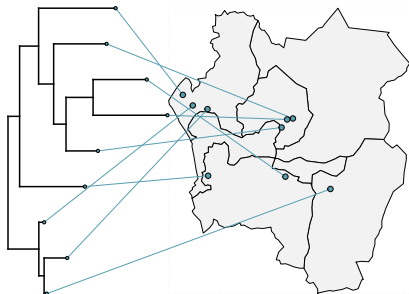
Supplemental Figure 3: Proportion of sequences belonging to the four major MTBC lineages and *M. bovis*.



Supplemental Figure 4: Time-resolved transmission networks with 5-10 sampled cases.

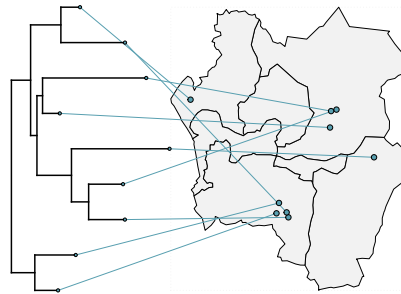
Network 5

Lineage 4, n = 9



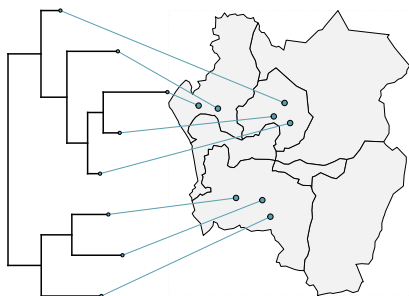
Network 74

Lineage 4, n = 9



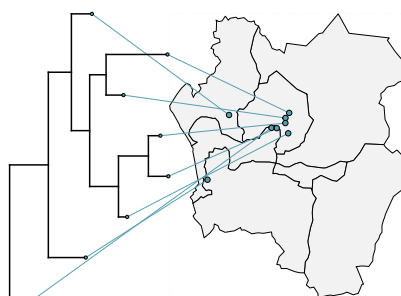
Network 61

Lineage 4, n = 8



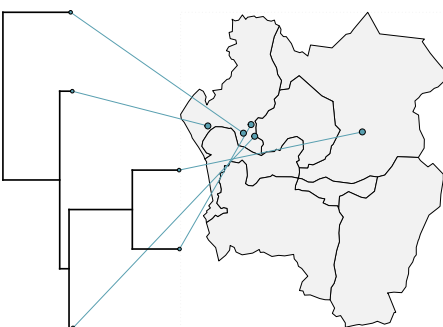
Network 80

Lineage 4, n = 8



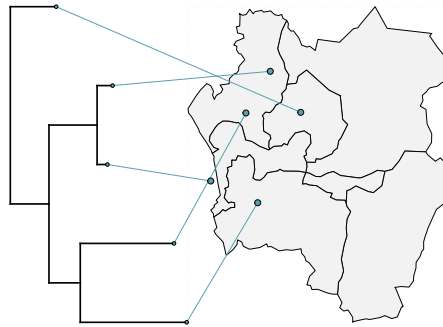
Network 1

Lineage 4, n = 5



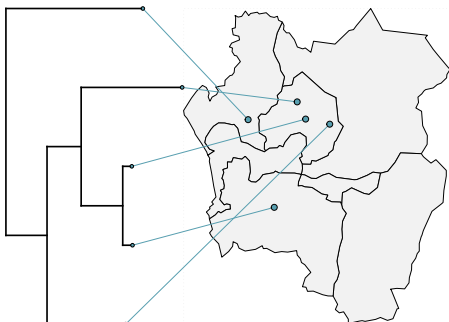
Network 13

Lineage 4, n = 5



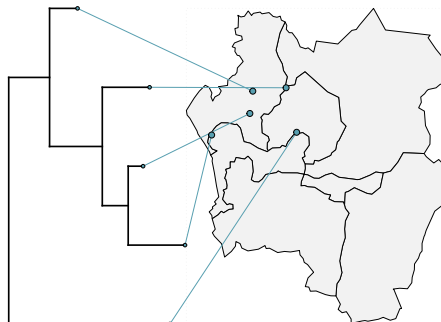
Network 19

Lineage 4, n = 5



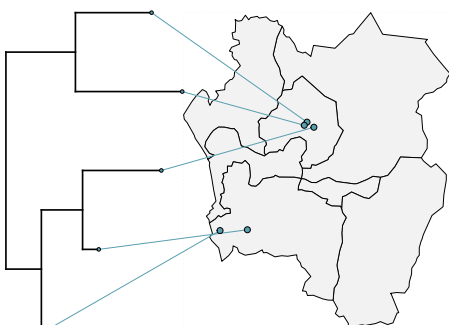
Network 28

Lineage 4, n = 5



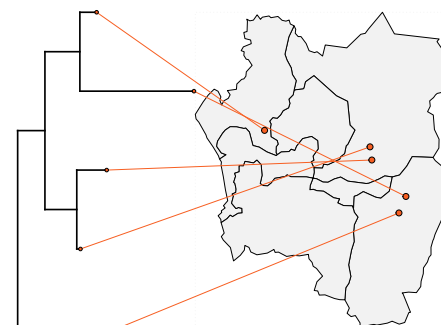
Network 40

Lineage 4, n = 5



Network 9

Lineage 3, n = 5



Supplemental Table 1: Sites that were excluded from the multi-sequence alignment.

START	STOP	CDS	Name	LocusTag
5240	7267	Drug_resistance	gyrB	Rv0005
7302	9818	Drug_resistance	gyrA	Rv0006
156578	157600	Drug_resistance	fbpC	Rv0129c
408634	409173	Drug_resistance	.	Rv0340
409362	410801	Drug_resistance	iniB	Rv0341
410838	412760	Drug_resistance	iniA	Rv0342
412757	414238	Drug_resistance	iniC	Rv0343
759807	763325	Drug_resistance	rpoB	Rv0667
781560	781934	Drug_resistance	rpsL	Rv0682
1416181	1417347	Drug_resistance	embR	Rv1267c
1471846	1473382	Drug_resistance	rrs	rrs
1673440	1674183	Drug_resistance	fabG1	Rv1483
1674202	1675011	Drug_resistance	inhA	Rv1484
1792400	1793740	Drug_resistance	.	Rv1592c
1917940	1918746	Drug_resistance	tlyA	Rv1694
2006636	2006947	Drug_resistance	.	Rv1772
2101651	2103042	Drug_resistance	ndh	Rv1854c
2153889	2156111	Drug_resistance	katG	Rv1908c
