Evolution of extensively drug-resistant tuberculosis over four decades revealed by whole genome sequencing of Mycobacterium tuberculosis from KwaZulu-Natal, South Africa

<table>
<thead>
<tr>
<th>Manuscript Number:</th>
<th>PMEDICINE-D-15-00672R1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Title:</td>
<td>Evolution of extensively drug-resistant tuberculosis over four decades revealed by whole genome sequencing of Mycobacterium tuberculosis from KwaZulu-Natal, South Africa</td>
</tr>
<tr>
<td>Short Title:</td>
<td>Whole genome sequencing reveals 40-year evolution of epidemic XDR-TB</td>
</tr>
<tr>
<td>Article Type:</td>
<td>Research Article</td>
</tr>
<tr>
<td>Keywords:</td>
<td>Tuberculosis; Drug resistance; MDR; XDR; Tugela Ferry; KwaZulu-Natal; South Africa; whole genome sequencing; rifampicin; rifampin; compensatory; compensation; evolution</td>
</tr>
</tbody>
</table>
| Corresponding Author: | Ashlee M. Earl  
                      Broad Institute of MIT & Harvard  
                      Cambridge, UNITED STATES |
| Corresponding Author's Institution: | Broad Institute of MIT & Harvard |
| First Author:      | Keira A. Cohen |
| Order of Authors:  | Keira A. Cohen  
                      Thomas Abeel  
                      Abigail Manson McGuire  
                      Christopher A. Desjardins  
                      Vanisha Munsamy  
                      Terrance P. Shea  
                      Bruce J. Walker  
                      Nokqubela Bantubani  
                      Deepak V. Almeida  
                      Lucia Alvarado  
                      Sinead B. Chapman  
                      Nomonde R. Mvelase  
                      Eamon Y. Duffy  
                      Michael G. Fitzgerald  
                      Pamla Govender  
                      Sharvari Gujja  
                      Susanna Hamilton  
                      Clint Howarth  
                      Jeffrey D. Larimer |
Background: The largest global outbreak of extensively drug-resistant (XDR) tuberculosis (TB) was identified in Tugela Ferry, KwaZulu-Natal, South Africa in 2005. Understanding the antecedents and timing of emergence of drug resistance in this fatal outbreak of XDR will inform the control and prevention of drug-resistant TB.

Methods and Findings: We performed whole genome sequencing and drug susceptibility testing on 337 clinical isolates of Mycobacterium tuberculosis (M. tb) collected in KwaZulu-Natal from 2008 to 2013, in addition to historical isolates, including one from the Tugela Ferry outbreak. We utilized an array of whole genome comparative techniques to assess the relatedness among strains, to establish the order of acquisition of drug resistance mutations including the timing of acquisitions leading to XDR-TB in the LAM4 spoligotype, and to calculate the number of independent evolutionary emergences of MDR and XDR. Our sequencing and analysis revealed a 50-member clone of XDR M. tb that was highly related to the Tugela Ferry XDR outbreak strain. We estimated that mutations conferring isoniazid and streptomycin resistance in this clone were acquired 50 years prior to the Tugela Ferry outbreak, with the subsequent emergence of multidrug resistance (MDR) and XDR occurring 30 and 40 years later, respectively. We observed frequent de novo evolution of MDR and XDR, with 56 and 9 independent evolutionary events, respectively. Isoniazid resistance evolved before rifampicin resistance 46 times, whereas rifampicin resistance evolved prior to isoniazid only twice. We identified additional putative compensatory mutations to rifampicin in this dataset.

Conclusions: In the first whole genome-based analysis of the emergence of drug resistance among clinical isolates of M. tb, we show that the ancestral precursor of the LAM4 XDR outbreak strain in Tugela Ferry gained mutations to first line drugs at the beginning of the antibiotic era. Subsequent accumulation of step-wise resistance mutations, occurring over decades and prior to the explosion of HIV in this region, yielded MDR and XDR, permitting the emergence of compensatory mutations. Our results suggest that drug-resistant strains circulating today reflect not only vulnerabilities of current TB control efforts, but also those that date back fifty years. In drug-resistant TB, isoniazid resistance was overwhelmingly the initial resistance mutation to be acquired, which would not be detected by current rapid molecular diagnostics employed in South Africa that assess only rifampicin resistance.

Stewart Cole
Ecole Polytechnique Federale de Lausanne
stewart.cole@epfl.ch
A microbiologist and current director of the Global Health Institute at EPFL, Dr. Cole led the research group that first sequenced the genome of Mycobacterium tuberculosis.
in its entirety. His familiarity with both microbiology and genomics as it relates to mycobacteria will be helpful in a providing a critical review of this manuscript.

