Spatial patterns of extensively drug-resistant tuberculosis transmission in KwaZulu-Natal, South Africa

Kristin N. Nelson,[1] N. Sarita Shah,[1,2] Barun Mathema,[3] Nazir Ismail,[4,5] James C.M. Brust,[6] Tyler S. Brown,[7] Sara C. Auld,[1,8] Shaheed Valley Omar,[4] Natashia Morris,[9] Angie Campbell,[1] Salim Allana,[1] Pravi Moodley,[10,11] Koleka Mlisana,[10,11] Neel R. Gandhi[1,8]

1. Emory University Rollins School of Public Health, Atlanta, GA, USA
2. Centers for Disease Control and Prevention, Atlanta, GA, USA
3. Columbia University Mailman School of Public Health, New York, NY, USA
4. National Institute for Communicable Diseases, Johannesburg, South Africa
5. University of Pretoria, Pretoria, South Africa
6. Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, NY, USA
7. Massachusetts General Hospital, Infectious Diseases Division, Boston, MA, USA
8. Emory University School of Medicine, Atlanta, GA, USA
9. Environment and Health Research Unit, South African Medical Research Council, Johannesburg, South Africa
10. National Health Laboratory Service, Durban, South Africa
11. School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal, Durban, South Africa

Abstract word count: 200 words (limit: 200 words)

Manuscript word Count: 3255 words (limit: 3500 words)

Running title: Geospatial patterns of XDR TB transmission

Corresponding Author:

Neel R. Gandhi, MD

Phone: +1 404 727-2317; Fax: (404) 727-8737; Email: neel.r.gandhi@emory.edu

Abstract:

Background: Transmission is driving the global drug-resistant tuberculosis epidemic; nearly three-quarters of drug-resistant tuberculosis cases are attributable to transmission. Geographic patterns of disease incidence, combined with information on probable transmission links, can define the spatial scale of transmission and generate hypotheses about factors driving transmission patterns.

Methods: We combined whole-genome sequencing data with home GPS coordinates from 344 participants with extensively drug-resistant (XDR) tuberculosis in KwaZulu-Natal, South Africa diagnosed from 2011-2014. We aimed to determine if genomically linked (≤5 single nucleotide polymorphisms [SNP] differences) cases lived close to one another, which would suggest a role for local community settings in transmission.

Results: 182 study participants were genomically linked, comprising 1084 case-pairs. The median distance between case-pairs' homes was 108 km (IQR: 64-162 km). Between-district, as compared to within-district, links accounted for the majority (912/1084, 84%) of genomic links. Half (526, 49%) of genomic links involved a case from Durban, the urban center of KwaZulu-Natal.

Conclusions: The high proportions of between-district links with Durban provide insight into possible drivers of province-wide XDR-tuberculosis transmission, including urban-rural migration. Further research should focus on characterizing the contribution of these drivers to overall XDR-tuberculosis transmission in KwaZulu-Natal to inform design of targeted strategies to curb the drug-resistant tuberculosis epidemic.

Key words: tuberculosis, drug-resistance; extensively drug-resistant tuberculosis; molecular epidemiology; whole genome sequencing; transmission; geospatial analysis

Introduction

Drug-resistant tuberculosis is a global crisis, causing an estimated 1.2 million cases each year.[1] Extensively drug-resistant (XDR) tuberculosis has now been reported from 123 countries and is associated with mortality rates from 50-90%.[2-4] Although drug-resistant tuberculosis strains are initially created by selection of drug-resistant mutants during treatment (acquired resistance), recent studies show that the majority of drug-resistant tuberculosis cases now arise due to transmission of already drug-resistant strains.[5, 6] This shift makes clear the urgent need for interventions to prevent transmission.

Molecular epidemiology studies have consistently shown that close contacts account for only a minority of secondary tuberculosis cases in settings with high tuberculosis incidence, suggesting that a substantial proportion of transmission may occur as a result of 'casual' contact in the community. [7-11] Although modeling and social mixing studies support this hypothesis and point to public transportation, schools and workplaces as likely transmission sites in high tuberculosis incidence settings, this has not been demonstrated directly.[12-14] Understanding the role of contacts proximate to or distant from the home can generate hypotheses about the modes of contact driving transmission. The advent of bacterial whole-genome sequencing (WGS) offers new opportunities to identify tuberculosis cases that are likely to be linked through transmission, by discriminating between TB isolates at the level of single nucleotide polymorphisms (SNPs). Isolates from different patients that differ by small numbers of SNPs are considered likely to represent a transmission event. Recent studies have employed WGS to identify probable transmission events, map chains of transmission in tuberculosis outbreaks, and describe the burden of tuberculosis disease due to recent infection as compared to reactivation.[10, 15-20] However, WGS has been underutilized to describe broader, population-level patterns of transmission in tuberculosis-endemic settings.

The spatial scale of disease transmission can provide insight into the settings and, by extension, the modes of contact that contribute to transmission. Tuberculosis transmission

requires air exchange—and therefore close proximity—between an infectious and susceptible person. The nature and location of these interactions define the relevant geographic scale for person-to-person interactions resulting in transmission.[21, 22] For example, short distances between transmission-linked cases may indicate that local contacts in, or close to, the household are most important in transmission. Alternatively, transmission links found across longer distances may indicate that long-distance contacts, and perhaps migration, may play an important role in disseminating disease. Previous geospatial analyses in tuberculosis have focused on the spatial distribution of *cases*, rather than the spatial scale of transmission *links*. Combining geospatial analysis with WGS data has the potential to provide more comprehensive information about the dynamic process of disease transmission.

We combined *Mycobacterium tuberculosis* (*Mtb*) whole-genome sequencing and geographic data to, first, evaluate the spatial scale of XDR-tuberculosis transmission in KwaZulu-Natal, South Africa, and second, quantify the proportion of transmission occurring within and between municipal districts in KwaZulu-Natal. Understanding the spatial scale and patterns of transmission can identify specific geographic areas and demographic groups that contribute to ongoing transmission and towards which interventions can be targeted.

Methods

Setting

South Africa has among the highest rates of tuberculosis globally, with 59% of tuberculosis patients co-infected with HIV. [1, 23] KwaZulu-Natal province, which comprises 11 districts and has a population of 10.3 million persons, has the highest tuberculosis and XDR-tuberculosis burden (3 per 100,000) in South Africa.[24, 25] [26] The most populous district, eThekwini, is home to the city of Durban, a common destination for employment and educational opportunities. The population in KwaZulu-Natal is highly mobile— a recent study found that over a third of the population had changed residence in the past two years.[27]

Study design and procedures

The Transmission of HIV-Associated XDR-tuberculosis (TRAX) study is a cross-sectional study that enrolled culture-confirmed XDR-tuberculosis patients diagnosed from 2011 to 2014 in KwaZulu-Natal. Detailed methods of the TRAX study have been previously published.[5] Briefly, we identified XDR-tuberculosis cases through the single referral laboratory that conducts drug-susceptibility testing (DST) for all public healthcare facilities in KwaZulu-Natal. All participants provided written informed consent; for deceased or severely ill participants, consent was obtained from next-of-kin.

We interviewed participants and performed medical record review to collect demographic information and medical history. Participants reported the locations of residences, schools, employment, hospital admissions and other congregate locations frequented in the five years preceding XDR-tuberculosis diagnosis. A global position system (GPS) coordinate location was collected at the location of each participant's home residence.