2156149	2156592	Drug_resistance	furA	Rv1909c
2288681	2289241	Drug_resistance	pncA	Rv2043c
2515304	2516548	Drug_resistance	srmR	Rv2242
2516787	2517695	Drug_resistance	fabD	Rv2243
2518115	2519365	Drug_resistance	kasA	Rv2245
2520743	2522164	Drug_resistance	accD6	Rv2247
2725571	2726087	Drug_resistance	oxyR	Rv2427a
2726193	2726780	Drug_resistance	ahpC	Rv2428
3073680	3074471	Drug_resistance	thyA	Rv2764c
3153039	3154631	Drug_resistance	efpA	Rv2846c
3489506	3490375	Drug_resistance	moaR1	Rv3124
3490476	3491651	Drug_resistance	PPE49	Rv3125c
3491808	3492122	Drug_resistance	.	Rv3126c
3505363	3506769	Drug_resistance	fadE24	Rv3139
3644898	3645977	Drug_resistance	manB	Rv3264c
3646895	3647809	Drug_resistance	rmlD	Rv3266c
4007331	4008182	Drug_resistance	nhoA	Rv3566c
4239863	4243147	Drug_resistance	embC	Rv3793
4243233	4246517	Drug_resistance	embA	Rv3794
4246514	4249810	Drug_resistance	embB	Rv3795
4264563	4265462	Drug_resistance	fbpD	Rv3803c
4326004	4327473	Drug_resistance	ethA	Rv3854c
4407528	4408202	Drug_resistance	gid	Rv3919c
105324	106715	PE/PPE	PPE1	Rv0096
131382	132872	PE/PPE	PE_PGRS1	Rv0109
149533	150996	PE/PPE	PE_PGRS2	Rv0124
177543	179309	PE/PPE	PE1	Rv0151c
179319	180896	PE/PPE	PE2	Rv0152c
187433	188839	PE/PPE	PE3	Rv0159c
188931	190439	PE/PPE	PE4	Rv0160c
307877	309547	PE/PPE	PPE2	Rv0256c
333437	336310	PE/PPE	PE_PGRS3	Rv0278c
336560	339073	PE/PPE	PE_PGRS4	Rv0279c
339364	340974	PE/PPE	PPE3	Rv0280
349624	349932	PE/PPE	PE5	Rv0285
349935	351476	PE/PPE	PPE4	Rv0286
351525	351818	PE/PPE	esxG	Rv0287
361334	363109	PE/PPE	PE_PGRS5	Rv0297
366150	372764	PE/PPE	PPE5	Rv0304c
372820	375711	PE/PPE	PPE6	Rv0305c
399535	400050	PE/PPE	PE6	Rv0335c
424269	424694	PE/PPE	PPE7	Rv0354c
424777	434679	PE/PPE	PPE8	Rv0355c
467459	468001	PE/PPE	PPE9	Rv0388c