Valerie Mizrahi  
Institute of Infectious Diseases and Molecular Medicine  
valerie.mizrahi@uct.ac.za  
A premier South African molecular microbiologist and Director of the Institute of Infectious Disease and Molecular Medicine (IDM) in Cape Town, Dr. Mizrahi has extensive knowledge of tuberculosis, molecular epidemiology and the South African context.

Sven Hoffner  
Karolinska Institutet  
sven.hoffner@ki.se  
A pre-eminent bacteriologist, who directs the WHO Supranational Reference Laboratory for drug-resistant TB has invaluable expertise in this field, and would likely serve as a savvy reviewer.

**Opposed Reviewers:**

Megan Murray  
Harvard School of Public Health  
Direct competitor

Sebastian Gagneux  
Swiss Tropical and Public Health Insitute  
Direct competitor

Rob Warren  
Stellenbosch University  
Direct competitor.

**Additional Information:**

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Financial Disclosure</strong></td>
<td>Please describe all sources of funding that have supported your work. A complete funding statement should do the following:</td>
</tr>
<tr>
<td></td>
<td>Include grant numbers and the URLs of any funder's website. Use the full name, not acronyms, of funding institutions, and use initials to identify authors who received the funding.</td>
</tr>
<tr>
<td></td>
<td>Describe the role of any sponsors or funders in the study design, data collection and analysis, decision to publish, or preparation of the manuscript. If they had no role in any of the above, include this sentence at the end of your statement: &quot;The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.&quot;</td>
</tr>
<tr>
<td></td>
<td>If the study was unfunded, provide a statement that clearly indicates this, for example: &quot;The author(s) received no specific funding for this work.&quot;</td>
</tr>
<tr>
<td></td>
<td>1. National Institute of Allergy and Infectious Diseases Contract No.:HHSN272200900018C and Grant Number U19AI110818 to TA AMM CAD TS BJW LA SC MF SG SH CH JL MP MP QZ SKY JW BWB AME; K23 AI098479-01A1 to MRO U19 Ai51794 to NP and MRO</td>
</tr>
<tr>
<td></td>
<td>National Heart Blood and Lung Institute T32HL007633 to KAC</td>
</tr>
<tr>
<td></td>
<td>Stony Wold-Herbert Foundation to MRO</td>
</tr>
<tr>
<td></td>
<td>U.S. Centers for Disease Control and Prevention to ASP</td>
</tr>
<tr>
<td></td>
<td>2. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.</td>
</tr>
<tr>
<td><strong>Competing Interests</strong></td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td></td>
</tr>
<tr>
<td>The authors have declared that no competing interests exist.</td>
<td></td>
</tr>
</tbody>
</table>

You are responsible for recognizing and disclosing on behalf of all authors any competing interest that could be perceived to bias their work, acknowledging all financial support and any other relevant financial or non-financial competing interests.

Do any authors of this manuscript have competing interests (as described in the PLOS Policy on Declaration and Evaluation of Competing Interests)?

**If yes**, please provide details about any and all competing interests in the box below. Your response should begin with this statement: *I have read the journal's policy and the authors of this manuscript have the following competing interests:*

**If no** authors have any competing interests to declare, please enter this statement in the box: *"The authors have declared that no competing interests exist."

<table>
<thead>
<tr>
<th><strong>Data Availability</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes - all data are fully available without restriction</td>
</tr>
</tbody>
</table>

PLOS journals require authors to make all data underlying the findings described in their manuscript fully available, without restriction and from the time of publication, with only rare exceptions to address legal and ethical concerns (see the PLOS Data Policy and FAQ for further details). When submitting a manuscript, authors must provide a Data Availability Statement that describes where the data underlying their manuscript can be found.

Your answers to the following constitute your statement about data availability and will be included with the article in the event of publication. **Please note that simply stating 'data available on request from the author' is not acceptable. If, however, your data are only available**
Do the authors confirm that all data underlying the findings described in their manuscript are fully available without restriction?

Please describe where your data may be found, writing in full sentences. Your answers should be entered into the box below and will be published in the form you provide them, if your manuscript is accepted. If you are copying our sample text below, please ensure you replace any instances of XXX with the appropriate details.

If your data are all contained within the paper and/or Supporting Information files, please state this in your answer below. For example, “All relevant data are within the paper and its Supporting Information files.”

If your data are held or will be held in a public repository, include URLs, accession numbers or DOIs. For example, “All XXX files are available from the XXX database (accession number(s) XXX, XXX).” If this information will only be available after acceptance, please indicate this by ticking the box below. If neither of these applies but you are able to provide details of access elsewhere, with or without limitations, please do so in the box below. For example:

“Data are available from the XXX Institutional Data Access / Ethics Committee for researchers who meet the criteria for access to confidential data.”

