Pre-detection history of extensively drug-resistant tuberculosis in KwaZulu-Natal, South Africa

Tyler S. Brown[a,b], Lavanya Challagundla[c], Evan H. Baugh[d], Shaheed Vally Omar[e], Arkady Mustaev[f], Sara C. Auld[g,h], N. Sarita Shah[h,i], Barry N. Kreiswirth[f], James C. M. Brust[j], Kristin N. Nelson[h], Apurva Narechania[k], Natalia Kurepina[f], Koleka Mlisana[l,m], Richard Bonneau[n], Vegard Eldholm[o], Nazir Ismail[e,p,q], Sergios-Orestis Kolokotronis[k,r], D. Ashley Robinson[c], Neel R. Gandhi[h,s], and Barun Mathema[a,1]

[a]Department of Epidemiology, Mailman School of Public Health, Columbia University, New York, NY 10032; [b]Infectious Diseases Division, Massachusetts General Hospital, Boston, MA 02114; [c]Department of Microbiology and Immunology, University of Mississippi Medical Center, Jackson, MS 39216; [d]Institute for Genomic Medicine, Columbia University, New York, NY 10032; [e]Centre for Tuberculosis, National Institute for Communicable Diseases, Johannesburg 2131, South Africa; [f]Public Health Research Institute of New Jersey Medical School, Rutgers University, Newark, NJ 07103; [g]Division of Pulmonary, Allergy, Critical Care and Sleep Medicine, Emory University School of Medicine, Atlanta, GA 30322; [h]Department of Epidemiology, Rollins School of Public Health, Emory University, Atlanta, GA 30322; [i]Global Tuberculosis Branch, Division of Global HIV and Tuberculosis, Center for Global Health, Centers for Disease Control and Prevention, Atlanta, GA 30329; [j]Divisions of General Internal Medicine and Infectious Diseases, Department of Medicine, Albert Einstein College of Medicine, Bronx, NY 10467; [k]Institute for Comparative Genomics, American Museum of Natural History, New York, NY 10024; [l]Department of Medical Microbiology, School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal, Durban 4001, South Africa; [m]Academic Affairs, Research and Quality Assurance Office, National Health Laboratory Service, Johannesburg 2192, South Africa; [n]Center for Computational Biology, Flatiron Institute, New York, NY; [o]Division of Infection Control and Environmental Health, Norwegian Institute of Public Health, 0456 Oslo, Norway; [p]Department of Medical Microbiology, Faculty of Health Sciences, University of Pretoria, Pretoria 0084, South Africa; [q]Department of Internal Medicine, University of Witwatersrand, Johannesburg 2000, South Africa; [r]Department of Epidemiology and Biostatistics, School of Public Health, SUNY Downstate Health Sciences University, Brooklyn, NY 11203; and [s]Division of Infectious Diseases, Department of Medicine, Emory School of Medicine, Emory University, Atlanta, GA 30322

Edited by Erwin Schurr, McGill University, Montreal, QC, Canada, and accepted by Editorial Board Member Carl F. Nathan October 3, 2019 (received for review April 17, 2019)

Antimicrobial-resistant (AMR) infections pose a major threat to global public health. Similar to other AMR pathogens, both historical and ongoing drug-resistant tuberculosis (TB) epidemics are characterized by transmission of a limited number of predominant *Mycobacterium tuberculosis* (*Mtb*) strains. Understanding how these predominant strains achieve sustained transmission, particularly during the critical period before they are detected via clinical or public health surveillance, can inform strategies for prevention and containment. In this study, we employ whole-genome sequence (WGS) data from TB clinical isolates collected in KwaZulu-Natal, South Africa to examine the pre-detection history of a successful strain of extensively drug-resistant (XDR) TB known as LAM4/KZN, first identified in a widely reported cluster of cases in 2005. We identify marked expansion of this strain concurrent with the onset of the generalized HIV epidemic 12 y prior to 2005, localize its geographic origin to a location in northeastern KwaZulu-Natal ∼400 km away from the site of the 2005 outbreak, and use protein structural modeling to propose a mechanism for how strain-specific *rpoB* mutations offset fitness costs associated with rifampin resistance in LAM4/KZN. Our findings highlight the importance of HIV coinfection, high preexisting rates of drug-resistant TB, human migration, and pathoadaptive evolution in the emergence and dispersal of this critical public health threat. We propose that integrating whole-genome sequencing into routine public health surveillance can enable the early detection and local containment of AMR pathogens before they achieve widespread dispersal.

infectious disease | epidemics | tuberculosis | antimicrobial resistance | population genetics

Prevention of antimicrobial-resistant (AMR) infections is a central strategic priority for global public health. Epidemics of AMR pathogens are typically characterized by transmission of a limited number of successful strains distributed over distinct geographic territories. Examples include carbapenem-resistant *Klebsiella pneumoniae* ST258 (1) and methicillin-resistant *Staphylococcus aureus* USA300 in the United States (2), and the multiple epidemic strains of drug-resistant *Mycobacterium tuberculosis* (*Mtb*) that have originated in South America (3), the former Soviet Union (4), and the United States (5). Recent studies have implicated a single successful epidemic strain in the majority of extensively drug-resistant tuberculosis (XDR-TB; defined as TB disease involving resistance to at least 4 first- and second-line antibiotics) cases in KwaZulu-Natal, South Africa (6, 7). Understanding the mechanisms through which certain AMR pathogens achieve sustained, widespread transmission remains an active and important area of scientific inquiry.

Multiple lines of evidence suggest that the fitness cost of drug-resistance mutations, namely the predicted decrease in transmissibility in drug-resistant pathogens versus their drug-susceptible counterparts, strongly influences the epidemiological success of

Significance

Epidemics of AMR pathogens are often only identified years or decades after they first evolved and distant from their place of origin. Consequently, evidence-based strategies for early containment of AMR epidemics are limited. This study employs whole-genome sequence data to reconstruct the "pre-detection" evolutionary and epidemiological history of an extensively drug-resistant *Mycobacterium tuberculosis* strain in KwaZulu-Natal, South Africa. We localize the geographic origin of this strain to an area hundreds of kilometers away from where the first clinical cases were reported and identify key host- and pathogen-specific factors that contributed to the rise of this important threat to global tuberculosis control. We propose that similar strategies can support the early identification and containment of AMR pathogens in the future.

Author contributions: T.S.B. and B.M. designed research; T.S.B., L.C., E.H.B., S.V.O., A.M., S.C.A., N.S.S., J.C.M.B., K.N.N., A.N., N.K., K.M., N.I., S.-O.K., D.A.R., and B.M. performed research; T.S.B., L.C., E.H.B., A.M., S.C.A., N.S.S., K.N.N., A.N., N.K., R.B., V.E., S.-O.K., D.A.R., and B.M. analyzed data; and T.S.B., S.-O.K., D.A.R., N.R.G., and B.M. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. E.S. is a guest editor invited by the Editorial Board.

Data deposition: Sequencing data reported in this paper have been deposited in the National Center for Biotechnology Information (NCBI) Sequence Read Archive (BioProject accession no. PRJNA476470).

[1]To whom correspondence may be addressed. Email: bm2055@cumc.columbia.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1906636116/-/DCSupplemental.

First published October 28, 2019.

drug-resistant pathogens (8, 9). Pathogen-specific adaptations that offset the fitness costs of drug resistance have been studied extensively (10–14). However, these pathogen-specific adaptations are often not exclusive to a predominant epidemic strain (15–19), suggesting that additional mechanisms (including the natural stochasticity of infectious disease transmission) likely influence the process through which certain drug-resistant strains become more prevalent than others. Local environmental or ecological factors, including differences between local host populations or contact networks, may also help explain the distinct geographic specificity of most epidemic strains (20, 21).