530751	532214	PE/PPE	PPE10	Rv0442c
543174	544730	PE/PPE	PPE11	Rv0453
622793	624577	PE/PPE	PE_PGRS6	Rv0532
671996	675916	PE/PPE	PE_PGRS7	Rv0578c
832981	833508	PE/PPE	PE_PGRS8	Rv0742
835701	838052	PE/PPE	PE_PGRS9	Rv0746
838451	840856	PE/PPE	PE_PGRS10	Rv0747
846159	847913	PE/PPE	PE_PGRS11	Rv0754
848103	850040	PE/PPE	PPE12	Rv0755c
924951	925364	PE/PPE	PE_PGRS12	Rv0832
925361	927610	PE/PPE	PE_PGRS13	Rv0833
927837	930485	PE/PPE	PE_PGRS14	Rv0834c
968424	970244	PE/PPE	PE_PGRS15	Rv0872c
976872	978203	PE/PPE	PPE13	Rv0878c
1020058	1021329	PE/PPE	PPE14	Rv0915c
1021344	1021643	PE/PPE	PE7	Rv0916c
1090373	1093144	PE/PPE	PE_PGRS16	Rv0977
1093361	1094356	PE/PPE	PE_PGRS17	Rv0978c
1095078	1096451	PE/PPE	PE_PGRS18	Rv0980c
1130191	1131111	PE/PPE	ispE	Rv1011
1161297	1162472	PE/PPE	PPE15	Rv1039c
1162549	1163376	PE/PPE	PE8	Rv1040c
1188421	1190424	PE/PPE	PE_PGRS19	Rv1067c
1190757	1192148	PE/PPE	PE_PGRS20	Rv1068c
1211560	1213863	PE/PPE	PE_PGRS21	Rv1087
1214513	1214947	PE/PPE	PE9	Rv1088
1214769	1215131	PE/PPE	PE10	Rv1089
1216469	1219030	PE/PPE	PE_PGRS22	Rv1091
1262272	1264128	PE/PPE	PPE16	Rv1135c
1298764	1299804	PE/PPE	PPE17	Rv1168c
1299822	1300124	PE/PPE	lipX	Rv1169c
1301755	1302681	PE/PPE	PE12	Rv1172c
1339003	1339302	PE/PPE	PE13	Rv1195
1339349	1340524	PE/PPE	PPE18	Rv1196
1357293	1357625	PE/PPE	PE14	Rv1214c
1384989	1386677	PE/PPE	PE_PGRS23	Rv1243c
1488154	1489965	PE/PPE	PE_PGRS24	Rv1325c
1532443	1533633	PE/PPE	PPE19	Rv1361c
1561464	1561772	PE/PPE	PE15	Rv1386
1561769	1563388	PE/PPE	PPE20	Rv1387
1572127	1573857	PE/PPE	PE_PGRS25	Rv1396c
1606386	1607972	PE/PPE	PE16	Rv1430
1618209	1619684	PE/PPE	PE_PGRS26	Rv1441c
1630638	1634627	PE/PPE	PE_PGRS27	Rv1450c
1636004	1638229	PE/PPE	PE_PGRS28	Rv1452c
1655609	1656721	PE/PPE	PE_PGRS29	Rv1468c
1751297	1753333	PE/PPE	PPE21	Rv1548c
1855764	1856696	PE/PPE	PE17	Rv1646
1862347	1865382	PE/PPE	PE_PGRS30	Rv1651c
1931497	1932654	PE/PPE	PPE22	Rv1705c
1932694	1933878	PE/PPE	PPE23	Rv1706c
1981614	1984775	PE/PPE	PPE24	Rv1753c
1989833	1992577	PE/PPE	wag22	Rv1759c
2000614	2002470	PE/PPE	PE_PGRS31	Rv1768
2025301	2026398	PE/PPE	PPE25	Rv1787
2026477	2026776	PE/PPE	PE18	Rv1788
2026790	2027971	PE/PPE	PPE26	Rv1789
2028425	2029477	PE/PPE	PPE27	Rv1790
2029904	2030203	PE/PPE	PE19	Rv1791
2039453	2041420	PE/PPE	PPE28	Rv1800
2042001	2043272	PE/PPE	PPE29	Rv1801
2043384	2044775	PE/PPE	PPE30	Rv1802
2044923	2046842	PE/PPE	PE_PGRS32	Rv1803c
2048072	2048371	PE/PPE	PE20	Rv1806
2048398	2049597	PE/PPE	PPE31	Rv1807
2049921	2051150	PE/PPE	PPE32	Rv1808