“Data are from the XXX study whose authors may be contacted at XXX.”

Additional data availability information:

All sequencing data are available from the Sequence Read Archive NCBI under the following umbrella BioProject identifiers: PRJNA183624 and PRJNA235615.
March 19th, 2015

U.S. Headquarters
PLOS
1160 Battery Street,
Koshland Building East, Suite 100
San Francisco, CA 94111
United States

Re: Submission to PLOS Medicine

Dear Editor,

Thank you for considering our manuscript entitled “Evolution of extensively drug-resistant tuberculosis over four decades revealed by whole genome sequencing of Mycobacterium tuberculosis from KwaZulu-Natal, South Africa” for publication in PLOS Medicine.

Our manuscript describes results from the largest whole genome sequencing effort conducted to date of drug-resistant Mycobacterium tuberculosis (M. tb) strains from South Africa. Our work was done in the context of an important historical outbreak of XDR in Tugela Ferry, KwaZulu-Natal (KZN), the first and largest outbreak of its kind to be reported (Gandhi et al., The Lancet, 2006). Other large-scale whole genome sequencing studies of M. tb have focused on identifying novel drug resistance and compensatory mutations (Farhat et al., Nature Genetics, 2012; Zhang et al., Nature Genetics, 2012; Casali et al., Nature Genetics 2014). In contrast, we dissect, for the first time, the temporal evolution of drug resistance in M. tb, providing critical missing information as to how drug-resistant TB develops within an endemic setting and shedding light on a historical epidemic of XDR TB. Using whole genome sequencing-based comparative analyses, we show that:

Using whole genome sequence-based comparative analyses, we show that:

1. Mutations leading to resistance and MDR evolved prior to the HIV epidemic in South Africa, and were acquired at the beginning of the antibiotic era. These early mutations persisted for decades within M. tb populations, are still circulating today and account for a large fraction of today’s drug resistant TB.

2. MDR and XDR evolved via stereotypical patterns, which are not currently exploited by frontline molecular diagnostics.

3. Whole genome sequencing, combined with phylogenetic and parsimony analytical approaches, can illuminate novel putative compensatory mutations that would have been missed using previously published methods.
We believe that our findings have significant implications for programmatic and clinical management of TB, both in South Africa and beyond. We anticipate that these findings will be highly cited and would build upon PLOS Medicine’s strong tradition of publishing cutting-edge, high quality research relating to TB and drug resistance.

Each co-author has read and approved the manuscript. This information is entirely new and is not under review for publication elsewhere. None of the authors have conflicting commercial interests.

On behalf of my coauthors, I thank you for considering this paper.

Sincerely yours,

Ashlee M. Earl, Ph.D.
Group Leader, Bacterial Genomics
Genome Sequencing & Analysis Platform
The Broad Institute of MIT & Harvard
Evolution of extensively drug-resistant tuberculosis over four decades revealed by whole genome sequencing of *Mycobacterium tuberculosis* from KwaZulu-Natal, South Africa

Authors:
Keira A. Cohen MD1,2*, Thomas Abeel PhD3,4*, Abigail Manson McGuire PhD1, Christopher A. Desjardins PhD3, Vanisha Munsamy2, Terrance P. Shea2, Bruce J. Walker3, Nonkqubela Bantubani MPH2, Deepak V. Almeida PhD2,11, Lucia Alvarado3, Sinéad B. Chapman3, Nomonde R. Mvelase MD6,7, Eamon Y. Duffy2, Michael G. Fitzgerald3, Pamla Govender2, Sharvari Gujja3, Susanna Hamilton3, Clinton Howarth3, Jeffrey D. Larimer3, Kashmeel Maharaj2, Matthew D. Pearson3, Margaret E. Priest3, Qiandong Zeng PhD3, Nesri Padayatchi MD8, Jacques Grosset MD2,11, Sarah K. Young3, Jennifer Wortman3, Koleka Mlisana MD6,7, Max R. O'Donnell MD8,9,10, Bruce W. Birren PhD3, William R. Bishai MD, PhD11, Alexander S. Pym MD, PhD3‡ and Ashlee M. Earl PhD3†

Affiliations:
1. Division of Pulmonary and Critical Care Medicine, Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA
2. KwaZulu-Natal Research Institute for TB and HIV (K-RITH), Durban, South Africa
3. Broad Institute of MIT and Harvard, Cambridge, MA, USA
4. Delft Bioinformatics Lab, Delft University of Technology, Delft, The Netherlands
5. Medical Research Council, Durban, South Africa
6. School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal, Durban, South Africa
7. National Health Laboratory Service, Durban, South Africa
8. Centre for the AIDS Programme of Research in South Africa (CAPRISA), Durban, South Africa
9. Division of Pulmonary, Allergy, and Critical Care Medicine, Columbia University College of Physicians and Surgeons, New York, USA
10. Department of Epidemiology, Columbia Mailman School of Public Health, New York, USA
11. Center for Tuberculosis Research, Johns Hopkins School of Medicine, Baltimore, MD, USA