Compared with the expansive body of research on adaptive evolution in AMR pathogens, relatively less is known about these potential extrinsic drivers of epidemiological success. This is in part because many of the key processes in the emergence of drug-resistant pathogens take place years or decades before they are first noticed as public health threats (the so-called pre-detection period). In addition, many drug-resistant pathogens are detected only after they have dispersed into a larger area away from their geographic origin, confounding attempts to identify local environmental or host-related factors driving emergence. Addressing these knowledge gaps, and better understanding this important pre-detection period, can help inform strategies for early identification and local containment of AMR pathogens.

In this report, we employ whole-genome sequence (WGS) data to reconstruct the joint evolutionary and epidemiological histories of a highly successful strain of XDR-TB in KwaZulu-Natal, South Africa (KZN). This strain, known as LAM4/KZN, was first identified during an outbreak of TB cases among cohospitalized HIV patients in Tugela Ferry, KwaZulu-Natal in 2005 (22), and has since become the predominant cause of XDR-TB in the province. Leveraging a large cross-sectional collection of clinical XDR-TB isolates, including a subset from the 2005 Tugela Ferry outbreak, and detailed patient geographic information, we localize the origin of LAM4/KZN to an outlying district in northeastern KwaZulu-Natal, distant from Tugela Ferry. We identify local environmental factors in this area, including high existing rates of drug-resistant TB and HIV coinfection, as important contributing factors in the emergence of this important strain. We use Bayesian phylogenetic inference to estimate when widespread transmission of LAM4/KZN began, implicating population-level modification of host susceptibility due to syndemic HIV infection as a key driver in this epidemic. In addition, we employ protein structural modeling to propose a mechanism for how strain-specific *rpoB* mutations offset fitness costs associated with rifampin resistance in LAM4/KZN, underscoring the multifactorial etiology of this important epidemic and AMR epidemics in general. We propose that routine genomic surveillance, enabled by the rapidly decreasing cost of pathogen sequencing, can provide a powerful adjunctive strategy for public health practitioners, specifically as a tool enabling the early identification and local containment of emerging AMR pathogens.

Results

XDR-TB in KwaZulu-Natal Is Diverse but Dominated by a Single Epidemic Strain. To understand the population structure of XDR-TB in KwaZulu-Natal, we first analyzed WGS data for 318 bacteriologically confirmed XDR-TB clinical isolates collected in the province between 2011 and 2014 (6, 23). XDR-TB strains from KwaZulu-Natal were genetically diverse and included isolates from multiple distantly related phylogenetic lineages. The majority of XDR-TB isolates (250 of 318, 79%) belonged to the genetically monomorphic LAM4/KZN clade (mean within-clade pairwise single-nucleotide polymorphism [SNP] difference: 21.07 SNPs; range: 0 to 104 SNPs) (Fig. 1), which is also identified as the 4.3.3 sublineage in the SNP-based classification system described by Coll et al. (24). Terminal branch lengths were significantly shorter among LAM4/KZN isolates (n = 248; mean terminal branch length: 6.93) when compared with non–LAM4/KZN strains (n = 70; mean terminal branch length: 39.43) by 2-sample Kolmogorov–Smirnov test (P = 1.997E-13) and a modified version of this test that accounts for differences in sample sizes (*SI Appendix*, Fig. S1). In this context, shorter terminal branch lengths are consistent with higher relative rates of transmission among LAM4/KZN isolates versus other XDR-TB strains in KwaZulu-Natal. Although mutation-rate heterogeneity between *Mtb* lineages has been reported, this variation is small (25) and not expected to measurably change the number of SNP differences accrued during the relatively short time since divergence in terminal branches. Isolated long terminal branches, indicative of genetically distinct isolates within subclades, were absent in LAM4/KZN (Fig. 1*A*), consistent with the highly clonal nature of this strain.

Seventy-eight percent of XDR-TB isolates were obtained from patients with HIV coinfection, and 31% of isolates were from patients with a history of prior treatment for multidrug-resistant (MDR) TB. We find no evidence of population differentiation between HIV-seronegative and HIV-seropositive cases (Weir and Cockerham's F_{ST} = 0.0043, P = 0.1065) or patients with prior history of MDR-TB and those with no prior history of MDR-TB (F_{ST} = −0.0028, P = 0.7997). These findings suggest that ongoing transmission of LAM4/KZN likely occurs in networks that include both HIV-seropositive and -seronegative patients, and among patients with and without prior history of MDR-TB infection.

Phylogenetic Reconstruction Supports Early Emergence of LAM4/KZN. To compare the timescale over which different XDR-TB lineages evolved, we compared the estimated time to the most recent common ancestor (TMRCA) for drug resistance-associated mutations in *rpoB* (a characteristic mutation in MDR-TB that mediates resistance to rifampin) and *gyrA* (a characteristic mutation in XDR-TB that mediates resistance to fluoroquinolones) across different XDR-TB sublineages on the phylogenetic tree (Fig. 1*B*). *rpoB* mutations arose before *gyrA* mutations in each of the predominant XDR-TB strains across all sublineages, consistent with the introduction of rifampin for clinical use ∼20 y before fluoroquinolones. *gyrA* mutations were all relatively recently acquired, with TMRCA for resistance-associated mutations observed in the 2.2.1 clade (*gyrA* D94G; median: 2008; 95% highest posterior density [HPD]: 2001 to 2011) and the oldest divergence among those from the 4.8 clade (*gyrA* D94N; median: 1982; 95% HPD: 1972 to 1989).

To date the stepwise acquisition of drug-resistance mutations in LAM4/KZN, we next reconstructed a dated phylogenetic tree using the LAM4/KZN subsample described in *Methods*. We estimated that LAM4/KZN acquired mutations that define the MDR-TB phenotype (*katG* S315T, which confers resistance to isoniazid, and *rpoB* L452P, which confers resistance to rifampin) in ∼1961 and 1983, respectively. Mutations required for the complete XDR phenotype, *gyrA* A90V and *rrs* a1401g (which confer resistance to fluoroquinolones and kanamycin, respectively) and 2 additional *rpoB* mutations (D435G and I1106T) were acquired approximately 10 y later (TMRCA: 1993; 95% HPD: 1988 to 1998; *SI Appendix*, Fig. S2), still well before the 2005 Tugela Ferry outbreak that first focused public attention on this epidemic. Multiple *gyrA* mutations are found among LAM4 *rpoB* single mutants (i.e., those carrying L452P but not D435G and I1106T), indicating that fluoroquinolone resistance emerged multiple times in MDR LAM4 isolates sharing a recent common ancestor with the *rpoB* triple-mutant clone. The order in which LAM4/KZN acquired drug-resistance mutations is consistent with both treatment guidelines (which outline that patients receive a fluoroquinolone only after they are found to have rifampin-resistant TB) and the historical timeline of when each corresponding antibiotic was introduced for clinical use (7). A nonsynonymous mutation in *drrA* (R262G) and a mutation in the Rv1144-mmpL13a