Whole genome sequencing

The diagnostic XDR *Mtb* isolate was obtained for all participants and re-cultured on Löwenstein-Jensen slants. We conducted population sweeps, extracted genomic DNA, and prepared sequencing libraries using Nextera DNA kits (Illumina, San Diego, CA). Raw paired-end sequencing reads were generated on the Illumina (MiSeq) platform and aligned to the H37Rv reference genome (NC_000962.3) using the Burrows-Wheeler Aligner. All isolates had reads covering >99% of the reference genome, and the lowest mean coverage depth for any isolate was 15X. SNPs were detected using standard pairwise resequencing techniques (Samtools v0.1.19) against the reference and filtered for quality, read consensus (>75% reads for the alternate allele) and proximity to indels (>50 base-pairs from any indel). SNPs in or within 50 base pairs of hypervariable PPE/PE gene families, repeat regions, and mobile elements

were excluded.[28] Alignment files can be found at NCBI Bioproject PRJNA476470.

Analysis

We defined a genomic link as a pair of XDR-tuberculosis cases ('case-pair') with 5 or fewer SNP differences between their *Mtb* sequences.[20, 29, 30] We mapped and calculated median geographic distance between the home residences of genomically linked cases using the *sp* and *geosphere* packages in R 3.4.1.[31] [32]

We stratified distances between genomically linked cases by sex, given historically distinct migratory behavior among male and females in sub-Saharan Africa. We also stratified by HIV coinfection, since the influence of HIV on the susceptibility, progression, and transmissibility of tuberculosis remains uncertain.[9, 29, 33-35] Lastly, we stratified by strain type, by comparing pairs of the most common *Mtb* strain type in KwaZulu-Natal, LAM4, with other strain types. We conducted our analysis at varying SNP thresholds (≤3 SNPs, ≤1 SNP) to assess the robustness of results to this choice.

To describe patterns of transmission by district of residence, we classified each case according to the district of their home residence and calculated the proportion of between- and within-district genomic links for all districts. We also calculated the proportion of pairs in each district with links to the urban district of eThekwini.

Sensitivity analysis of differential enrollment in TRAX by district

To assess whether our results were sensitive to differential enrollment of XDR cases by district, we compared our results to those we might have observed had we enrolled all cases. We used the complete register of diagnosed XDR-tuberculosis cases from the referral laboratory to calculate the fraction of diagnosed cases from each district that participated in TRAX (enrollment fraction). For within-district links, we adjusted the number of genomic links by a factor of the inverse enrollment fraction. For between-district links, we adjusted the number of

links using the mean of the inverse enrollment fractions for both districts. We compared the proportions of within- and between-district links calculated using these enrollment fractions to the proportions we observed.

As cases from rural areas may have reduced access to high-quality healthcare services, we hypothesized they may be underdiagnosed, and thus included in TRAX, compared to cases from urban eThekwini district.[36, 37] To examine the effect of this potential source of bias, we varied our assumptions about the extent of this over-enrollment (assuming the enrollment fraction was anywhere from 20-40% higher in eThekwini than in other districts) and repeated our analysis of between and within-district links.

Ethical Considerations

The study was approved by the Institutional Review Boards of Emory University, Albert Einstein College of Medicine, and the University of KwaZulu-Natal, and by CDC's National Center for HIV, Hepatitis, STDs and Tuberculosis.

Results

Between 2011 and 2014, we screened 521 (51%) of 1027 culture-confirmed XDR-tuberculosis patients diagnosed in KwaZulu-Natal and enrolled 404 (78% of screened) (Figure 1). TRAX participants were similar to all diagnosed XDR-tuberculosis cases in terms of age ($p=0.52$), sex ($p=0.76$), and district of diagnosing facility ($p=0.70$). Among the 404 participants, 234 (58%) were female, with a median age of 34 years (interquartile range [IQR]: 28-43). Three hundred eleven (77%) participants were HIV-positive, of whom 236 (76%) were on antiretroviral therapy and 155 (50%) were virologically suppressed at enrollment (viral load < 400 copies/mL) (Table 1). Half (n = 204, 50%) of participants reported living in urban sub-districts, and 133 (33%) participants lived in eThekwini district. Mobility of TRAX participants was high, with 89 (22%) participants reporting living at a different residence than their current residence in the

Figure 1. Selection of study participants and identification of genomic links using whole genome sequencing (WGS) and varying single nucleotide polymorphism (SNP) thresholds.

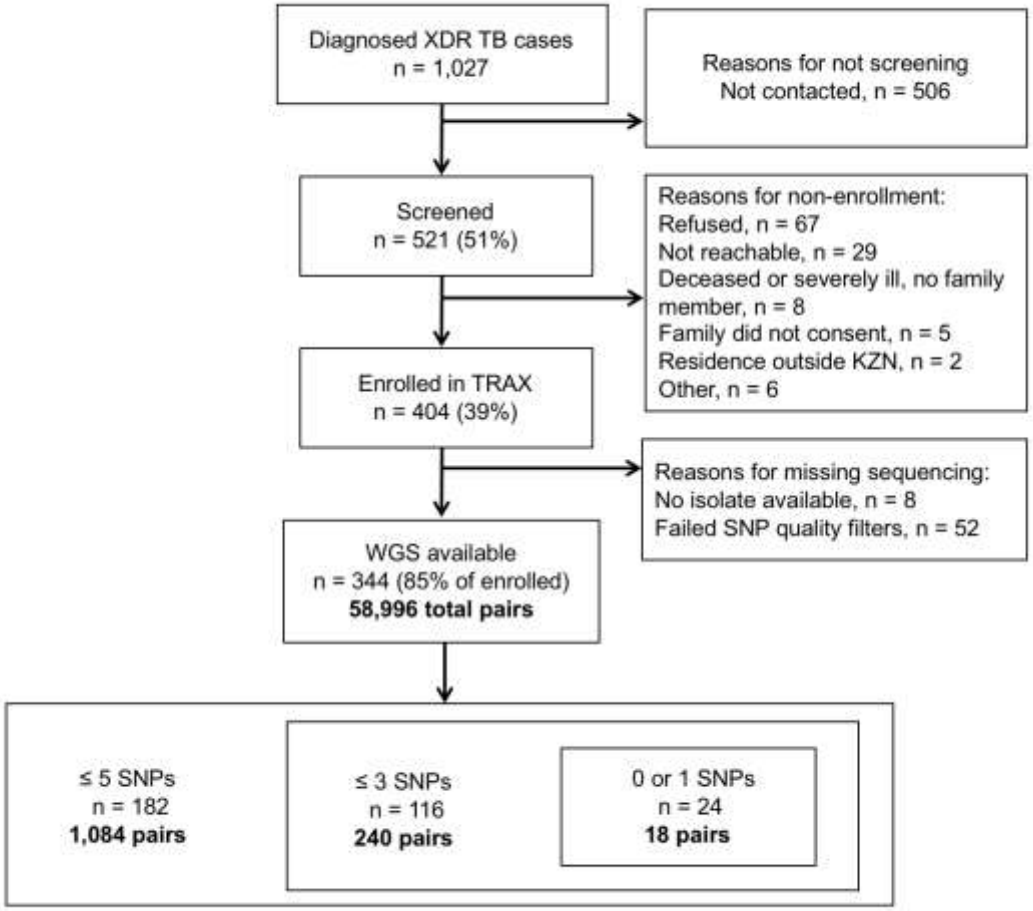


Table 1. Characteristics of participants in TRAX cohort, and comparison to subset with Whole Genome Sequencing (WGS) results and with genomic links – KwaZulu-Natal Province, South Africa