*These authors contributed equally
†These authors also contributed equally

Corresponding authors:
Dr. Ashlee M. Earl
The Broad Institute of MIT & Harvard
415 Main Street
Cambridge, MA 02142, USA
E-mail: aearl@broadinstitute.org
Phone: +1 617 714 7927

Dr. Alexander S. Pym
KwaZulu-Natal Research Institute for TB and HIV (K-RITH)
719 Umbilo Road
Berea, 4075, South Africa
E-mail: alex.pym@k-rith.org
Phone: +27 31 260 4991
Evolution of extensively drug-resistant tuberculosis over four decades revealed by whole genome sequencing of *Mycobacterium tuberculosis* from KwaZulu-Natal, South Africa

**Brief title:** Whole genome sequencing reveals 40-year evolution of epidemic XDR-TB

**ABSTRACT**

**Background:** The largest global outbreak of extensively drug-resistant (XDR) tuberculosis (TB) was identified in Tugela Ferry, KwaZulu-Natal, South Africa in 2005. Understanding the antecedents and timing of emergence of drug resistance in this fatal outbreak of XDR will inform the control and prevention of drug-resistant TB.

**Methods and Findings:** We performed whole genome sequencing and drug susceptibility testing on 337 clinical isolates of *Mycobacterium tuberculosis* (*M. tb*) collected in KwaZulu-Natal from 2008 to 2013, in addition to historical isolates, including one from the Tugela Ferry outbreak. We utilized an array of whole genome comparative techniques to assess the relatedness among strains, to establish the order of acquisition of drug resistance mutations including the timing of acquisitions leading to XDR-TB in the LAM4 spoligotype, and to calculate the number of independent evolutionary emergences of MDR and XDR. Our sequencing and analysis revealed a 50-member clone of XDR *M. tb* that was highly related to the Tugela Ferry XDR outbreak strain. We estimated that mutations conferring isoniazid and streptomycin resistance in this clone were acquired 50 years prior to the Tugela Ferry outbreak, with the subsequent emergence of multidrug resistance (MDR) and XDR occurring 30 and 40 years later, respectively. We observed frequent de novo evolution of MDR and XDR, with 56 and 9 independent evolutionary events, respectively. Isoniazid resistance evolved before rifampicin resistance 46 times, whereas rifampicin resistance evolved prior to isoniazid only twice. We identified additional putative compensatory mutations to rifampicin in this dataset.

**Conclusions:** In the first whole genome-based analysis of the emergence of drug resistance among clinical isolates of *M. tb*, we show that the ancestral precursor of the LAM4 XDR outbreak strain in Tugela Ferry gained mutations to first line drugs at the beginning of the antibiotic era. Subsequent accumulation of step-wise resistance mutations, occurring over decades and prior to the explosion of HIV in this region, yielded MDR and XDR, permitting the emergence of compensatory mutations. Our results suggest that drug-resistant strains circulating today reflect not only vulnerabilities of current TB control efforts, but also those that date back fifty years. In drug-resistant TB, isoniazid resistance was overwhelmingly the initial resistance mutation to be acquired, which would not be detected by current rapid molecular diagnostics employed in South Africa that assess only rifampicin resistance.
INTRODUCTION

The global burden of tuberculosis (TB) remains high with an estimated 9 million active disease cases and 1.5 million deaths in 2013 [1]. Multidrug-resistant (MDR) TB, defined as *Mycobacterium tuberculosis* (*M. tb*) with *in vitro* resistance to both isoniazid and rifampicin, accounted for at least 480,000 incident cases and 210,000 attributed deaths in 2013 [1]. Extensively drug-resistant (XDR) TB, which is MDR with additional resistance to both quinolones and second-line injectable agents [2], has been reported in 100 countries [1]. With prohibitively high morbidity, XDR poses a dire threat to public health, particularly in populations with high HIV prevalence [1,3].

The incidence of TB in South Africa is estimated by the WHO to be 860 (776 – 980) per 100,000 population, which is among the highest in the world [1]. With a population of approximately 10 million, KwaZulu-Natal is the easternmost of South Africa’s nine provinces. While its provincial TB incidence is similar to that of the rest of South Africa (889 per 100,000 in 2012, based on treatment initiation data) [4], KwaZulu-Natal has been notable for disproportionately high rates of drug-resistant TB [4,5]. Compounding this epidemic, South Africa has seen a dramatic increase in HIV prevalence in the last 25 years. UNAIDS estimates that national adult HIV prevalence was only 0.3% in 1990, but rose to 19.1% in 2013 [6]. In KwaZulu-Natal HIV rates are particularly high, with 37.4% HIV prevalence documented among pregnant women in 2011 [4].