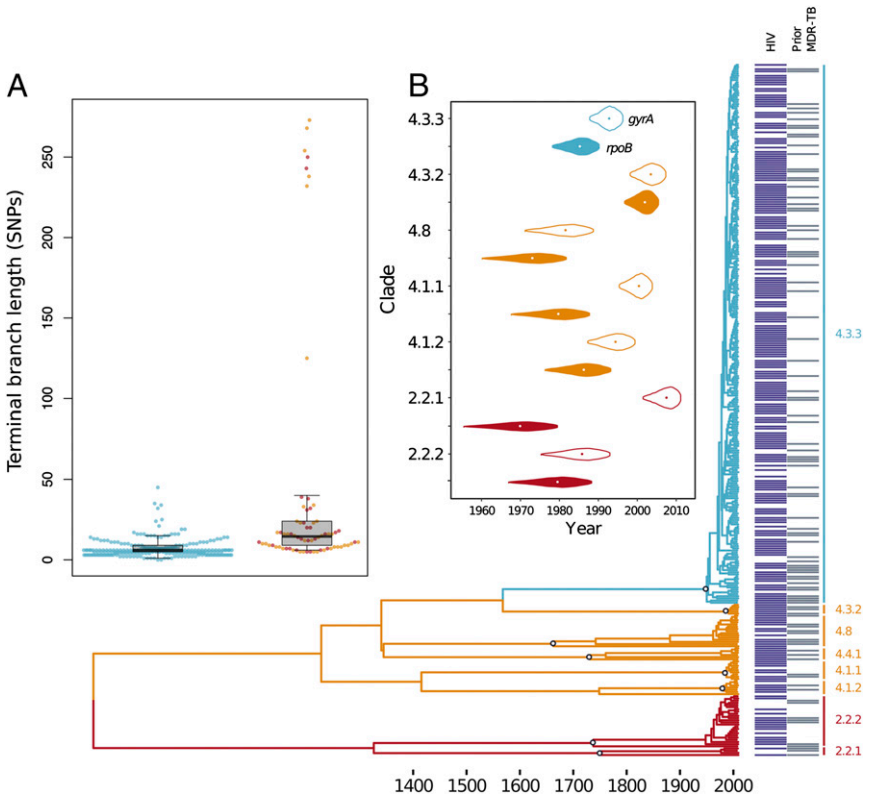

Fig. 1. Bayesian phylogenetic reconstruction for 318 XDR-TB isolates from KZN. Clades are labeled using previously described SNP-based classification for *Mtb* per Coll et al. (24) and annotated with HIV serostatus and history of prior MDR-TB (turquoise, LAM4/KZN/4.3.3; orange, non–LAM4/KZN lineage 4; red, lineage 2). (*A*) Terminal branch lengths for a BEAST-estimated phylogenetic tree by clade group. (*B*) TMRCA for drug-resistant clone groups for each clade (open violin plots, *gyrA* mutants; closed violin plots, *rpoB* mutants). All nodes corresponding to SNP-based clade labels (open circles) have posterior probability greater than 0.99.

intergenic region (–102C>A), both associated with XDR-TB in prior genome-wide association studies (26), were identified as co-occurring with *rpoB* L452P and the L452P/D435G/I1106T triple mutation, respectively. LAM4/KZN isolates also carry a clade-defining nonsynonymous mutation in *Rv2000* (L275P), an *Mtb* gene of unknown function (27) (*SI Appendix*, Fig. S3).

rpoB D435G and I1106T are rare mutations across *Mtb* populations globally and the L452P/D435G/I1106T triple mutation is synapomorphic (i.e., clade-defining) to LAM4/KZN. L452P and D435G are located adjacent to the RNA polymerase (RNAP)β active site (in the so-called rifampin resistance-determining region; RRDR). Protein structural modeling indicates that D435G alters the distribution of charges within the RNAP active site, both in isolation and in the presence of L452P (*SI Appendix*, Figs. S4 and S5), which may have an impact on polymerase activity or transcriptional targets (28). We did not identify any known putative *rpoC* or *rpoA* compensatory mutations among LAM4/KZN isolates, although *rpoC* mutations were common among non–KZN/LAM4 isolates (*SI Appendix*, Fig. S3). However, I1106T is located outside the RRDR but within the RNAPβ–RNAPβ′ (*rpoC*) binding interface, adjacent to the location of multiple previously reported RNAPβ′ compensatory mutations (*SI Appendix*, Fig. S6) (29).

LAM4/KZN Bacterial Populations Increased in Size Concurrent with the Onset of the HIV Generalized Epidemic. We used Bayesian skyline analysis to estimate the historical size of the LAM4/KZN population and thereby infer changes in transmission of this strain over time. We observed evidence of expansion in the effective population size (N_e) of this strain starting between 1995 and 2003 (Fig. 2). Comparing the posterior distribution of N_e estimates at different time points in the Bayesian skyline plot, we find statistically significant increases in N_e between 1999 and 2010 in the analysis including nonsynonymous sites (Fig. 2*A*) and between 1998 and 2010 in the analysis excluding nonsynonymous sites (Fig. 2*B*). This period corresponds to the onset of the generalized HIV epidemic and the increase in overall TB case notifications that occurred in the mid-1990s (30–32). This period also predates the first reported LAM4/KZN outbreak in Tugela Ferry by up to 10 y (22). Bayesian estimates of bacterial population size can be biased by several factors, including positive selection (33), and we thus undertook 2 additional analyses to examine population-size history in LAM4/KZN: 1) Model comparison via marginal-likelihood estimation favors the Bayesian skyline model and logistical population growth over constant population size, although not significantly so. In the exponential growth model, the posterior distribution of the population growth rate did not include 0, rejecting a constant population-size model (*SI Appendix*, Table S1). 2) We calculated genome-wide neutrality statistics based on the site-frequency spectrum of mutations in LAM4/KZN isolates, which demonstrated statistically significant departures from the expected-site frequency spectrum for a constant-sized population under no selective pressure. Taken together, neutrality statistics for the LAM4/KZN isolates are most consistent with population expansion occurring after a selective sweep (*SI Appendix*, Table S2).

Population-Genetic Signatures Localize the Most Likely Geographic Origin of LAM4/KZN. To investigate the geographic origin of LAM4/KZN within KwaZulu-Natal, we first used 244 LAM4/KZN isolates with associated GPS data (for each participant's residence) to test for geospatial genetic structure. When isolates

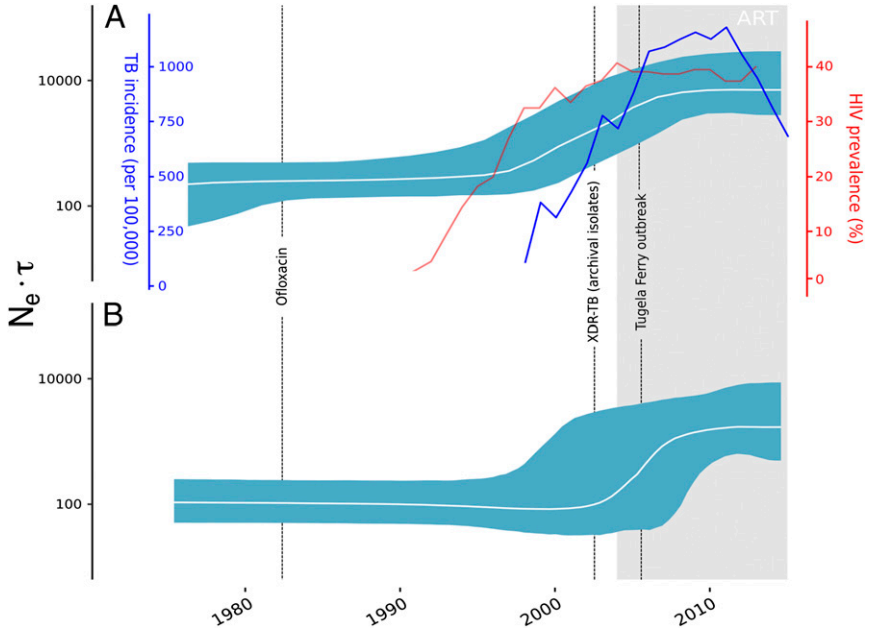

Fig. 2. Bayesian skyline plot for LAM4/KZN, HIV prevalence in antenatal clinics, TB incidence by year in KZN, and neutrality statistics for LAM4/KZN isolates. (*A*) BSP estimated using all variant sites. The white line indicates the median values for the product of the N_e, the effective population size, and τ, the generation time over time; 95% HPD range is indicated in turquoise. (*B*) BSP estimated using synonymous sites only. HIV prevalence and TB incidence data were obtained from publicly reported sources (66). The period following widespread introduction of combination antiretroviral therapy is shown as the gray region on the BSP.