2051282	2052688	PE/PPE	PPE33	Rv1809
2061178	2062674	PE/PPE	PE_PGRS33	Rv1818c
2087971	2089518	PE/PPE	PE_PGRS34	Rv1840c
2162932	2167311	PE/PPE	PPE34	Rv1917c
2167649	2170612	PE/PPE	PPE35	Rv1918c
2226244	2227920	PE/PPE	PE_PGRS35	Rv1983
2356729	2358033	PE/PPE	PE_PGRS36	Rv2098c
2358033	2358206	PE/PPE	PE21	Rv2099c
2367359	2367655	PE/PPE	PE22	Rv2107
2367711	2368442	PE/PPE	PPE36	Rv2108
2381071	2382492	PE/PPE	PPE37	Rv2123
2387202	2387972	PE/PPE	PE_PGRS37	Rv2126c
2423240	2424838	PE/PPE	PE_PGRS38	Rv2162c
2600731	2601879	PE/PPE	PE23	Rv2328
2617667	2618908	PE/PPE	PE_PGRS39	Rv2340c
2632923	2634098	PE/PPE	PPE38	Rv2352c
2634528	2635592	PE/PPE	PPE39	Rv2353c
2637688	2639535	PE/PPE	PPE40	Rv2356c
2651753	2651938	PE/PPE	PE_PGRS40	Rv2371
2692799	2693884	PE/PPE	PE_PGRS41	Rv2396
2706017	2706736	PE/PPE	PE24	Rv2408
2727336	2727920	PE/PPE	PPE41	Rv2430c
2727967	2728266	PE/PPE	PE25	Rv2431c
2795301	2797385	PE/PPE	PE_PGRS42	Rv2487c
2801254	2806236	PE/PPE	PE_PGRS43	Rv2490c
2835785	2837263	PE/PPE	PE26	Rv2519
2921551	2923182	PE/PPE	PE_PGRS44	Rv2591
2935046	2936788	PE/PPE	PPE42	Rv2608
2943600	2944985	PE/PPE	PE_PGRS45	Rv2615c
2960105	2962441	PE/PPE	PE_PGRS46	Rv2634c
3053914	3055491	PE/PPE	PE_PGRS47	Rv2741
3076894	3078078	PE/PPE	PPE43	Rv2768c
3078158	3078985	PE/PPE	PE27	Rv2769c
3079309	3080457	PE/PPE	PPE44	Rv2770c
3162268	3164115	PE/PPE	PE_PGRS48	Rv2853
3179368	3180531	PE/PPE	gcpE	Rv2868c
3200794	3202020	PE/PPE	PPE45	Rv2892c
3376939	3378243	PE/PPE	PPE46	Rv3018c
3378329	3378415	PE/PPE	PE27A	Rv3018A
3379036	3379329	PE/PPE	esxS	Rv3020c
3379376	3380452	PE/PPE	PPE47	Rv3021c
3380440	3380682	PE/PPE	PPE48	Rv3022c
3380679	3380993	PE/PPE	PE29	Rv3022A
3465778	3467091	PE/PPE	lipY	Rv3097c
3490476	3491651	PE/PPE	PPE49	Rv3125c
3501334	3501732	PE/PPE	PPE50	Rv3135
3501794	3502936	PE/PPE	PPE51	Rv3136
3510088	3511317	PE/PPE	PPE52	Rv3144c
3527391	3529163	PE/PPE	PPE53	Rv3159c
3729364	3736935	PE/PPE	PPE54	Rv3343c
3736984	3738438	PE/PPE	PE_PGRS49	Rv3344c
3738158	3742774	PE/PPE	PE_PGRS50	Rv3345c
3743711	3753184	PE/PPE	PPE55	Rv3347c
3755952	3767102	PE/PPE	PPE56	Rv3350c
3778568	3780334	PE/PPE	PE_PGRS51	Rv3367
3801653	3803848	PE/PPE	PE_PGRS52	Rv3388
3842239	3842769	PE/PPE	PPE57	Rv3425
3843036	3843734	PE/PPE	PPE58	Rv3426
3847165	3847701	PE/PPE	PPE59	Rv3429
3894093	3894389	PE/PPE	PE31	Rv3477
3894426	3895607	PE/PPE	PPE60	Rv3478
3926569	3930714	PE/PPE	PE_PGRS53	Rv3507
3931005	3936710	PE/PPE	PE_PGRS54	Rv3508
3939617	3941761	PE/PPE	PE_PGRS55	Rv3511
3941724	3944963	PE/PPE	PE_PGRS56	Rv3512
3945794	3950263	PE/PPE	PE_PGRS57	Rv3514