Characteristic	TRAX cohort, n=404 n (%)	Cases with WGS, n=344 n (%)	p-value[1]	Genomically linked cases (≤5 SNPs), n=182 n (%)	p-value[2]
Demographic					
Female	234 (58)	202 (59)	0.44	111 (61)	0.37
Age, median (IQR)	34 (28-43)	34 (29-43)	0.19	34 (29-44)	0.97
0-15 yr	16 (4)	12 (3)	0.21	9 (5)	0.47
16-34 yr	207 (51)	171 (50)		88 (48)	
35-54 yr	150 (37)	134 (39)		71 (39)	
≥55 yr	31 (8)	27 (8)		14 (8)	
Monthly household income					
<R500	139 (34)	120 (35)	0.36	64 (35)	0.27
R500-R2,500	186 (46)	153 (44)		83 (46)	
>R2,500	79 (20)	71 (21)		35 (19)	
Clinical					
Current or former smoker	39 (10)	35 (10)	0.47	18 (10)	0.98
Diabetes	23 (6)	22 (6)	0.15	10 (5)	0.47
HIV positive	311 (77)	266 (77)	0.70	145 (80)	0.27
Receiving antiretroviral therapy	236 (76)	204 (77)	0.49	108/145	0.32
CD4 cell count (median, IQR)	340 (117–431)	240 (111-425)	0.26	233 (104-316)	0.54
Virologic suppression (<400 copies/mL)	155 (50)	134 (39)	0.56	74 (41)	0.49
Cough					
Patients with cough	333 (82)	284 (83)	0.87	147 (81)	0.35
Median duration of cough	8 (4-12)	8 (4-12)	0.22	8 (4-12)	0.39
Sputum smear positive for acid-fast bacilli	270 (67)	235 (68)	0.31	118 (65)	0.16
Previous treatment for any tuberculosis	291 (72)	247 (72)	0.81	127 (70)	0.38
Previous treatment for multidrug-resistant tuberculosis	124 (31)	105 (31)	0.86	45 (25)	0.01

[1]p-values compare cases with WGS (n=344) to all TRAX participants (n=404)
[2]p-values compare linked cases (n=182) to all cases with WGS (n=344)

previous five years; 41 (46%) of those residences were in a district other than their current residence. Inter-district movement was also common—of those participants that reported spending >2 hours per week at congregate locations (n=254), 93 (37%) named a congregate location in a different district than their current residence.

Mtb isolates from 344 (85%) participants passed all sequencing quality filters and were available for analysis, creating a total of 58,996 unique case-pairs. Cases with WGS were similar to all enrolled cases (Table 1). Among these case-pairs, 1084 (1.8%) differed by 5 or fewer SNPs, indicating a genomic link; these case-pairs involved 182 unique participants (Figure 1). Among these 182 cases, the median number of genomic links per case was 6 (IQR: 2-17), with 63 (35%) participants having greater than 10 genomic links (Supplemental Figure 1). These 182 participants reported residences across all eleven districts in KwaZulu-Natal province, and were demographically similar to non-linked cases (Table 1-2, Figure 1).

Geographic distance between genomically linked participants

Among the 1084 genomically linked case-pairs, the homes of 3 (0.3%) case-pairs were within 1 km of one another, 12 (1%) were within 5 km of one another, and 29 (3%) were within 10 km. The majority of case-pairs' homes (871, 80%) were ≥ 50 km apart, and the homes of over half (589, 54%) of case-pairs were more than 100km apart. The median distance between the home residences of genomically linked cases was 108 km (IQR: 64-162 km). This distance was similar when we increased the stringency of the threshold for genomic links: among pairs with fewer than 3 SNPs, the median distance was 117 km (IQR 67-162); among pairs with fewer than 1 SNP difference, the median distance was 127 km (IQR 59-152) (Figure 3). The median distance between case-pairs homes' was >95 km for all strata of sex, HIV status, and strain type. (Supplemental Table 1).

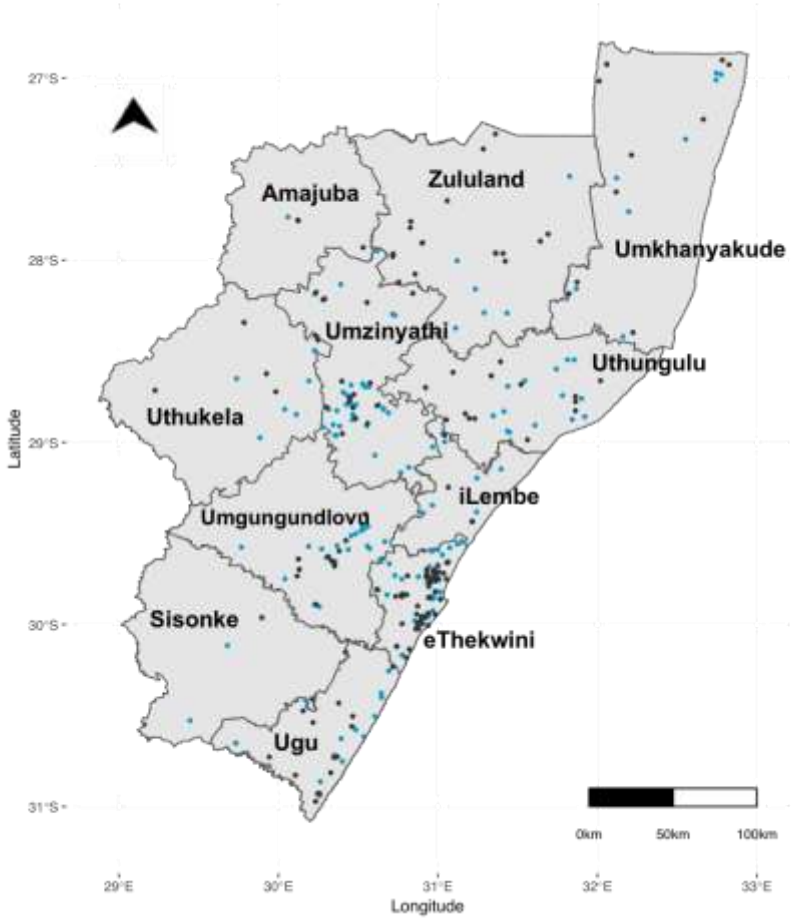
Since some cases had multiple genomic links, we wanted to determine whether cases with distant links also had links close to home. We selected the geographically closest link for each

Table 2. Geographic distribution of XDR-tuberculosis cases by district.