In 2005, the identification of an outbreak of XDR-TB at the Church of Scotland Hospital in KwaZulu-Natal, Tugela Ferry, raised global alarm and called attention to the prospect of dissemination of potentially untreatable TB [7]. Not only was resistance to four or more classes of antibiotics noted in these strains, but also the disease, in the context of HIV co-infection, was rapidly fatal with 98% mortality [7]. Traditional genotyping by IS6110 fingerprinting and spoligotyping identified a predominant strain of the spoligotype, ST60, later termed LAM4/F15/KZN (henceforth referred to as LAM4), suggestive of a clonal outbreak [7–11]. Targeted sequencing of resistance mutations in a subset of these XDR strains revealed identical mutations [9], further supporting the theory of acquisition of XDR-level resistance and subsequent transmission. Nosocomial spread was deemed likely by a social network analysis [9].

Since the events at Church of Scotland Hospital, which still stands as the largest outbreak of XDR-TB ever reported, XDR-TB has been reported in the majority of hospitals across KwaZulu-Natal [12], and more than 516 XDR-TB cases have been reported within Tugela Ferry alone [13]. In addition, XDR-TB caused by strains not falling within the LAM4 spoligotype have been seen indicating repeated evolutionary emergences of XDR among strains circulating within the region [10,14]; however, the relative contribution of *de novo* versus vertically inherited resistance of XDR-TB is unknown. It is also unknown how and when XDR-level drug resistance developed, information that could be exploited to detect and prevent higher-level resistances from emerging in South Africa and elsewhere in the world.

Whole genome sequencing efforts that target large collections of *M. tb* have provided critical insights into *M. tb* population dynamics including *M. tb* transmission and the molecular causes of drug resistance [15–18]. Although some strains from KwaZulu-Natal have been sequenced [19,20], there has been no large-scale sequencing project from this province, nor studies that
have systematically addressed the molecular evolution of XDR. In the largest compilation of whole genome sequences from clinical isolates of *M. tb* from South Africa, we used a combination of comparative genomic techniques to elucidate when and how epidemic XDR drug resistance emerged and discuss the implications of these findings with respect to current and future TB control.

**Brief Methods:**

**Specimen Collection and Characterization**

We selected 337 clinical isolates of *M. tb* with diverse drug susceptibility patterns. Strains were collected both retrospectively and prospectively from 2008 to 2013 from all 11 districts of KwaZulu-Natal (Table S1). Biomedical Research Ethics Council (BREC) approval from the University of KwaZulu-Natal was granted for whole genome sequencing of clinical strains. On all study isolates, drug susceptibility testing (DST) by critical concentration was performed for rifampicin, isoniazid, streptomycin, kanamycin and ofloxacin (Tables S2 and S3). Extended DST was performed for key isolates (Table S2). Subject metadata included age, gender, AFB smear and HIV status, when available (Table S4). Study participants were assigned GPS coordinates corresponding to their home provincial district or site of sputum collection.

We also selected three historical strains previously collected in KwaZulu-Natal for re-sequencing [20,19]: KZN4207 (drug susceptible, collected in Durban in 1995), KZN1435 (MDR, collected in Durban in 1994), and KZN605 (XDR, collected in Tugela Ferry in 2005).

**Whole Genome Sequencing**

Genomic DNA was extracted using published methods [21]. The majority of strains were single colony selected prior to DNA isolation (Supplemental Methods and Table S4). Library preparation and whole genome sequencing (WGS) were performed as previously described on the Illumina HiSeq 2000 at the Broad Institute [22]. The median depth of sequencing was 143x and coverage of the H37Rv genome was 99.9%. Sequencing data were submitted to the Sequence Read Archive NCBI under the following umbrella BioProject identifiers: PRJNA183624 and PRJNA235615.

**Bioinformatic Analysis**

**Primary Analysis**

Reads were mapped onto a reference strain of H37Rv (GenBank accession number CP003248.2) using BWA version 0.5.9 [23]. In cases where read coverage of the reference was greater than 200x, reads were down-sampled using Picard [24] prior to mapping. Variants were identified using Pilon version 1.5 as described [22].