3969343	3970563	PE/PPE	PPE61	Rv3532
3970705	3972453	PE/PPE	PPE62	Rv3533c
3978059	3979498	PE/PPE	PPE63	Rv3539
3997980	3999638	PE/PPE	PPE64	Rv3558
4031404	4033158	PE/PPE	PE_PGRS58	Rv3590c
4036731	4038050	PE/PPE	PE_PGRS59	Rv3595c
4060648	4061889	PE/PPE	PPE65	Rv3621c
4061899	4062198	PE/PPE	PE32	Rv3622c
4091233	4091517	PE/PPE	PE33	Rv3650
4093632	4093946	PE/PPE	PE_PGRS60	Rv3652
4093940	4094527	PE/PPE	PE_PGRS61	Rv3653
4189285	4190232	PE/PPE	PPE66	Rv3738c
4190284	4190517	PE/PPE	PPE67	Rv3739c
4196171	4196506	PE/PPE	PE34	Rv3746c
4276571	4278085	PE/PPE	PE_PGRS62	Rv3812
4286721	4287935	PE/PPE	.	Rv3822
4350745	4351044	PE/PPE	PE35	Rv3872
4351075	4352181	PE/PPE	PPE68	Rv3873
4374484	4375683	PE/PPE	PPE69	Rv3892c
4375762	4375995	PE/PPE	PE36	Rv3893c
23173	23273	Repeat	.	.
79507	79551	Repeat	.	.
80236	80550	Repeat	.	.
103713	105215	Repeat	.	.
154073	154125	Repeat	.	.
154126	154178	Repeat	.	.
154179	154231	Repeat	.	.
206812	206850	Repeat	.	.
206869	206907	Repeat	.	.
272855	272955	Repeat	.	.
547488	547517	Repeat	.	.
580578	580654	Repeat	.	.
580655	580731	Repeat	.	.
580732	580808	Repeat	.	.
616828	616878	Repeat	.	.
642754	642811	Repeat	.	.
701247	701369	Repeat	.	.
703912	703985	Repeat	.	.
706790	706863	Repeat	.	.
709425	709548	Repeat	.	.
709585	709663	Repeat	.	.
711624	711702	Repeat	.	.
795467	795518	Repeat	.	.
802429	802477	Repeat	.	.
812835	812921	Repeat	.	.
812922	812975	Repeat	.	.
863155	863255	Repeat	.	.
889017	889020	Repeat	.	.
889021	889048	Repeat	.	.
890348	890375	Repeat	.	.
890376	890379	Repeat	.	.
960173	960225	Repeat	.	.
960226	960278	Repeat	.	.
960279	960333	Repeat	.	.
1025458	1025472	Repeat	.	.
1026879	1026893	Repeat	.	.
1027061	1027076	Repeat	.	.
1029345	1029360	Repeat	.	.
1164572	1165435	Repeat	.	.
1165532	1165549	Repeat	.	.
1179345	1179395	Repeat	.	.
1251621	1252945	Repeat	.	.
1276296	1277643	Repeat	.	.
1277843	1277846	Repeat	.	.
1277847	1277863	Repeat	.	.
1278800	1278816	Repeat	.	.