District	n (% of total)	Population (thousands)	Genomically linked (% of total)
Amajuba	4 (1.2)	500 (4.9)	1 (0.5)
eThekwini	115 (33)	3,400 (33)	53 (29)
iLembe	11 (3.2)	607 (5.9)	7 (4)
Sisonke	4 (1.2)	461 (4.5)	3 (2)
Ugu	32 (9.3)	722 (7.0)	14 (8)
UMgungundlovu	37 (10.8)	1,018 (10)	26 (14)
Umkhanyakude	19 (5.5)	626 (6.1)	9 (5)
Umzinyathi	53 (15.4)	510 (5.0)	37 (20)
Uthukela	15 (4.4)	669 (6.5)	9 (5)
Uthungulu	30 (8.7)	908 (8.8)	16 (9)
Zululand	24 (7.0)	840 (8.2)	7 (4)
Total	344	10,261	182

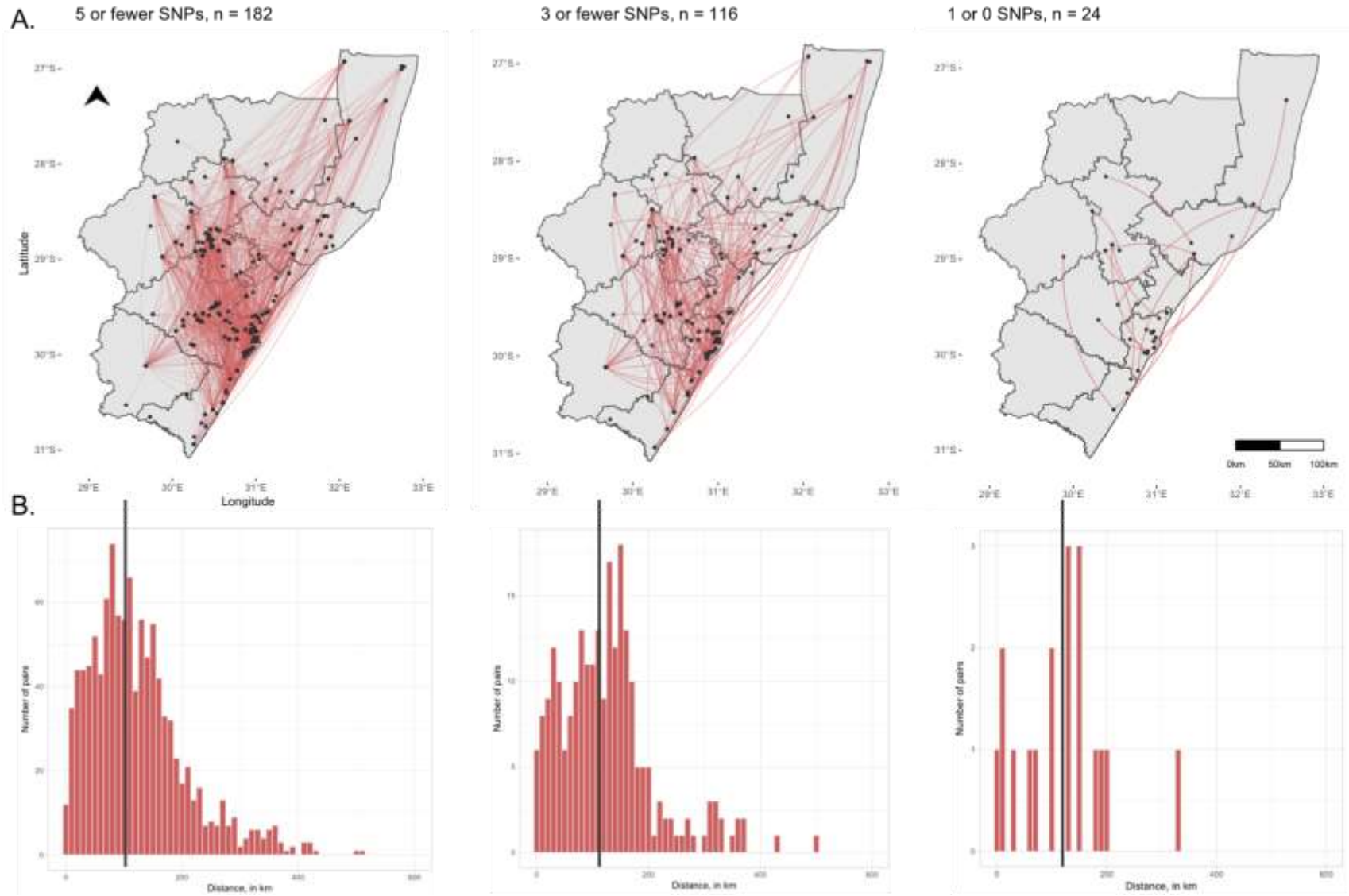
Population by district and percent of cases in each district with at least one genomic link. Population statistics sourced from the Statistics South Africa 2011 Census (http://www.statssa.gov.za/)

Figure 2. Geographic distribution of XDR-tuberculosis cases with genomic links in KwaZulu-Natal province, South Africa**.**



Blue dots indicate georeferenced locations of reported home residences of TRAX cases who are genomically linked; black dots indicate those not genomically linked. The eleven districts of KwaZulu-Natal are labeled. The most populous district in KwaZulu-Natal is eThekwini, which includes the city of Durban. Note: As of 2015, Sisonke district is known as Harry Gwala district and as of 2016, Uthungulu district is known as King Cetshwayo district.

Figure 3. Map and distribution of geographic distances between home residences of genomically linked case-pairs in KwaZulu-Natal.



A. Black dots indicate home residences of XDR-tuberculosis cases; red lines represent genomic links between cases. B. Black lines on histograms indicate the median distance between homes of genomically linked case-pairs at each SNP threshold. Note differences in y axis range across plots.

case. Among the 182 cases involved in genomically linked case-pairs, 20 (11%) cases lived within 5 km of their closest link, 40 (22%) lived within 10 km; 68 (37%) lived more than 50 km from their closest geographic link, and 22 (12%) of cases lived over 100km from their closest link. The median distance to the closest geographic link was 32 km (Supplemental Figure 2).

Within- and between-district links

Overall, 16% of genomic links were among case-pairs residing within the same district (172/1084), while 84% of genomically linked case-pairs lived in different districts of KwaZulu-Natal province (912/1084) (Figure 4, Table 3). Three districts had no within-district genomic links (Amajuba, iLembe, and Sisonke) and eThekwini had the highest proportion of within-district links (17%). Proportions of within- and between-district links were similar when the SNP threshold was reduced to fewer than 3 SNPs and fewer than 1 SNP (Supplemental Table 2).

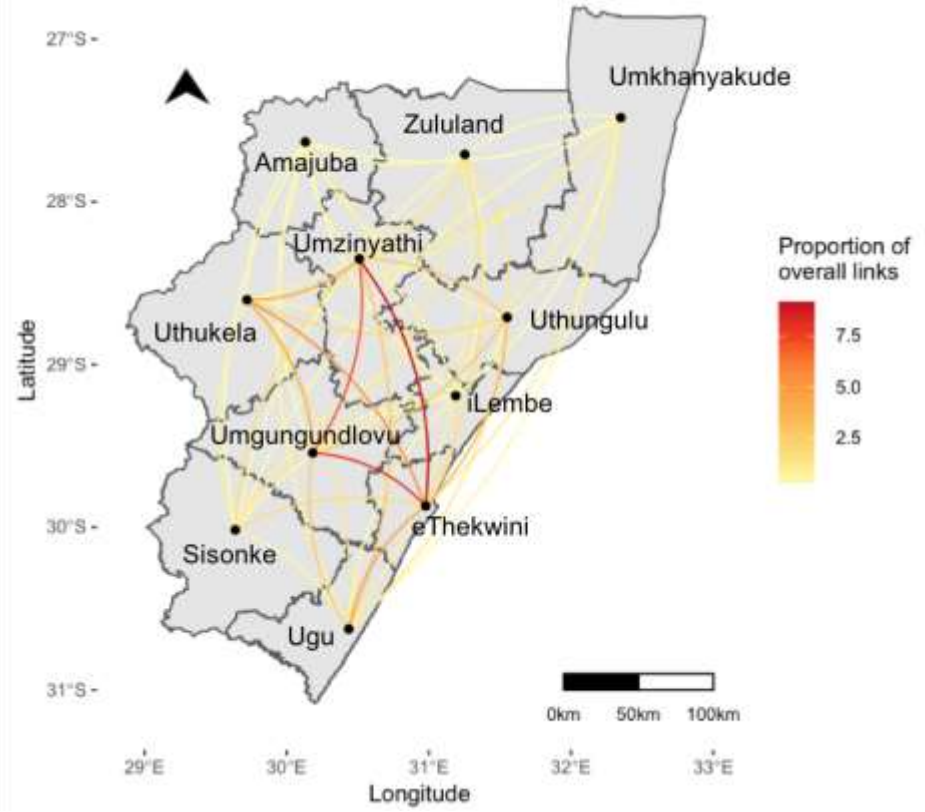
Approximately half (n=526, 49%) of all case-pairs were linked to the urban district of eThekwini. In every district except for two (Sisonke and Amajuba), the plurality of genomic links included a case that lived in eThekwini (Figure 4, Table 3, Supplemental Table 3). eThekwini district had the highest proportion (20%) of links with Umzinyathi.