were grouped by district, a Mantel test revealed a significant positive correlation between groups' pairwise F_{ST} and geographic distance, as measured by the shortest road distance between groups ($r = 0.4500$, $P = 0.0139$) (Fig. 3*A*). This pattern, known as isolation by distance, was also observed for the analysis when isolates were grouped using hierarchical clustering on interisolate distances and when the haversine great-circle distance between groups was used instead of road distance (*SI Appendix*, Fig. S7). The most genetically differentiated subpopulation was the northeastern uMkhanyakude district, followed by the Ugu and Uthukela districts around the periphery of KwaZulu-Natal (*SI Appendix*, Fig. S8). The eThekwini district that includes the largest city in the province, Durban, and the remaining southern and coastal districts around eThekwini had relatively lower pairwise F_{ST} values (Fig. 3*D*), suggesting that groups of LAM4/KZN isolates in these districts are more strongly linked to one another via internal migration.

The pattern of isolation by distance can be caused by multiple processes. The possibility that LAM4/KZN represents an older population at equilibrium between migration and drift is unlikely given the estimate of a short TMRCA and the results of neutrality tests that reject equilibrium models (Fig. 1 and *SI Appendix*, Table S2). Another possibility is that LAM4/KZN represents a nonequilibrium population that has undergone a recent geographic range expansion from a common origin. We tested this range-expansion hypothesis by using the π and ψ statistics as described in *Methods*. When π was regressed against shortest road distance from an origin, the strongest negative correlation occurred with the uMkhanyakude district as origin (Pearson $r = -0.7674$, $P = 0.0079$; Fig. 3*B*). Similarly, for ψ, the strongest positive correlation occurred with the uMkhanyakude district as origin (Pearson $r = 0.6766$, $P = 0.0227$; Fig. 3*C*). Spatial trends consistent with range expansion from an origin in uMkhanyakude are also observed if the π and ψ values for this location are excluded from the regressions (Fig. 3 *B* and *C*, dashed lines), although these trends are not statistically significant (π: $R = -0.205$, $P = 0.3128$; ψ: $R = 0.450$, $P = 0.1316$). The relationships between π, ψ, and distance are reversed when uMzinyathi district is tested as the origin (π: $R = 0.734$, $P = 0.9878$; ψ: $R = -0.481$, $P = 0.905$), providing strong evidence that Tugela Ferry was not the origin location for the XDR LAM4/KZN epidemic. By interpolating the correlations between the districts, we found the best evidence for an origin subpopulation in the northeastern uMkhanyakude district (Fig. 3 *E* and *F*). When the analysis was done using subpopulations defined by hierarchical clustering rather than by districts, the results were also consistent with uMkhanyakude as a well-differentiated subpopulation that represented the most likely origin for LAM4/KZN: significant isolation by distance (Mantel $r = 0.4300$, $P = 0.0329$), significant negative correlation for π with distance from uMkhanyakude (Pearson $r = -0.6732$, $P = 0.0164$), and a trending positive correlation for ψ with distance from uMkhanyakude ($r = 0.4838$, $P = 0.0782$; *SI Appendix*, Fig. S9). Importantly, these results were not explained by differences in sample sizes, differences in estimated human population densities between sites (*SI Appendix*, Fig. S10), or temporal confounding of the observed spatial genetic trends (*SI Appendix*, Table S3). Thus, these results suggest an origin of LAM4/KZN ~400 km away from the first reported cases at Tugela Ferry.

Discussion

Multiple environmental and pathogen-specific factors must align such that novel AMR pathogens are able to establish sustained transmission, propagate among susceptible hosts, and disperse into geographically separated populations (20, 34). Our results indicate that XDR-TB strain LAM4/KZN emerged in the early 1990s, most likely in a district with high preexisting rates of drug-resistant TB located hundreds of kilometers away from where this strain was first discovered (22). We estimate that LAM4/KZN acquired the mutations required for the XDR phenotype approximately 12 y before the Tugela Ferry outbreak and 8 y before the first known XDR LAM4/KZN clinical isolate was collected in KwaZulu-Natal (35), indicating that this strain emerged years before it was identified by existing clinical and