1278817	1278820	Repeat	.	.
1305495	1305556	Repeat	.	.
1305557	1305618	Repeat	.	.
1305619	1305661	Repeat	.	.
1456585	1456627	Repeat	.	.
1457453	1457504	Repeat	.	.
1457505	1457557	Repeat	.	.
1468143	1468161	Repeat	.	.
1469633	1469651	Repeat	.	.
1507531	1507581	Repeat	.	.
1541949	1541951	Repeat	.	.
1543307	1543309	Repeat	.	.
1612558	1612578	Repeat	.	.
1612579	1612599	Repeat	.	.
1612600	1612620	Repeat	.	.
1612621	1612641	Repeat	.	.
1612642	1612662	Repeat	.	.
1625366	1625418	Repeat	.	.
1644261	1644313	Repeat	.	.
1644314	1644364	Repeat	.	.
1779266	1779277	Repeat	.	.
1779959	1780047	Repeat	.	.
1780485	1780573	Repeat	.	.
1788514	1789811	Repeat	.	.
1830074	1830125	Repeat	.	.
1907460	1907515	Repeat	.	.
1907516	1907571	Repeat	.	.
1927218	1928589	Repeat	.	.
1938093	1938145	Repeat	.	.
1944756	1944808	Repeat	.	.
1982965	1983042	Repeat	.	.
1983043	1983120	Repeat	.	.
1983121	1983198	Repeat	.	.
1983199	1983276	Repeat	.	.
1983277	1983354	Repeat	.	.
1987703	1987730	Repeat	.	.
1989030	1989057	Repeat	.	.
1996101	1996128	Repeat	.	.
1997428	1997455	Repeat	.	.
1998584	1998597	Repeat	.	.
1999800	1999813	Repeat	.	.
2059441	2059498	Repeat	.	.
2059518	2059575	Repeat	.	.
2163393	2163461	Repeat	.	.
2163462	2163530	Repeat	.	.
2163741	2163809	Repeat	.	.
2163810	2163878	Repeat	.	.
2163879	2163947	Repeat	.	.
2163948	2164016	Repeat	.	.
2164017	2164085	Repeat	.	.
2195989	2197350	Repeat	.	.
2330147	2330225	Repeat	.	.
2365414	2365441	Repeat	.	.
2366741	2366768	Repeat	.	.
2372437	2372492	Repeat	.	.
2372494	2372549	Repeat	.	.
2430117	2430144	Repeat	.	.
2431444	2431471	Repeat	.	.
2458392	2458449	Repeat	.	.
2493801	2493818	Repeat	.	.
2522173	2522230	Repeat	.	.
2523184	2523236	Repeat	.	.
2531898	2531950	Repeat	.	.
2531951	2532003	Repeat	.	.
2532004	2532056	Repeat	.	.
2532057	2532109	Repeat	.	.