At the individual case level, nearly a third of genomically linked cases (53, 29%) lived in the metropolitan district of eThekwini. Of note, 37 (70%) of these 53 cases were genomically linked to at least one other case within eThekwini, and nearly all (n=51, 96%) were genomically linked to at least one case outside of eThekwini. Among the 129 cases who lived outside of eThekwini, approximately half (n=59, 46%) had at least one genomic link within their home district. Nearly all (n=127, 96%) had at least one genomic link outside their home district, and 76 (61%) of those cases had at least one genomic link with a case in eThekwini.

Table 3. Proportions of within- and between-district genomic links (≤5 SNPs) in KwaZulu-Natal

District	Total links	Within-district links (%)	Between-district links (%)	Links with eThekwini (%)
Amajuba	1	0 (0)	1 (100)	0 (0)
eThekwini	526	91 (17)	435 (83)	--
iLembe	32	0 (0)	32 (100)	10 (31)
Sisonke	61	0 (0)	61 (100)	12 (20)
Ugu	236	12 (5)	224 (95)	75 (32)
UMgungundlovu	313	23 (7)	290 (93)	100 (32)
Umkhanyakude	97	1 (1)	96 (99)	25 (26)
Umzinyathi	334	32 (10)	302 (90)	104 (31)
Uthukela	160	7 (4)	153 (96)	45 (28)
Uthungulu	171	5 (3)	166 (97)	50 (29)
Zululand	65	1(2)	64 (99)	14 (22)

Figure 4. Genomic links (≤5 SNPs) within and between districts in KwaZulu-Natal



The proportion of genomic links occurring between each district out of the total number of links involving that district is represented by the color of the line. Amajuba district, which had only one genomic link, was excluded from this analysis.

Adjustment for differential enrollment by district

Enrollment fractions, based on the total number of diagnosed cases in each district, ranged from 0.22 in Sisonke and Amajuba to 0.50 in Umkhanyakude. Adjusting for enrollment, the proportion of within- and between-district links were 15% and 85%, respectively, which is nearly identical to the proportions in the unadjusted analysis. District-specific proportions of within- and between-district links were also similar to the unadjusted proportions (Supplemental Table 4). When we varied the proportion of cases enrolled in eThekwini relative to other districts (assuming enrollment was up to 40% higher in eThekwini than in other districts), eThekwini still accounted for the plurality of links in all but two districts.

Discussion

We aimed to define the spatial scale and identify geographic patterns of XDR-tuberculosis transmission in KwaZulu-Natal, South Africa. We found that genomically linked pairs of XDR-tuberculosis cases generally lived far apart, and that the majority (84%) of genomic links were between cases who lived in different districts. Nearly half of all genomically linked case-pairs involved a case in eThekwini district. Taken together, this evidence suggests that movement across districts, as well as into and out of eThekwini, may play a central role in the dissemination of XDR-tuberculosis across the province.

The median geographic distance between genomically liked cases was 108 km, which is remarkably high considering that tuberculosis cases with genetically similar strains have been found to be geographically clustered in other settings.[38, 39] We found similarly high geographic distances at more stringent thresholds of 3 and 1 SNP. Although there is no universal SNP threshold for defining a direct transmission link, there is general agreement that the threshold should be tailored to local tuberculosis epidemiology.[40, 41] Further, we also examined median distance by strain type, given that the genomic epidemiology of XDR-tuberculosis in KwaZulu-Natal is dominated by a single, highly clonal strain (LAM4).[42] The

median distance between genomically linked cases was similarly high among pairs of cases with the LAM4 strain and among non-LAM4 pairs. Although the LAM4 strain accounted for the majority of genomic links in our study, the phenomenon of the predominance of an individual clone is common in other settings with a high prevalence of drug-resistant tuberculosis.[43, 44].

The high proportions of between-district links and links with eThekwini suggest that cross-district movement, and perhaps eThekwini, plays a central role in patterns of XDR-tuberculosis transmission in KwaZulu-Natal. While previous studies have shown concentrations of tuberculosis cases in urban areas, suggesting that these settings are conducive to transmission, they have not examined the role of urban settings in driving transmission patterns and incidence in broader geographic areas.[13, 45] Although our convenience sample of XDR-tuberculosis cases diagnosed during the study period (n=404, 39%) does not provide a complete set of transmission links, we performed several analyses to assess whether our results are robust to potential selection bias. First, the demographic characteristics of TRAX cases were similar to all diagnosed cases in terms of age, sex, and the district of diagnosing facility. Second, our bias analysis showed that the proportions of between-district links and links with eThekwini remained high under scenarios of differential enrollment by district. Lastly, given that most cases of TB progress to active disease within two years of infection, it is likely that we captured the majority of relevant transmission links among TRAX cases, and that these links reflect larger transmission patterns in KwaZulu-Natal. [46]

Collectively, these findings provide insight into possible drivers of XDR-tuberculosis transmission in KwaZulu-Natal. Human movement and migration can transport pathogens across long distances, resulting in transmission that occurs far from an individual's home. Cyclical migration between rural and urban areas for employment is common in South Africa and in other rapidly developing countries, and effectively creates 'bridge' populations between urban and rural areas. This type of migration, which has previously been linked to HIV transmission, could also be driving tuberculosis transmission.[33] As such, it could explain both

the large distances between the homes of genomically linked cases and that cases were more likely to be linked to eThekwini district than to another case in their home district.

In addition to migration for employment, individuals may move between districts for other reasons. A previous analysis of TRAX participants showed that 36% of cases who were diagnosed with XDR-tuberculosis in eThekwini lived in a different district, indicating that travel from rural to urban areas for healthcare is common.[47] Importantly, travel to seek tuberculosis diagnosis and treatment is likely to coincide with an individual's infectious period, potentially providing abundant opportunities for transmission. Inter-district travel, be it for employment, healthcare, or other reasons, expands the geographic range of settings that are relevant for transmission. Indeed, almost a quarter of congregate locations reported by TRAX participants were outside of their home district, further suggesting that many locations that are potential settings of exposure or transmission may be distant from home.