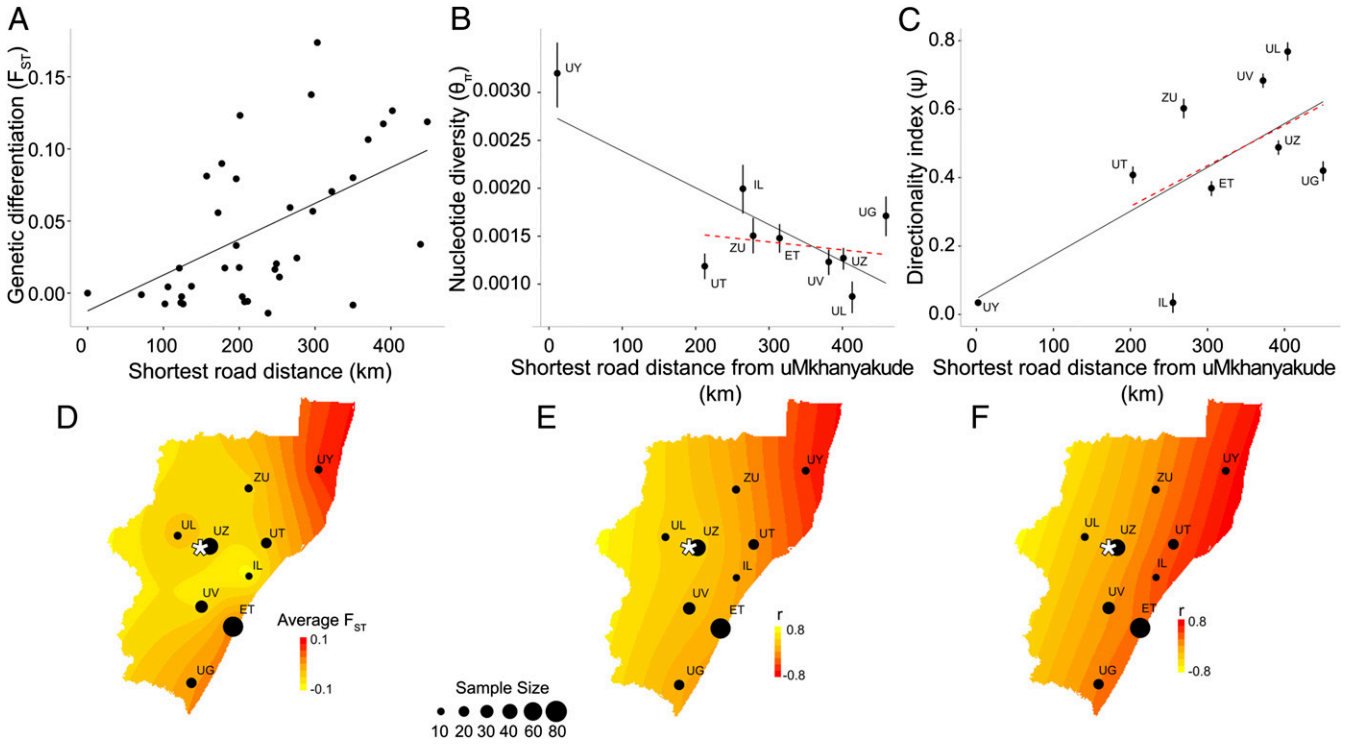

Fig. 3. Population-genetic signatures of range expansion from a common origin for LAM4/KZN isolates. (*A*) Pairwise F_{ST} vs. shortest road distance between isolates grouped by district. (*B* and *C*) Linear regression of nucleotide diversity (π) or the directionality index (ψ) vs. shortest road distance from uMkhanyakude district. Black solid lines show regression of π or ψ versus distance from uMkhanyakude for all locations. Red dashed lines show the regression of π or ψ versus distance from uMkhanyakude but excluding the π or ψ values for uMkhanyakude. (*D*) Average pairwise F_{ST} estimates for districts, with kriging interpolation between sampling points; red color indicates greater differentiation. (*E* and *F*) Spatial distribution of the correlations in *B* and *C*, with kriging interpolation between sampling points; red color indicates better evidence of origin. Districts are abbreviated as follows: eThekwini (ET), iLembe (IL), Ugu (UG), uThukela (UL), uThungulu (UT), uMgungundlovu (UV), uMkhanyakude (UY), uMzinyathi (UZ), and Zululand (ZU). Error bars indicate the SD of π or ψ based on 1,000 bootstrap replicates. The location of Tugela Ferry is indicated with a star.

public health surveillance activities. LAM4/KZN subsequently underwent marked population expansion and concurrent geographic dispersal following the onset of the generalized HIV epidemic. Default from MDR-TB treatment was common during this period (36) and susceptibility testing for antibiotics used for the treatment of MDR-TB was limited. We find evidence that early acquisition of unique *rpoB* mutations and other mutations (including *drrA* R262G, Rv1144-mmpL13a −102C>A, and *Rv2000* L275P) may have acted in multiple ways to augment the overall pathoadaptive fitness of LAM4/KZN and drive the epidemiological success of this strain. These findings underscore the complex, multifactorial processes behind the XDR-TB epidemic in KwaZulu-Natal, and highlight multiple important priorities for future efforts to contain drug-resistant TB and AMR pathogens in general.

HIV coinfection has had wide-reaching impacts on global TB control. Expansion of LAM4/KZN concurrent with the onset of widespread HIV infection may in part reflect dramatic increases in overall HIV-associated TB transmission (including both susceptible and other drug-resistant strains) that occurred in South Africa between 2000 and 2010. However, this phenomenon alone fails to explain the differential expansion of LAM4/KZN (versus other XDR-TB strains in KwaZulu-Natal). Prior work has suggested that LAM4/KZN may be less virulent than closely related drug-susceptible strains, and that decreased virulence may have promoted increased transmission between HIV-coinfected patients (37). Nearly all XDR-TB cases identified early during the Tugela Ferry outbreak were HIV-coinfected (22, 38), whereas 32% of XDR-TB cases in our more recent study were HIV-negative. We found no evidence of genetic differentiation between isolates from patients with and without HIV coinfection, suggesting ongoing linked transmission between these patient populations. Importantly, we note that case fatality rates in the Tugela Ferry outbreak were nearly 100% among patients with HIV (39) and it is possible that our more recent study (started ∼6 y after the Tugela Ferry outbreak) failed to capture sizable numbers of early clustered cases among HIV-coinfected patients. Taken in context, our results indicate that although HIV coinfection was undoubtedly important early in the epidemic, linked transmission between HIV-positive and HIV-negative patients has played an important part in the more recent dispersal of LAM4/KZN (40).

Earlier studies attributed the majority of XDR-TB cases in Tugela Ferry to nosocomial transmission between cohospitalized patients with HIV (38). However, there is increasing recognition that community transmission is an important driver in the expansion and dispersal of specific TB strains (41, 42), and evidence from recent studies indicates that LAM4/KZN currently circulates at the community level throughout KwaZulu-Natal (6, 7, 43). We have found genomic evidence supporting marked expansion of LAM4/KZN ∼10 y before the Tugela Ferry outbreak, suggesting that widespread community transmission is not only ongoing but significantly predates the first clinical reports of XDR-TB in the province. Multiple processes likely contributed to the rapid dispersal of LAM4/KZN during this pre-detection period (44), including limited availability of drug-susceptibility testing and provision of ineffective antibiotic therapy that likely prolonged the duration over which individual patients were infectious, traveling, and mixing with susceptible populations in other locations.

Determining where the XDR-TB epidemic began is an essential first step for identifying priority areas for public health surveillance and understanding the networks underlying geographic dispersal of XDR-TB. Our analysis indicates that LAM4/KZN most likely originated in uMkhanyakude, a rural district in northeastern KwaZulu-Natal that borders Mozambique and eSwatini. We note that uMkhanyakude consistently records the highest rates of drug-resistant TB cases (45) and the highest HIV prevalence in the province (46). The frequency with which new drug-resistance mutations arise increases with increasing population size (34), and it is not unexpected that XDR-TB would arise in an environment where other forms of drug-resistant TB are already highly prevalent and where province-wide default and failure rates for MDR-TB treatment are high (36). We find no evidence to suggest that LAM4/KZN originated in Tugela Ferry; rather, isolates from uMzinyathi district (where Tugela Ferry is located) exhibit genetic signatures consistent with an *Mtb* subpopulation derived from another, more diverse origin population. High mortality rates in Tugela Ferry, driven by HIV coinfection, and resultant detection bias may explain why LAM4/KZN was first identified only after dispersal from its origin in uMkhanyakude (22). Expanding and intensifying genomic surveillance efforts may provide opportunities to identify and preemptively contain future drug-resistant TB outbreaks in their origin locations.

Recent studies have suggested that rural–urban migration between densely populated eThekwini (which includes Durban) and other districts, rather than local transmission within districts, likely drives geospatial patterns of XDR-TB transmission in KwaZulu-Natal (43). Consistent with these findings, our results suggest that introduction of LAM4/KZN from its origin in outlying, sparsely populated uMkhanyakude into urban areas like eThekwini was likely a key event in the epidemiological history of this subpopulation.

Similar to other epidemics in which rural–urban migration is an important driver (47, 48), we propose that urban areas in KwaZulu-Natal likely act as hubs for the amplification and dispersal of XDR-TB across the province, as suggested by the low level of genetic differentiation we observed between eThekwini and other central districts. We note that if geographic dispersal occurs more rapidly than separate pathogen populations accrue genetic evidence of differentiation, conventional strategies for inferring transmission based on differentiation (for example, using SNP differences between patient isolates) may be difficult. Additional work is needed to further investigate this hypothesis and characterize the contact and mobility networks underlying geographic dispersal of LAM4/KZN.

Pathogen-specific factors that contribute to epidemiological success—via enhanced transmissibility (10), adaptive recovery of pathogen fitness (11, 12), and epistasis between drug-resistance mutations (13)—have been studied extensively. There is strong contextual evidence suggesting that secondary mutations in *rpoB* have played an important part in the epidemiological success of LAM4/KZN. Although LAM4/KZN causes ∼78% of all XDR-TB cases in KwaZulu-Natal, the ancestral population from which this strain evolved causes only 25% of multidrug-resistant TB cases (i.e., TB resistant to 2 first-line antibiotics) in the province (49). This MDR-TB subpopulation is characterized by a single nonsynonymous *rpoB* variant at L452P, a mutation that is associated with low-level rifampin resistance and delayed in vitro growth (50) that may explain the relatively limited transmission of this strain. We also note that XDR isolates carrying L452P but lacking D435G and I1106T are rare in our sample compared with the abundantly prevalent L452P/D435G/I1106T triple mutant. We infer from Bayesian skyline analysis that the LAM4/KZN subpopulation expanded only after acquiring 2 important strain-specific *rpoB* variants (D435G and I1106T). Notably, I1106T is located at the RNAPβ–RNAPβ′ interface, adjacent to the site of known putative *rpoC* compensatory mutations.