2532110	2532162	Repeat	.	.
2532163	2532212	Repeat	.	.
2550011	2550013	Repeat	.	.
2550014	2550041	Repeat	.	.
2551341	2551368	Repeat	.	.
2551369	2551371	Repeat	.	.
2636904	2636931	Repeat	.	.
2687128	2687179	Repeat	.	.
2687180	2687257	Repeat	.	.
2716315	2716391	Repeat	.	.
2720644	2720656	Repeat	.	.
2721844	2721856	Repeat	.	.
2762762	2763061	Repeat	.	.
2763397	2763696	Repeat	.	.
2784614	2784642	Repeat	.	.
2785942	2785970	Repeat	.	.
2800671	2800918	Repeat	.	.
2806368	2806625	Repeat	.	.
2972106	2972108	Repeat	.	.
2972109	2972136	Repeat	.	.
2973436	2973463	Repeat	.	.
2973464	2973466	Repeat	.	.
2983019	2983033	Repeat	.	.
2996003	2996053	Repeat	.	.
2996054	2996104	Repeat	.	.
2996105	2996155	Repeat	.	.
3007063	3007115	Repeat	.	.
3007116	3007168	Repeat	.	.
3007169	3007221	Repeat	.	.
3013612	3013687	Repeat	.	.
3073055	3073112	Repeat	.	.
3119185	3119294	Repeat	.	.
3119335	3119370	Repeat	.	.
3119411	3119446	Repeat	.	.
3119484	3119519	Repeat	.	.
3119556	3119591	Repeat	.	.
3119627	3119662	Repeat	.	.
3119701	3119736	Repeat	.	.
3119777	3119812	Repeat	.	.
3119848	3119883	Repeat	.	.
3119921	3119956	Repeat	.	.
3119995	3120030	Repeat	.	.
3120068	3120103	Repeat	.	.
3120141	3120176	Repeat	.	.
3120213	3120248	Repeat	.	.
3120285	3120320	Repeat	.	.
3120359	3120394	Repeat	.	.
3120433	3120468	Repeat	.	.
3120504	3120523	Repeat	.	.
3121882	3121897	Repeat	.	.
3121938	3121973	Repeat	.	.
3122013	3122048	Repeat	.	.
3122086	3122121	Repeat	.	.
3122158	3122193	Repeat	.	.
3122230	3122265	Repeat	.	.
3122303	3122338	Repeat	.	.
3122375	3122410	Repeat	.	.
3122436	3122471	Repeat	.	.
3122513	3122548	Repeat	.	.
3122585	3122620	Repeat	.	.
3122661	3122696	Repeat	.	.
3122738	3122773	Repeat	.	.
3122811	3122846	Repeat	.	.
3122882	3122917	Repeat	.	.
3122955	3122990	Repeat	.	.
3123029	3123064	Repeat	.	.

3123102	3123137	Repeat	.	.
3123173	3123208	Repeat	.	.
3123248	3123283	Repeat	.	.
3123318	3123353	Repeat	.	.
3123390	3123425	Repeat	.	.
3123467	3123502	Repeat	.	.
3123541	3123576	Repeat	.	.
3155874	3155927	Repeat	.	.
3155928	3155981	Repeat	.	.
3155982	3156035	Repeat	.	.
3156036	3156089	Repeat	.	.
3160522	3160583	Repeat	.	.
3171468	3171518	Repeat	.	.
3171522	3171572	Repeat	.	.
3171576	3171616	Repeat	.	.
3181794	3181836	Repeat	.	.
3192202	3192254	Repeat	.	.
3192255	3192307	Repeat	.	.
3192308	3192360	Repeat	.	.
3318835	3318889	Repeat	.	.
3319468	3319568	Repeat	.	.
3319569	3319666	Repeat	.	.
3333768	3333773	Repeat	.	.
3335787	3335792	Repeat	.	.
3381351	3381365	Repeat	.	.
3382660	3382674	Repeat	.	.
3481399	3481413	Repeat	.	.
3482708	3482722	Repeat	.	.
3551227	3551229	Repeat	.	.
3551230	3551257	Repeat	.	.
3552557	3552584	Repeat	.	.
3552585	3552587	Repeat	.	.
3552710	3552712	Repeat	.	.
3552713	3552740	Repeat	.	.
3554040	3554067	Repeat	.	.
3554068	3554070	Repeat	.	.
3591493	3591569	Repeat	.	.
3626614	3626666	Repeat	.	.
3658658	3658715	Repeat	.	.
3704895	3705004	Repeat	.	.
3710382	3710409	Repeat	.	.
3711709	3711736	Repeat	.	.
3743198	3743510	Repeat	.	.
3769514	3769720	Repeat	.	.
3769754	3769862	Repeat	.	.
3770994	3771091	Repeat	.	.
3795058	3795085	Repeat	.	.
3796385	3796412	Repeat	.	.
3799987	3800011	Repeat	.	.
3801530	3801554	Repeat	.	.
3883550	3884921	Repeat	.	.
3890779	3890806	Repeat	.	.
3892106	3892133	Repeat	.	.
3945098	3945597	Repeat	.	.
3950830	3951329	Repeat	.	.
3991568	3991625	Repeat	.	.
4052949	4052966	Repeat	.	.
4052971	4052994	Repeat	.	.
4052995	4053021	Repeat	.	.
4053106	4053216	Repeat	.	.
4053217	4053327	Repeat	.	.
4053328	4053438	Repeat	.	.
4053439	4053549	Repeat	.	.
4075615	4075630	Repeat	.	.
4077735	4077750	Repeat	.	.
4078506	4078518	Repeat	.	.