**Strain Diversity and Biogeography**

We conducted phylogenetic analyses for both the entire set of 340 strains, as well as for a subset of 111 strains belonging to the LAM4 spoligotype. For each set, all sites with unambiguous single nucleotide polymorphisms (SNPs) were used to generate a midpoint rooted phylogenetic tree in RAxML (version 7.3.3) [25] using a GTR+Gamma substitution model with 1,000 bootstrap replicates. Each strain’s spoligotype was predicted by statistically testing for the presence of each of 43 unique spacer sequences used in classical spoligotyping from sequence reads. Results were matched to spacer pattern profiles at SITVITWEB to generate a named
spoligotype (Supplemental Methods) [26]. Clonal strains were identified using a density based clustering algorithm [27] that groups strains that differ by no more than 10 SNPs to at least one other member within a clone [28–30](Supplemental Methods).

Mantel tests were performed to evaluate the relationship between genetic and geographic distances among strains using the ZT software v1.1 [31]. Pairwise genetic distances were calculated as the number of SNP differences between strains, and geographic distances were calculated using the haversine formula [32] and points of origin for strain pairs.

**Ordering and Dating Evolution of Drug Resistance**

A curated list of genomic polymorphisms associated with drug resistance was defined for each tested drug based on a literature review (Supplemental Methods). Polymorphisms associated with compensatory mechanisms to isoniazid, rifampicin and ethambutol were also defined (Supplemental Methods). Strains with predicted resistance were identified based on the carriage of mutations from the curated list. We used PAUP [33] to reconstruct the patterns of drug resistance mutation gains and losses throughout the phylogenetic tree representing all 340 strains. PAUP was run using a cost matrix that assigned a 10x greater cost for a loss event relative to a gain event. We used BEAST [34] to estimate a mutation rate and to determine dates for the acquisition of mutations within the LAM4 spoligotype. BEAST was run for 50 million iterations using the relaxed lognormal clock model, the GTR+Gamma substitution model and a starting value for the mean mutation rate of 0.35 SNPs/genome/year [28,30,35–37]. We assayed a range of values for the starting mean mutation rate, covering the range of values previously reported in the literature, with little difference in the output. Estimated dates are given with 95% confidence intervals as determined by the highest posterior density.

**RESULTS**

Our study included 337 participants with an average age of 33.8 ± 10.7 years, of whom 165 (49%) were male (Table 1). Overall, 140 patients were HIV positive, 51 were HIV negative, and 146 had unknown HIV status. Baseline characteristics were similar among HIV positive and negative individuals, with the exception that HIV negative individuals were younger (p = 0.0030), more likely to be smear positive (p = 0.0139) and to derive outside of eThekwini, the provincial capital (p = 0.0132).

A clinical sample was obtained from each patient; *M. tb* was isolated using standard approaches and phenotypic drug susceptibility testing was performed on each isolate using standard methodology (Supplemental Methods). Phenotypic DST revealed 88 susceptible, 23 mono-drug-resistant (defined as phenotypic resistance to only one drug), 19 poly-drug-resistant (defined as phenotypic resistance to two drugs that does not meet criteria for MDR), 140 MDR *sensu stricto*, and 67 XDR *M. tb* strains (Figure 1). Phenotypic MDR and XDR-TB cases were identified in all 11 districts of KwaZulu-Natal. While we observed a trend toward HIV negative individuals harboring more drug susceptible TB, this observation did not meet statistical significance (p = 0.0542).

**Table 1: Demographic characteristics of participants and phenotypic drug susceptibility of strains.** Data are n (%) or mean ± SD.
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>HIV positive (N = 140)</th>
<th>HIV negative (N=51)</th>
<th>p-value*</th>
<th>HIV unknown (N=146)</th>
<th>All patients (N=337)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (y); mean ± SD</td>
<td>34.0 ± 8.5</td>
<td>31.8 ± 13.4</td>
<td>0.0030</td>
<td>34.3 ± 11.5</td>
<td>33.8 ± 10.7</td>
</tr>
<tr>
<td>Male</td>
<td>65 (46)</td>
<td>30 (59)</td>
<td>0.1435</td>
<td>70 (48)</td>
<td>165 (49)</td>
</tr>
<tr>
<td>Smear status</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>smear positive</td>
<td>89 (64)</td>
<td>42 (82)</td>
<td>0.0139</td>
<td>51 (35)</td>
<td>182 (54)</td>
</tr>
<tr>
<td>smear negative</td>
<td>50 (36)</td>
<td>9 (18)</td>
<td>0.0209</td>
<td>24 (16)</td>
<td>83 (25)</td>
</tr>
<tr>
<td>smear unknown</td>
<td>1 (1)</td>
<td>0 (0)</td>
<td>1.0000</td>
<td>71 (49)</td>
<td>72 (21)</td>
</tr>
<tr>
<td>Phenotypic DST results</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Susceptible</td>
<td>39 (28)</td>
<td>22 (43)</td>
<td>0.0542</td>
<td>27 (18)</td>
<td>88 (26)</td>
</tr>
<tr>
<td>Mono-drug resistant</td>
<td>7 (5)</td>
<td>2 (4)</td>
<td>1.0000</td>
<td>14 (10)</td>
<td>23 (7)</td>
</tr>
<tr>
<td>Poly-drug resistant</td>
<td>4 (3)</td>
<td>0 (0)</td>
<td>0.5751</td>
<td>15 (10)</td>
<td>19 (6)</td>
</tr>
<tr>
<td>MDR sensu stricto</td>
<td>54 (39)</td>
<td>16 (31)</td>
<td>0.3996</td>
<td>70 (48)</td>
<td>140 (42)</td>
</tr>
<tr>
<td>XDR</td>
<td>36 (26)</td>
<td>11 (22)</td>
<td>0.7044</td>
<td>20 (14)</td>
<td>67 (20)</td>
</tr>
<tr>
<td>Tugela Ferry XDR</td>
<td>27 (75)</td>
<td>8 (73)</td>
<td>0.6750</td>
<td>14 (70)</td>
<td>49 (73)</td>
</tr>
<tr>
<td>other XDR</td>
<td>9 (25)</td>
<td>3 (27)</td>
<td>1.0000</td>
<td>6 (30)</td>
<td>18 (27)</td>
</tr>
<tr>
<td>Geography</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>eThekwini</td>
<td>73 (52)</td>
<td>36 (71)</td>
<td>0.0312</td>
<td>69 (47)</td>
<td>178 (53)</td>
</tr>
</tbody>
</table>