We hypothesize that these secondary *rpoB* mutations likely contributed to enhanced transmission and dispersal of KZN/LAM4 via enhanced resistance to rifampin and recovery of overall pathoadaptive fitness versus the L452P single mutant. L452P confers only low-level rifampin resistance, whereas mutations at position 435 (including D435Y and D435G) are associated with distinctly higher rifampin MICs (50), and in LAM4/KZN acquisition of D435G likely augmented rifampin resistance versus the L452P single mutant (51). Recent studies have implicated changes in protein stability (52) and transcriptional efficiency (53) as important determinants of how putative compensatory mutations impact fitness in rifampin-resistant *Mtb*. Although in our analysis D435G and I1106T only partially offset the destabilizing effects of L452P, multiple studies suggest that the compensatory effects of D435G may not be related exclusively to its effects on protein stability. Importantly, in vitro studies have demonstrated that the mutation homologous to D435G in *Escherichia coli* (D516G) is associated with increased growth and enhanced transcriptional efficiency versus wild type and confers compensatory recovery of fitness in rifampin-resistant *rpoB* mutants (54). In addition, we have shown that D435G is among the few RRDR mutations that induce changes in the distribution of charges near the RNA binding site, a property shared by H445Y, a variant associated with extensive, pleiotropic changes in surface lipid expression and host macrophage programming (18). In vitro studies are needed to characterize how D435G impacts pathogen fitness and host–pathogen interactions in L452P mutants. Lastly, we provide supporting evidence for the assertion that 2 previously identified mutations, *drrA* R262G and Rv1144-mmpL13a –102C>A, may be important contributors to the epidemiological success of LAM4/KZN. Prior work has hypothesized that *drrA*, a probable drug efflux pump, and Rv1144-mmpL13a –102C>A, located in a promoter region for an operon containing 2 transmembrane transport proteins (mmpL13a and mmpL13b), may enhance both antibiotic resistance and transport of important lipid products mediating virulence (including phthiocerol dimycocerosate) (55). In addition, we note that other less successful XDR-TB strains circulating in KwaZulu-Natal also harbor putative compensatory mutations (for example, in *rpoC*), underscoring how pathogen-specific factors alone may fail to fully explain the differential success of certain drug-resistant *Mtb* epidemic clones. These findings highlight how even the pathogen-specific mediators of epidemic success can only be properly understood as the convergence of multiple adaptive processes that enhance the selective advantage of a given strain with respect to a complex fitness landscape involving antibiotic selection pressure, virulence, and a complicated array of host–pathogen interactions.

Taken together, our results underscore how field surveillance efforts leveraging geo-located clinical isolates and phylogenetic and population-genetic analyses of high-resolution bacterial genomic data can delineate the spatiotemporal context and the drivers behind the rise of epidemic AMR pathogens. Two conclusions from our study are particularly relevant for future public health efforts: First, highly local ecological factors (in this case, high existing rates of drug-resistant TB and HIV coinfection in uMkhanyakude district) can influence the evolutionary conditions that promote the selection and expansion of drug-resistant mutants, and surveillance efforts should be designed to capture variation in these factors at the local level. Second, failure to contain emergent AMR epidemics due to lack of early recognition and appropriate treatment, coupled with internal human migration, can facilitate rapid, widespread geographic dispersal. Late detection, following dispersal between geographically distant populations, can confound efforts to identify local risk factors for AMR emergence. Although multiple logistical challenges still exist, and the efficacy and cost-effectiveness of whole-genome sequencing for this application has yet to be established, we

anticipate that routine integration of whole-genome sequencing into public health surveillance can enable interventions for early detection and containment of AMR epidemics, and thereby significantly circumscribe the potential impact of these major threats to global public health.

Methods

Study Design and Ethical Oversight. We conducted a prospective study of all culture-confirmed XDR-TB cases in KwaZulu-Natal between 2011 and 2014 (6); 521 patients with XDR-TB were identified during recruitment and 404 total patients were enrolled in the study. Written informed consent was obtained from all participants or, for deceased or severely ill participants, from the next of kin. Participants underwent serological testing for HIV, were interviewed regarding history of prior MDR-TB, and had the locations of their homes mapped by GPS. Participants were considered to have had a prior episode of MDR-TB if their most recent MDR-TB culture result was more than 3 mo before their initial XDR-TB culture result. The study protocol was approved by the institutional review boards of Columbia University Medical Center, Emory University, Albert Einstein College of Medicine, and the University of KwaZulu-Natal and by the US Centers for Disease Control and Prevention.

Genomic Characterization of *Mtb* Clinical Isolates. XDR-TB isolates (318) with WGS data meeting criteria for read depth and sequence coverage were included for analysis, of which 250 were LAM4/KZN (*SI Appendix*, Fig. S11). In addition, we included 54 MDR-TB and 52 pan-susceptible isolates collected during concurrent cross-sectional studies in KZN (56, 57) and 8 archival LAM4/KZN isolates collected between 1994 and 2010 (7, 27), with the goal of including drug-susceptible, MDR, and XDR isolates for each *Mtb* subpopulation in KZN. These isolates underwent whole-genome sequencing and targeted sequencing of drug-resistance genes as described previously (*SI Appendix, Methods*) (6). We used the SNP-based *Mtb* sublineage taxonomy described by Coll et al. (24) to label individual isolates and sublineages in our phylogenetic analysis. Sequencing data for this project are available in the NCBI Sequence Read Archive (BioProject accession no. PRJNA476470; *SI Appendix*, Table S4).

Phylogenetic Reconstruction and Dating of Drug-Resistance Mutations. We used Markov chain Monte Carlo (MCMC)-based Bayesian phylogenetic inference in BEAST v1.8.4 (58) to reconstruct dated phylogenetic trees and infer the time to the most recent common ancestor for isolates carrying drug-resistance mutations of interest. We used a strict molecular clock with an informative prior distribution on the mutation rate and the HKY nucleotide-substitution model (59) with among-site rate heterogeneity defined by a Γ distribution. We ran each MCMC chain for at least 200 million states with 10 to 25% burn-in (see *SI Appendix, Methods* for additional information). Nodes corresponding to the most recent common ancestor for relevant drug-resistance mutations were identified via maximum-likelihood ancestral state reconstruction in Mesquite v3.51 (60).

Bayesian Skyline Analysis and Population-Genetic Analysis. We generated nonparametric estimates of *Mtb* population history by constructing Bayesian skyline plots (BSPs) (61). Given possible confounding of BSP analyses by population structure (62), we selected a monophyletic clade for BSP analysis that includes only XDR LAM4/KZN sequences and closely related ancestral strains. We ran separate analyses including and excluding nonsynonymous SNPs to evaluate the possible effects of selection on amino acid changes on BSP-derived estimates of prior demographic history. To improve mixing and convergence during Bayesian skyline analysis, we analyzed a subsample of LAM4/KZN isolates (n = 50 randomly sampled from 250 total LAM4/KZN isolates) plus closely related MDR and drug-susceptible isolates belonging to the same clade (n = 89). Bayesian skyline plot results were similar for this subsample compared with the same analysis with all 250 LAM4/KZN isolates included (*SI Appendix*, Fig. S12).

We compared terminal branch lengths (measured as the number of SNPs mapped to each terminal branch) between LAM4/KZN isolates and XDR-TB from all other *Mtb* subpopulations using the 2-sample Kolmogorov–Smirnov test (10). To minimize differences in sampling variance introduced by differences in sample sizes between groups, we also applied the Kolmogorov–Smirnov test to 10,000 equal-sized bootstrap replicates from each group (*SI Appendix*, Fig. S1).