There are several limitations to this study. Underdiagnosis of XDR-tuberculosis remains a challenge in resource-limited settings where insensitive diagnostic tools are commonly used and limited laboratory capacity curbs access to comprehensive drug susceptibility testing. As a result, transmission patterns observed among diagnosed cases provide only a limited characterization of province-wide patterns. In this study, however, we employed WGS to identify case-pairs with a high likelihood of transmission based on stringent SNP thresholds. The spatial scale we observed suggests an important role of migration, even if intermediate cases in the transmission chain were not diagnosed or enrolled in TRAX. Second, we captured participant's homes as only one location. In a setting like KwaZulu-Natal where migration is common, individuals may have multiple 'current' or recent residences, all of which may be possible locations of tuberculosis exposure and transmission. Thus, the 22% of cases that reported living in a different residence in the past five years may represent a lower bound on the proportion of cases that occupy multiple residences throughout the year. Future studies should aim to understand the role of cyclical migratory patterns and multiple residences in defining the

settings relevant for tuberculosis exposure and transmission. Lastly, 'mixed' infections, or genetically distinct populations within the same host, present potential challenges for inferring transmission based on a single *Mtb* isolate.[48] Yet, we do not expect mixed infections to be differential with respect to participants' homes, suggesting that our results are robust to the potential effects of within-host bacterial heterogeneity.

Evidence that the drug-resistant tuberculosis epidemic is increasingly attributable to transmission of drug-resistant strains has highlighted the importance of understanding transmission patterns in order to prevent incident cases.[5, 6, 49, 50] Despite the challenges of measuring transmission, the use of next-generation bacterial sequencing technologies brings us a step closer to understanding the settings and modes of contact sustaining tuberculosis transmission in high-burden settings. By defining the spatial scale of transmission, we provide preliminary data about transmission patterns and lay the foundation for further studies that more explicitly examine associations between casual contact in urban settings, migratory behavior, and the ongoing spread of XDR-tuberculosis. Ultimately, this knowledge can inform the development of tailored prevention strategies that target geographic areas and demographic groups that contribute disproportionately to transmission.

Acknowledgments*: We are grateful to the study team at the University of KwaZulu-Natal for their tireless efforts in data collection, record abstraction, participant recruitment, and interviews. We thank the participants and their families who consented to participate in this study.*

Footnotes

Conflicts of Interest: *The authors have no conflicts of interests to declare.*

Disclaimer: *The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the official position of the funding agencies. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.*

Financial Support: *This study was primarily funded by grants from the US National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH): R01AI089349 (PI Gandhi and R01AI087465 (PI Gandhi). It was also supported in part by NIH/NIAID grants: K23AI083088 (PI Brust), K24AI114444 (PI Gandhi), K23AI134182 (PI Auld), Emory CFAR P30AI050409 (PI Curran), Einstein CFAR P30AI051519 (PI Goldstein), by Einstein/Montefiore ICTR UL1 TR001073 (PI Shamoon).*

Meeting presentation: *Findings from this study have been presented at the Conference on Retroviruses and Opportunistic Infections (CROI) in February 2017 in Seattle, Washington (abstract number 660).*

Contact Information*: Correspondence and requests for reprints may be directed to Neel Gandhi (neel.r.gandh@emory.edu) at 1518 Clifton Road NE, CNR 3031, Atlanta, GA 30322.*

References

1. Global Tuberculosis Report. Geneva, Switzerland: World Health Organization, **2017**.

2. Pietersen E, Ignatius E, Streicher EM, et al. Long-term outcomes of patients with extensively drug-resistant tuberculosis in South Africa: a cohort study. Lancet (London, England) **2014**; 383:1230-9.

3. O'Donnell MR, Padayatchi N, Kvasnovsky C, Werner L, Master I, Horsburgh CR, Jr. Treatment outcomes for extensively drug-resistant tuberculosis and HIV co-infection. Emerg Infect Dis **2013**; 19:416-24.

4. Gandhi NR, Moll A, Sturm AW, et al. Extensively drug-resistant tuberculosis as a cause of death in patients co-infected with tuberculosis and HIV in a rural area of South Africa. Lancet (London, England) **2006**; 368:1575-80.

5. Shah NS, Auld SC, Brust JCM, et al. Transmission of Extensively Drug-Resistant Tuberculosis in South Africa. New England Journal of Medicine **2017**; 376:243-53.

6. Yang C, Luo T, Shen X, et al. Transmission of multidrug-resistant Mycobacterium tuberculosis in Shanghai, China: a retrospective observational study using whole-genome sequencing and epidemiological investigation. The Lancet Infectious Diseases **2017**; 17:275-84.

7. Middelkoop K, Mathema B, Myer L, et al. Transmission of tuberculosis in a South African community with a high prevalence of HIV infection. The Journal of infectious diseases **2015**; 211:53-61.

8. Verver S, Warren RM, Munch Z, et al. Proportion of tuberculosis transmission that takes place in households in a high-incidence area. Lancet (London, England) **2004**; 363:212-4.

9. Crampin AC, Glynn JR, Traore H, et al. Tuberculosis transmission attributable to close contacts and HIV status, Malawi. Emerging infectious diseases **2006**; 12:729-35.

10. Glynn JR, Guerra-Assuncao JA, Houben RMGJ, et al. Whole Genome Sequencing Shows a Low Proportion of Tuberculosis Disease Is Attributable to Known Close Contacts in Rural Malawi. PloS one **2015**; 10:e0132840.

11. Martinez L, Shen Y, Mupere E, Kizza A, Hill PC, Whalen CC. Transmission of Mycobacterium Tuberculosis in Households and the Community: A Systematic Review and Meta-Analysis. Am J Epidemiol **2017**; 185:1327-39.

12. Johnstone-Robertson SP, Mark D, Morrow C, et al. Social mixing patterns within a South African township community: implications for respiratory disease transmission and control. American journal of epidemiology **2011**; 174:1246-55.

13. Andrews JR, Morrow C, Walensky RP, Wood R. Integrating social contact and environmental data in evaluating tuberculosis transmission in a South African township. The Journal of infectious diseases **2014**; 210:597-603.

14. Andrews JR, Morrow C, Wood R. Modeling the Role of Public Transportation in Sustaining Tuberculosis Transmission in South Africa. American Journal of Epidemiology **2013**; 177:556-61.

15. Bryant JM, Harris SR, Parkhill J, et al. Whole-genome sequencing to establish relapse or re-infection with Mycobacterium tuberculosis: a retrospective observational study. The Lancet Respiratory medicine **2013**; 1:786-92.

16. Gardy JL, Johnston JC, Ho Sui SJ, et al. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. The New England journal of medicine **2011**; 364:730-9.

17. Kato-Maeda M, Ho C, Passarelli B, et al. Use of whole genome sequencing to determine the microevolution of Mycobacterium tuberculosis during an outbreak. PloS one **2013**; 8:e58235.

18. Nikolayevskyy V, Kranzer K, Niemann S, Drobniewski F. Whole genome sequencing of Mycobacterium tuberculosis for detection of recent transmission and tracing outbreaks: A systematic review. Tuberculosis (Edinburgh, Scotland) **2016**; 98:77-85.

19. Pérez-Lago L, Comas I, Navarro Y, et al. Whole genome sequencing analysis of intrapatient microevolution in Mycobacterium tuberculosis: potential impact on the inference of tuberculosis transmission. The Journal of infectious diseases **2014**; 209:98-108.

20. Walker TM, Ip CL, Harrell RH, et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. The Lancet Infectious Diseases **2013**; 13:137-46.

21. Lessler J, Salje H, Grabowski MK, Cummings DAT. Measuring Spatial Dependence for Infectious Disease Epidemiology. PloS one **2016**; 11:e0155249.