4079786	4079798	Repeat	.	.
4134601	4134725	Repeat	.	.
4348721	4348773	Repeat	.	.
4348774	4348826	Repeat	.	.
4353280	4353330	Repeat	.	.
4353331	4353381	Repeat	.	.
4353382	4353432	Repeat	.	.

Supplemental Table 2: Population size estimates and sequencing coverage by zone.

Zone	Population (%)	Cases in Zone	Cases with Sequencing Data (%)
1	181,021 (20.6%)	1365	188 (14%)
2	166,388 (19%)	860	88 (10%)
3	9,586 (1.1%)	55	2 (4%)
4	124,605 (14.2%)	1249	187 (15%)
5	42,500 (4.8%)	305	42 (14%)
6	144,702 (16.5%)	1065	125 (12%)
7	208,753 (23.8%)	1678	209 (13%)

Supplemental Table 3: Characteristics of sequenced and not sequenced culture positive TB cases

A Characteristics of sequenced and not sequenced cases

	Not Sequenced (N=3005)	Sequenced (N=861)	Overall (N=3866)
Sex			
Male	1945 (64.7%)	597 (69.3%)	2542 (65.8%)
Female	1060 (35.3%)	264 (30.7%)	1324 (34.2%)
Age			
Mean (SD)	34.9 (12.1)	33.9 (11.5)	34.7 (12.0)
Median [Min, Max]	34.0 [0, 89.0]	34.0 [1.00, 89.0]	34.0 [0, 89.0]
HIV status			
Negative	1166 (38.8%)	324 (37.6%)	1490 (38.5%)
Positive	1770 (58.9%)	517 (60.0%)	2287 (59.2%)
Missing	69 (2.3%)	20 (2.3%)	89 (2.3%)

B Count and percentage of cases in each zone that were sequenced or not sequenced

	Not Sequenced (N=1532)	Sequenced (N=841)	Overall (N=2373)
Zone			
1	321 (63.1%)	188 (36.9%)	509
2	177 (66.8%)	88 (33.2%)	265
3	10 (83.3%)	2 (16.7%)	12
4	278 (59.8%)	187 (40.2%)	465
5	68 (61.8%)	42 (38.2%)	110
6	284 (69.4%)	125 (30.6%)	409
7	394 (65.3%)	209 (34.7%)	603

C Count and percentage of cases in each year that were sequenced or not sequenced

	2015 (N=999)	2016 (N=916)	2017 (N=821)	2018 (N=642)	2019 (N=488)
Sequenced					
Not Sequenced	876 (87.7%)	650 (71.0%)	547 (66.6%)	539 (84.0%)	393 (80.5%)
Sequenced	123 (12.3%)	266 (29.0%)	274 (33.4%)	103 (16.0%)	95 (19.5%)

Supplemental Table 4: Summary of transmission networks used in the DBM analysis

Transmission network	Total	HIV positive	Cases within hotspots	Cases within hotspots & HIV positive
#2	25	15 (60%)	-	-
#43	17	10 (59%)	6 (35%)	4 (67%)
#71	15	3 (20%)	4 (27%)	1 (25%)
#5	9	5 (56%)	-	-
#74	9	6 (67%)	44%	75%
#61	8	4 (50%)	-	-
#80	8	3 (38%)	6 (75%)	2 (33%)
#1	5	3 (60%)	0 (0%)	0 (0%)
#13	5	4 (80%)	-	-
#19	5	4 (80%)	-	-
#28	5	1 (20%)	-	-
#40	5	1 (20%)	-	-
#9	5	3 (60%)	-	-

Transmission networks with spatial foci of transmission are denoted with red text.

Supplemental Table 5: Estimated transmission flow (posterior means) between and within zones of Blantyre

		Infected Zone						
		1	2	3	4	5	6	7
Infector Zone	1	0.06	0.03	0.00	0.04	0.02	0.04	0.05
	2	0.02	0.04	0.00	0.01	0.00	0.03	0.03
	3	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	4	0.03	0.02	0.00	0.07	0.01	0.00	0.05
	5	0.02	0.00	0.00	0.01	0.00	0.00	0.02
	6	0.03	0.03	0.00	0.01	0.01	0.06	0.03
	7	0.04	0.02	0.00	0.05	0.01	0.03	0.08