Geospatial Population-Genetic Analysis. For geospatial analysis, groups of LAM4/KZN isolates were defined using 2 different approaches. First, we defined groups by the 11 municipal districts of KZN, using the average latitude and longitude of the participants' places of residence within the district to define the geographic center for each group. Second, we used Ward's hierarchical clustering method to cluster isolates into 15 populations based on their geographic distance from one another. With both approaches, each participant was represented by a single isolate. Groups with <4 sampled isolates were dropped from further analysis, which left 9 and 10 populations to be analyzed with these 2 approaches, respectively. All analyses were performed using both shortest road distance and haversine great-circle distance between the centers of each group. To test for geospatial genetic structure, a Mantel test with 10,000 permutations was performed using pairwise F_{ST} and geographic distance between group centers.

A serial founder model of range expansion was used to infer the geospatial origin of LAM4/KZN. This model predicts decreased genetic diversity and increased derived allele frequency, on average, as subpopulations are established farther from an origin population by a series of bottlenecks (63). Nucleotide diversity (π) was calculated with the R package *pegas* (64) and derived allele frequency skew (referred to as the "directionality index"; ψ) was calculated with modified R code of the *rangeExpansion* package (65). To identify the derived alleles, we used the LAM4/KZN sister clade 4.3.2 and assigned those alleles unique to LAM4/KZN as derived and those alleles shared with the sister clade as ancestral. To infer the origin under a serial founder model, π and ψ measurements were regressed against geographic distance from an origin population. Each subpopulation was tested as an origin population, and the best origin was identified based on the most negative correlation (for π) or positive correlation (for ψ) with geographic distance from that origin.

Data Availability. All data discussed in the paper will be made available to readers.

ACKNOWLEDGMENTS. This work was supported by grants from the National Institute of Allergy and Infectious Disease, National Institutes of Health (R01AI089349, R01AI087465, R01AI114304, and R01AI138646); NIH career development awards and training grants (K24AI114444 to N.R.G.; K23AI134182 to S.C.A.; T32AI007061 to T.S.B.); and grants from the Emory University Center for AIDS Research (P30AI050409), Albert Einstein College of Medicine Center for AIDS Research (P30AI124414), and Albert Einstein College of Medicine Institute for Clinical and Translational Research (UL1 TR001073). We would like to acknowledge in kind support from the American Museum of Natural History and the Flatiron Institute–Simons Foundation for providing computational resources. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the funding agencies.

1. F. R. Deleo *et al.*, Molecular dissection of the evolution of carbapenem-resistant multilocus sequence type 258 *Klebsiella pneumoniae*. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 4988–4993 (2014).
2. U. Seybold *et al.*, Emergence of community-associated methicillin-resistant *Staphylococcus aureus* USA300 genotype as a major cause of health care-associated blood stream infections. *Clin. Infect. Dis.* **42**, 647–656 (2006).
3. V. Eldholm *et al.*, Four decades of transmission of a multidrug-resistant *Mycobacterium tuberculosis* outbreak strain. *Nat. Commun.* **6**, 7119 (2015).
4. N. Casali *et al.*, Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat. Genet.* **46**, 279–286 (2014).
5. P. J. Bifani *et al.*, Origin and interstate spread of a New York City multidrug-resistant *Mycobacterium tuberculosis* clone family. *JAMA* **275**, 452–457 (1996).
6. N. S. Shah *et al.*, Transmission of extensively drug-resistant tuberculosis in South Africa. *N. Engl. J. Med.* **376**, 243–253 (2017).
7. K. A. Cohen *et al.*, Evolution of extensively drug-resistant tuberculosis over four decades: Whole genome sequencing and dating analysis of *Mycobacterium tuberculosis* isolates from KwaZulu-Natal. *PLoS Med.* **12**, e1001880 (2015).
8. B. R. Levin *et al.*, The population genetics of antibiotic resistance. *Clin. Infect. Dis.* **24** (suppl. 1), S9–S16 (1997).
9. D. I. Andersson, B. R. Levin, The biological cost of antibiotic resistance. *Curr. Opin. Microbiol.* **2**, 489–493 (1999).
10. K. E. Holt *et al.*, Frequent transmission of the *Mycobacterium tuberculosis* Beijing lineage and positive selection for the EsxW Beijing variant in Vietnam. *Nat. Genet.* **50**, 849–856 (2018).
11. I. Comas *et al.*, Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat. Genet.* **44**, 106–110 (2011).
12. M. Merker *et al.*, Compensatory evolution drives multidrug-resistant tuberculosis in Central Asia. *eLife* **7**, e38200 (2018).
13. S. Borrell *et al.*, Epistasis between antibiotic resistance mutations drives the evolution of extensively drug-resistant tuberculosis. *Evol. Med. Public Health* **2013**, 65–74 (2013).
14. D. R. Sherman *et al.*, Compensatory ahpC gene expression in isoniazid-resistant *Mycobacterium tuberculosis*. *Science* **272**, 1641–1643 (1996).