22. Riley S. Large-Scale Spatial-Transmission Models of Infectious Disease. Science **2007**; 316.

23. Ndjeka N. Strategic Overview of MDR-TB Care in South Africa.

24. Lim JR, Gandhi NR, Mthiyane T, et al. Incidence and Geographic Distribution of Extensively Drug-Resistant Tuberculosis in KwaZulu-Natal Province, South Africa. PloS one **2015**; 10:e0132076.

25. Shisana O RT, Simbayi L, Zuma KK, Jooste S, Zungu NP, Labadarios D, Onoya D, Wabiri N. South African National HIV Prevalence, Incidence, and Behaviour Survey, 2012.

26. Census 2011: Census in Brief. Pretoria, South Africa: Statistics South Africa, **2011**.

27. Camlin CS, Hosegood V, Newell M-L, McGrath N, Bärnighausen T, Snow RC. Gender, Migration and HIV in Rural KwaZulu-Natal, South Africa. PloS one **2010**; 5:e11539.

28. Eldholm V, Monteserin J, Rieux A, et al. Four decades of transmission of a multidrug-resistant Mycobacterium tuberculosis outbreak strain. **2015**; 6:7119.

29. Guerra-Assunção JA, Crampin AC, Houben RMGJ, et al. Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a high prevalence area. eLife **2015**; 4.

30. Casali N, Broda A, Harris SR, et al. Whole Genome Sequence Analysis of a Large Isoniazid-Resistant Tuberculosis Outbreak in London: A Retrospective Observational Study. PLoS medicine **2016**; 13:e1002137.

31. Pebesma EJ BR. Classes and methods for spatial data in R. R News **2005**; 5.

32. Hijmans R. geosphere: Spherical Trigonometry. R package version 1.5-5 ed, **2016**.

33. Lurie M, Harrison A, Wilkinson D, Karim SA. Circular migration and sexual networking in rural KwaZulu/Natal: implications for the spread of HIV and other sexually transmitted diseases. Health Transition Review **1997**; 7:17-27.

34. Huang CC, Tchetgen ET, Becerra MC, et al. The effect of HIV-related immunosuppression on the risk of tuberculosis transmission to household contacts. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2014**; 58:765-74.

35. Espinal MA, Peréz EN, Baéz J, et al. Infectiousness of Mycobacterium tuberculosis in HIV-1-infected patients with tuberculosis: a prospective study. Lancet (London, England) **2000**; 355:275-80.

36. Sullivan BJ, Esmaili BE, Cunningham CK. Barriers to initiating tuberculosis treatment in sub-Saharan Africa: a systematic review focused on children and youth. Global Health Action **2017**; 10:1290317.

37. South African Demographic and Health Survey 2003. Pretoria: Department of Health, Medical Research Council, **2007**.

38. Ribeiro FK, Pan W, Bertolde A, et al. Genotypic and Spatial Analysis of Mycobacterium tuberculosis Transmission in a High-Incidence Urban Setting. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2015**; 61:758-66.

39. Zelner JL, Murray MB, Becerra MC, et al. Identifying Hotspots of Multidrug-Resistant Tuberculosis Transmission Using Spatial and Molecular Genetic Data. The Journal of infectious diseases **2016**; 213:287-94.

40. Bryant JM, Schürch AC, van Deutekom H, et al. Inferring patient to patient transmission of Mycobacterium tuberculosis from whole genome sequencing data. BMC infectious diseases **2013**; 13:110.

41. Hatherell H-A, 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.

42. Gandhi NR, Brust JCM, Moodley P, et al. Minimal diversity of drug-resistant Mycobacterium tuberculosis strains, South Africa. Emerging infectious diseases **2014**; 20:426-33.

43. Chihota VN, Müller B, Mlambo CK, et al. Population Structure of Multi- and Extensively Drug-Resistant Mycobacterium tuberculosis Strains in South Africa. Journal of Clinical Microbiology **2012**; 50:995-1002.

44. Brown TS, Narechania A, Walker JR, et al. Genomic epidemiology of Lineage 4 Mycobacterium tuberculosis subpopulations in New York city and New Jersey, 1999–2009. BMC Genomics **2016**; 17:947.

45. Alene KA, Viney K, McBryde ES, et al. Spatial patterns of multidrug resistant tuberculosis and relationships to socio-economic, demographic and household factors in northwest Ethiopia. PloS one **2017**; 12:e0171800.

46. Targeted tuberculin testing and treatment of latent tuberculosis infection. This official statement of the American Thoracic Society was adopted by the ATS Board of Directors, July 1999. This is a Joint Statement of the American Thoracic Society (ATS) and the Centers for Disease Control and Prevention (CDC). This statement was endorsed by the Council of the Infectious Diseases Society of America. (IDSA), September 1999, and the sections of this statement. American journal of respiratory and critical care medicine **2000**; 161:S221-47.

47. Kapwata T, Morris N, Gandhi N, et al. Spatial distribution of extensively drug-resistant tuberculosis (XDR-TB) patients in KwaZulu-Natal, South Africa. bioRxiv **2017**.

48. Cohen T, van Helden PD, Wilson D, et al. Mixed-Strain Mycobacterium tuberculosis Infections and the Implications for Tuberculosis Treatment and Control. Clinical Microbiology Reviews **2012**; 25:708-19.

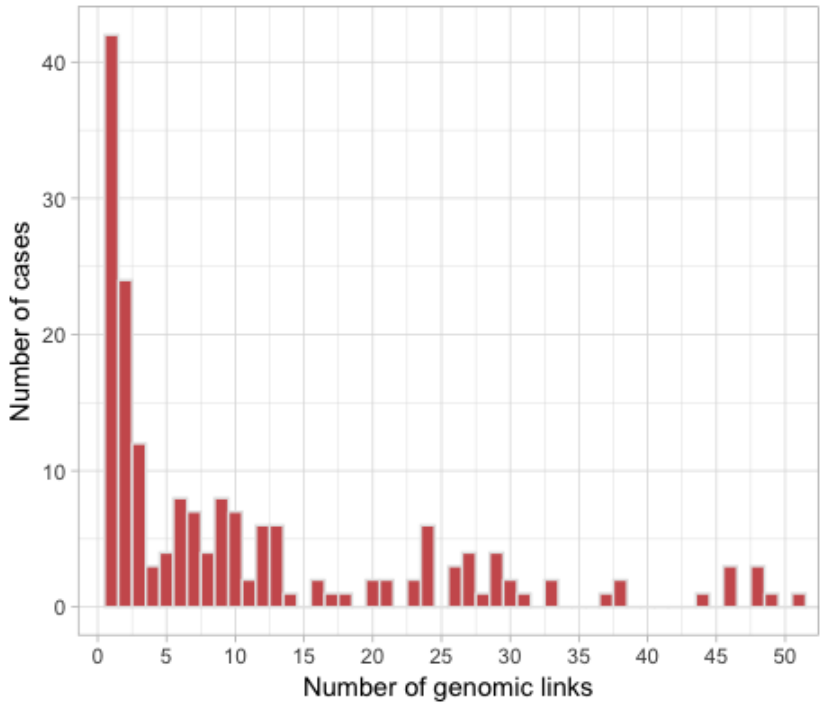
49. Becerra MC, Appleton SC, Franke MF, et al. Tuberculosis burden in households of patients with multidrug-resistant and extensively drug-resistant tuberculosis: a retrospective cohort study. Lancet (London, England) **2011**; 377:147-52.

50. Devaux I, Kremer K, Heersma H, Van Soolingen D. Clusters of Multidrug-Resistant Mycobacterium tuberculosis Cases, Europe. Emerging Infectious Diseases **2009**; 15:1052-60.