*HIV positive and HIV negative individuals were compared using Fisher’s exact test for categorical variables and non-parametric Mann-Whitney test for continuous variables. Results having a p value <0.05 were considered statistically significant.

We performed WGS on all 337 clinical strains, as well as three historical strains isolated prior to the study collection period. We assessed the diversity and phylogenetic relatedness among strains using information from 17,232 variable sites with single nucleotide polymorphisms (SNPs) relative to the H37Rv reference genome (Figure 1; Figure S1). The resulting phylogenetic tree revealed four of the seven previously described main global lineages of M. tb circulating in KwaZulu-Natal during the sampling time frame [38–40]. The vast majority of isolates (95%) belonged to lineages 2 and 4, with lesser representation from lineages 1 and 3. A computational spoligotype prediction was performed and 17 unique spoligotypes were identified [26] (Table S4).

**Figure 1: Diverse strains contribute to drug resistance in KwaZulu-Natal**
Midpoint rooted maximum-likelihood phylogeny of 340 M. tb isolates. Four of the seven known M. tb lineages were identified: CAS (Lin1), Beijing (Lin2), Euro-American (Lin3), and EAI (Lin4). Digital spoligotyping identified 17 unique spoligotypes in the dataset;
spoligotypes are shown on this figure if they are represented by three or more strains. Corresponding spoligotypes and phenotypes are reported for all strains in Table S4. Phenotypic XDR, MDR, poly- and mono-drug resistance (labeled ‘Drug Resistant other’) and pan-susceptible strains are indicated by colored tick marks at the tip of each leaf node.

We defined a ‘clone’ as a set of strains in which each member differs by no more than 10 SNPs to at least one other member, which is similar to definitions used in previous genomic studies of M. tb transmission [28–30] (Figure S2, Supplemental Methods). Nearly one third of strains (107 of 340, 31%) belonged to 11 such clones (Table S5), which were distributed across six spoligotypes and three lineages (Figure S3). All clones were phenotypically drug-resistant, indicating recent person-to-person spread of a diverse set of drug-resistant strains that included both HIV positive and HIV negative individuals. Using a Mantel test, we determined that there was very low correlation between geographic and genetic distances among strains (r = - 0.067906, p = 0.001760) indicating that strains did not cluster geographically. Older transmission events and/or high patient mobility between districts likely account for this pattern (Figure S4).

The ‘historical’ Tugela Ferry XDR strain, KZN605, was nested phylogenetically within a large clone of 50 LAM4 strains with predominantly phenotypic XDR (Figure 2). All of the strains within this clone (henceforth referred to as the Tugela Ferry XDR Clone) possessed the characteristic drug resistance mutations that were previously identified in XDR-TB strains circulating in Tugela Ferry during the outbreak [9,19], further indicating this clone’s continued prevalence within KwaZulu-Natal. Patients in whom the Tugela Ferry XDR Clone was isolated were from 10 of the 11 districts within the province (Figure 3). In addition, the Tugela Ferry XDR Clone was not overrepresented among HIV-positive patients, (p = 0.6750) (Table 1), suggesting that strains within this clone were neither geographically constrained nor restricted to immuno-deficient hosts.