15. L. S. Munoz-Price *et al*., Clinical epidemiology of the global expansion of *Klebsiella pneumoniae* carbapenemases. *Lancet Infect. Dis.* **13**, 785–796 (2013).
16. C. L. Wielders, A. C. Fluit, S. Brisse, J. Verhoef, F. J. Schmitz, mecA gene is widely disseminated in *Staphylococcus aureus* population. *J. Clin. Microbiol.* **40**, 3970–3975 (2002).
17. S. Takala-Harrison *et al*., Independent emergence of artemisinin resistance mutations among *Plasmodium falciparum* in Southeast Asia. *J. Infect. Dis.* **211**, 670–679 (2015).
18. N. C. Howard *et al*., *Mycobacterium tuberculosis* carrying a rifampicin drug resistance mutation reprograms macrophage metabolism through cell wall lipid changes. *Nat. Microbiol.* **3**, 1099–1108 (2018).
19. M. T. Zaw, N. A. Emran, Z. Lin, Mutations inside rifampicin-resistance determining region of rpoB gene associated with rifampicin-resistance in *Mycobacterium tuberculosis*. *J. Infect. Public Health* **11**, 605–610 (2018).
20. A. H. Holmes *et al*., Understanding the mechanisms and drivers of antimicrobial resistance. *Lancet* **387**, 176–187 (2016).
21. M. Lipsitch, The rise and fall of antimicrobial resistance. *Trends Microbiol.* **9**, 438–444 (2001).
22. N. R. Gandhi *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* **368**, 1575–1580 (2006).
23. N. Sarita Shah *et al*., Extensively drug-resistant Mycobacterium tuberculosis. National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/bioproject/PRJNA476470. Deposited 3 July 2018.
24. F. Coll *et al*., A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat. Commun.* **5**, 4812 (2014).
25. C. B. Ford *et al*., *Mycobacterium tuberculosis* mutation rate estimates from different lineages predict substantial differences in the emergence of drug-resistant tuberculosis. *Nat. Genet.* **45**, 784–790 (2013).
26. F. Coll *et al*., Genome-wide analysis of multi- and extensively drug-resistant *Mycobacterium tuberculosis*. *Nat. Genet.* **50**, 307–316 (2018).
27. T. R. Ioerger *et al*., Genome analysis of multi- and extensively-drug-resistant tuberculosis from KwaZulu-Natal, South Africa. *PLoS One* **4**, e7778 (2009).
28. V. Molodtsov, N. T. Scharf, M. A. Stefan, G. A. Garcia, K. S. Murakami, Structural basis for rifamycin resistance of bacterial RNA polymerase by the three most clinically important RpoB mutations found in *Mycobacterium tuberculosis*. *Mol. Microbiol.* **103**, 1034–1045 (2017).
29. T. Song *et al*., Fitness costs of rifampicin resistance in *Mycobacterium tuberculosis* are amplified under conditions of nutrient starvation and compensated by mutation in the β′ subunit of RNA polymerase. *Mol. Microbiol.* **91**, 1106–1119 (2014).
30. S. D. Lawn, L. G. Bekker, K. Middelkoop, L. Myer, R. Wood, Impact of HIV infection on the epidemiology of tuberculosis in a peri-urban community in South Africa: The need for age-specific interventions. *Clin. Infect. Dis.* **42**, 1040–1047 (2006).
31. S. Hermans, C. R. Horsburgh Jr, R. Wood, A century of tuberculosis epidemiology in the Northern and Southern Hemisphere: The differential impact of control interventions. *PLoS One* **10**, e0135179 (2015).
32. S. S. Abdool Karim, G. J. Churchyard, Q. A. Karim, S. D. Lawn, HIV infection and tuberculosis in South Africa: An urgent need to escalate the public health response. *Lancet* **374**, 921–933 (2009).
33. M. Lapierre, C. Blin, A. Lambert, G. Achaz, E. P. Rocha, The impact of selection, gene conversion, and biased sampling on the assessment of microbial demography. *Mol. Biol. Evol.* **33**, 1711–1725 (2016).
34. P. A. zur Wiesch, R. Kouyos, J. Engelstädter, R. R. Regoes, S. Bonhoeffer, Population biological principles of drug-resistance evolution in infectious diseases. *Lancet Infect. Dis.* **11**, 236–247 (2011).
35. M. Pillay, A. W. Sturm, Evolution of the extensively drug-resistant F15/LAM4/KZN strain of *Mycobacterium tuberculosis* in KwaZulu-Natal, South Africa. *Clin. Infect. Dis.* **45**, 1409–1414 (2007).
36. J. C. Brust, N. R. Gandhi, H. Carrara, G. Osburn, N. Padayatchi, High treatment failure and default rates for patients with multidrug-resistant tuberculosis in KwaZulu-Natal, South Africa, 2000-2003. *Int. J. Tuberc. Lung Dis.* **14**, 413–419 (2010).
37. K. L. Smith *et al*., Reduced virulence of an extensively drug-resistant outbreak strain of *Mycobacterium tuberculosis* in a murine model. *PLoS One* **9**, e94953 (2014).
38. N. R. Gandhi *et al*., Nosocomial transmission of extensively drug-resistant tuberculosis in a rural hospital in South Africa. *J. Infect. Dis.* **207**, 9–17 (2013).
39. N. R. Gandhi *et al*., Risk factors for mortality among MDR- and XDR-TB patients in a high HIV prevalence setting. *Int. J. Tuberc. Lung Dis.* **16**, 90–97 (2012).
40. V. Eldholm *et al*., Impact of HIV co-infection on the evolution and transmission of multidrug-resistant tuberculosis. *eLife* **5**, e16644 (2016).
41. J. R. Glynn *et al*., Whole genome sequencing shows a low proportion of tuberculosis disease is attributable to known close contacts in rural Malawi. *PLoS One* **10**, e0132840 (2015).
42. J. A. Guerra-Assunção *et al*., Large-scale whole genome sequencing of *M. tuberculosis* provides insights into transmission in a high prevalence area. *eLife* **4**, e05166 (2015).
43. K. N. Nelson *et al*., Spatial patterns of extensively drug-resistant tuberculosis (XDR-tuberculosis) transmission in KwaZulu-Natal, South Africa. *J. Infect. Dis.* **218**, 1964–1973 (2018).
44. S. Basu *et al*., Averting epidemics of extensively drug-resistant tuberculosis. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 7672–7677 (2009).
45. K. Wallengren *et al*., Drug-resistant tuberculosis, KwaZulu-Natal, South Africa, 2001-2007. *Emerg. Infect. Dis.* **17**, 1913–1916 (2011).
46. National Department of Health, *The National Antenatal Sentinel HIV & Syphilis Prevalence Survey in South Africa, 2011* (South African National Department of Health, Pretoria, South Africa, 2012).
47. M. N. Lurie, B. G. Williams, Migration and health in southern Africa: 100 years and still circulating. *Health Psychol. Behav. Med.* **2**, 34–40 (2014).
48. M. P. Fallah, L. A. Skrip, J. Enders, Preventing rural to urban spread of Ebola: Lessons from Liberia. *Lancet* **392**, 279–280 (2018).
49. N. R. Gandhi *et al*., Minimal diversity of drug-resistant *Mycobacterium tuberculosis* strains, South Africa. *Emerg. Infect. Dis.* **20**, 426–433 (2014).
50. P. Miotto, A. M. Cabibbe, E. Borroni, M. Degano, D. M. Cirillo, Role of disputed mutations in the *rpoB* gene in interpretation of automated liquid MGIT culture results for rifampin susceptibility testing of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **56**, e01599-17 (2018).
51. N. Dookie, A. W. Sturm, P. Moodley, Mechanisms of first-line antimicrobial resistance in multi-drug and extensively drug resistant strains of *Mycobacterium tuberculosis* in KwaZulu-Natal, South Africa. *BMC Infect. Dis.* **16**, 609 (2016).
52. S. Portelli, J. E. Phelan, D. B. Ascher, T. G. Clark, N. Furnham, Understanding molecular consequences of putative drug resistant mutations in *Mycobacterium tuberculosis*. *Sci. Rep.* **8**, 15356 (2018).
53. M. A. Stefan, F. S. Ugur, G. A. Garcia, Source of the fitness defect in rifamycin-resistant *Mycobacterium tuberculosis* RNA polymerase and the mechanism of compensation by mutations in the β′ subunit. *Antimicrob. Agents Chemother.* **62**, e00164-18 (2018).
54. M. G. Reynolds, Compensatory evolution in rifampin-resistant *Escherichia coli*. *Genetics* **156**, 1471–1481 (2000).
55. P. A. Black *et al*., Energy metabolism and drug efflux in *Mycobacterium tuberculosis*. *Antimicrob. Agents Chemother.* **58**, 2491–2503 (2014).
56. J. C. M. Brust *et al*., Improved survival and cure rates with concurrent treatment for multidrug-resistant tuberculosis-human immunodeficiency virus coinfection in South Africa. *Clin. Infect. Dis.* **66**, 1246–1253 (2018).
57. A. Nanoo *et al*., Nationwide and regional incidence of microbiologically confirmed pulmonary tuberculosis in South Africa, 2004-12: A time series analysis. *Lancet Infect. Dis.* **15**, 1066–1076 (2015).
58. A. J. Drummond, A. Rambaut, BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
59. M. Hasegawa, H. Kishino, T. Yano, Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**, 160–174 (1985).
60. W. P. Maddison, D. R. Maddison, Mesquite: A Modular System for Evolutionary Analysis (Version 3.51, 2018). http://www.mesquiteproject.org. Accessed 10 October 2019.
61. A. J. Drummond, A. Rambaut, B. Shapiro, O. G. Pybus, Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).
62. R. Heller, L. Chikhi, H. R. Siegismund, The confounding effect of population structure on Bayesian skyline plot inferences of demographic history. *PLoS One* **8**, e62992 (2013).
63. L. Challagundla *et al*., Range expansion and the origin of USA300 North American epidemic methicillin-resistant *Staphylococcus aureus*. *MBio* **9**, e02016-17 (2018).
64. E. Paradis, pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics* **26**, 419–420 (2010).
65. B. M. Peter, M. Slatkin, Detecting range expansions from genetic data. *Evolution* **67**, 3274–3289 (2013).
66. Health Systems Trust, Health Indicator Database. https://www.hst.org.za/healthindicators. Accessed 20 July 2017.