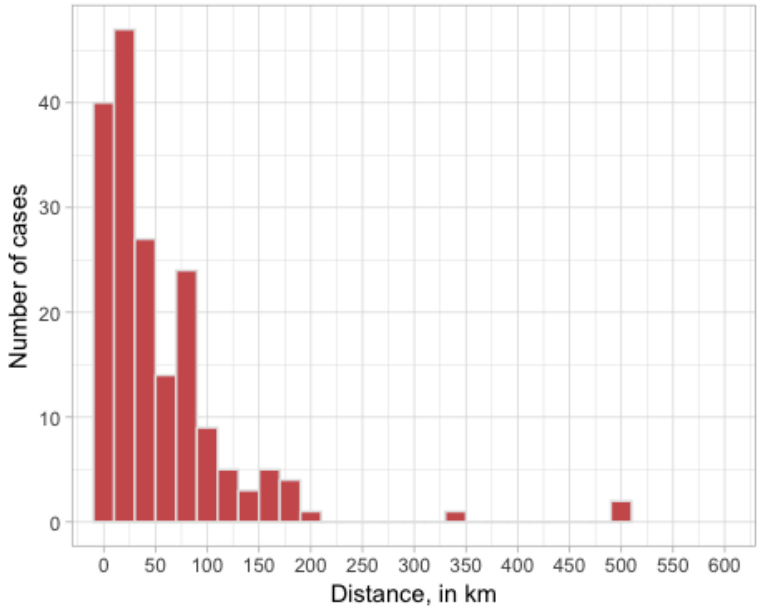
Supplemental Table 1. Genomic links (≤5 SNPs) by sex, HIV status, and strain type

Pair	Total links (% of total)	Median distance, in km (IQR)
Sex		
Female / Female	382 (35)	96 (49 – 150)
Female / Male	534 (49)	107 (69 – 165)
Male / Male	168 (15)	131 (84 – 150)
HIV status		
HIV+ / HIV+	654 (60)	104 (58 – 155)
HIV+ / HIV-	377 (35)	117 (70 – 166)
HIV- / HIV-	53 (5)	136 (80 – 197)
Strain type		
LAM4 / LAM4	1075 (99)	127 (68 – 147)
Non-LAM4 / Non-LAM4	9 (1)	108 (64 – 162)

Supplemental Figure 1. Number of genomic links (≤5 SNPs) per case



Supplemental Figure 2. Shortest geographic link among genomic links (≤5 SNPs) for each case



Supplemental Table 2. Within and between-district links at ≤3 and ≤1 SNPs

	≤ 5 SNPs				≤ 3 SNPs				≤ 1 SNP			
District	Total links	Within-district (%)	Between-district (%)	With eThekwini (%)	Total links	Within-district (%)	Between-district (%)	With eThekwini (%)	Total links	Within-district (%)	Between-district (%)	With eThekwini (%)
Amajuba	1	0 (0)	1 (100)	0 (0)	0	-	-	-	0	-	-	-
eThekwini	526	91 (17)	435 (83)	--	115	24 (21)	91 (79)	--	11	4 (36)	7 (64)	--
iLembe	32	0 (0)	32 (100)	10 (31)	10	(0)	10 (100)	3 (30)	0	-	-	-
Sisonke	61	0 (0)	61 (100)	12 (20)	15	(0)	15 (100)	4 (27)	0	-	-	-
Ugu	236	12 (5)	224 (95)	75 (32)	74	50 (7)	24 (93)	25 (34)	4	0 (0)	4 (100)	1 (25)
UMgungundlovu	313	23 (7)	290 (93)	100 (32)	47	1 (2)	46 (98)	9 (19)	2	0 (0)	2 (100)	1 (50)
Umkhanyakude	97	1 (1)	96 (99)	25 (26)	20	1 (5)	19 (95)	6 (30)	2	0 (0)	2 (100)	1 (50)
Umzinyathi	334	32 (10)	302 (90)	104 (31)	64	50 (8)	14 (92)	21 (33)	6	0 (0)	6 (100)	2 (33)
Uthukela	160	7 (4)	153 (96)	45 (28)	37	3 (8)	34 (92)	9 (24)	2	0 (0)	2 (100)	0 (0)
Uthungulu	171	5 (3)	166 (97)	50 (29)	39	1 (3)	38 (97)	12 (31)	5	0 (0)	5 (100)	2 (40)
Zululand	65	1(2)	64 (99)	14 (22)	18	1 (6)	17 (94)	2 (11)	0	-	-	-

Supplemental Table 3. Within and between-district genomic links (≤5 SNPs)

	Amajuba	eThekwini	iLembe	Sisonke	Ugu	UMgungundlovu	Umkhanyakude	Umzinyathi	Uthukela	Uthungulu	Zululand	Total number of links involving each district	Within-district links (%)	Between-district links (%)	Links with eThekwini (%)
Amajuba	0	0	0	0	0	0	0	0	0	1	0	1	0 (0)	1 (100)	0 (0)
eThekwini		91	10	12	75	100	25	104	45	50	14	526	91 (17)	435 (83)	--
iLembe			0	2	3	3	2	6	2	4	0	32	0 (0)	32 (100)	10 (31)
Sisonke				0	6	9	4	15	6	5	2	61	0 (0)	61 (100)	12 (20)
Ugu					12	45	13	39	16	17	10	236	12 (5)	224 (95)	75 (32)
UMgungundlovu						23	16	54	27	27	9	313	23 (7)	290 (93)	100 (32)
Umkhanyakude							1	10	11	11	4	97	1 (1)	96 (99)	25 (26)
Umzinyathi								32	32	29	13	334	32 (10)	302 (90)	104 (31)
Uthukela									7	12	2	160	7 (4)	153 (96)	45 (28)
Uthungulu										5	10	171	5 (3)	166 (97)	50 (29)
Zululand											1	65	1(2)	64 (99)	14 (22)

Supplemental Table 4. Enrollment fraction bias analysis of between- and within-district genomic links

District	Sampling fraction	Inverse sampling fraction	Total number of links involving each district	Within-district links (%)	Between-district links (%)	Links with eThekwini (%)
Amajuba	0.22	4.50	24.2	2.3 (9.3)	22.0 (90.7)	1.8 (7.3)
eThekwini	0.39	2.54	1377.9	230.7 (16.7)	1147.2 (83.3)	230.7 (16.7)
iLembe	0.28	3.57	106.0	1.8 (1.7)	104.2 (98.3)	30.5 (28.8)
Sisonke	0.22	4.50	223.6	2.3 (1.0)	221.3 (99.0)	42.2 (18.9)
Ugu	0.42	2.37	598.0	2.3 (4.8)	595.8 (95.3)	183.3 (30.7)
UMgungundlovu	0.37	2.68	839.4	28.4 (7.3)	811.0 (92.7)	260.6 (31.0)
Umkhanyakude	0.50	2.00	232.1	61.6 (0.9)	170.5 (99.1)	56.7 (24.4)
Umzinyathi	0.36	2.80	915.6	2.0 (9.8)	913.6 (90.2)	277.3 (30.3)
Uthukela	0.30	3.38	554.6	89.5 (4.3)	465.1 (95.7)	133.0 (24.0)
Uthungulu	0.40	2.48	499.6	23.6 (2.5)	476.0 (97.5)	125.4 (25.1)
Zululand	0.38	2.61	206.3	12.4 (1.3)	193.9 (98.7)	36.0 (17.4)

Total inflated links = 3017.2