**Figure 2: Molecular evolution and dating of drug resistance emergence within the Tugela Ferry XDR Clone**

Midpoint rooted maximum-likelihood phylogeny of 107 M. tb isolates of the LAM4 spoligotype. The gray shaded box identifies the Tugela Ferry XDR Clone. KZN605, the historical XDR strain collected in Tugela Ferry during the outbreak, is a member of this clone. Two additional historical isolates, KZN1435 and KZN4207, are not members of the Tugela Ferry XDR Clone. Each evolutionary gain of a drug resistance mutation was assigned to its position on the phylogenetic tree by parsimony (colored circles). A-E traces the stepwise order of drug resistance acquisition in the Tugela Ferry XDR Clone and estimates the year when each mutation was gained. A) katG S315T (isoniazid); gidB 130 bp deletion (streptomycin); 1957 (95% CI: 1937-1971); B) inhA promoter -8 (isoniazid and ethionamide); 1964 (95% CI: 1948-1976); C) embB M306V (ethambutol); 1967 (95% CI: 1950-1978); D) rpoB L452P (rifampicin); pncA 1bp insertion (pyrazinamide); 1984 (95% CI: 1974-1992); and E) rpoB D435G (rifampicin); rrs 1400 (kanamycin); gyrA A90V (ofloxacin); 1995 (95% CI: 1988-1999). The accumulation of individual drug-resistant mutations within a strain is denoted to the right of the phylogenetic tree. Four additional LAM4 strains on a distant branch were not included in this figure due to size constraints. Bootstrap values are provided for lettered nodes, and bootstrap values for all nodes are shown in Figure S5.
Figure 3: Wide geographic spread of XDR and the Tugela Ferry XDR Clone members

Pie charts indicate the fraction of sequenced XDR *M. tb* from each of the 11 districts of KwaZulu-Natal. The size of the pie chart indicates the relative number of strains sequenced from each of the 11 districts within KwaZulu Natal. Tugela Ferry, in the uMzinyathi district, is indicated in red.

Many of the sequenced LAM4 strains were closely related to the Tugela Ferry XDR Clone, but had different DST profiles (Figure 2, Figure S2 and Table S6) giving us an opportunity to finely dissect the order of acquisition of mutations giving rise to the Tugela Ferry XDR Clone [9,19]. LAM4 strain phylogeny was recalculated using data from only LAM4 strains and parsimony was used to place the origin of known resistance-conferring mutations on the tree. The recalculated LAM4 tree was consistent with our previous tree containing data from all strains and was well supported by bootstrapping analysis. This was particularly true for internal nodes representing more distant evolutionary events, which enabled us to confidently assign evolutionary ordering of drug resistance mutation acquisition (Figures 2 and S5).

As shown in Figure 2, the first step towards XDR-level resistance in this epidemic clone was the acquisition of isoniazid and streptomycin resistance-conferring mutations in *katG* and *gidB*, respectively, which were gained at node A of the phylogenetic tree (100% bootstrap support). With accumulation of successive mutations, the ancestral strain (and its descendants) gained i) additional poly-drug resistance to ethionamide and ethambutol via mutations in the *inhA* promoter and *embB* (nodes B and C, respectively, 100% and 89%); ii) MDR via mutations in *rpoB* and *pncA* that conferred resistance to rifampicin and pyrazinamide (node D, 100%); and iii) XDR via mutations in *rrs* and *gyrA*, which conferred resistance to kanamycin and ofloxacin, respectively, and an additional *rpoB* mutation (node E, 97%). This ordering was highly supported by bootstrapping in the phylogenetic reconstruction. Thus, the first step towards XDR-level drug resistance in this epidemic clone was the acquisition of isoniazid and streptomycin resistance followed by ethambutol and ethionamide resistance, then rifampicin and pyrazinamide resistance, and, ultimately, kanamycin and ofloxacin resistance.

Because we had dates of isolation for all sequenced strains—including strains that were isolated more than 20 years ago—we applied a Bayesian statistical approach to estimate when mutations leading to the Tugela Ferry XDR Clone emerged. Using this approach, which takes into account the phylogeny of LAM4 strains, the dates of their isolation and published mutation rates for *M. tb* [28,30,35–37], we calculated that LAM4 in KZN mutated at a rate of 0.61 SNPs/genome/year. This mutation rate was higher than other previously published mutation rates, regardless of which literature prior was used as the starting mean mutation rate. Applying this rate, we estimated that drug resistance mutations at node A were acquired in 1957 (95% CI: 1937 – 1971), soon after the introduction of streptomycin and isoniazid into clinical practice. MDR-level resistance was acquired in 1984 (95% CI: 1974 – 1992; node D) and XDR-level resistance was acquired in 1995 (95% CI: 1988-1999; node E), ten years prior to its acute recognition in 2005 in Tugela Ferry (Figure 2).

We also observed multiple drug resistance mutations within LAM4 that emerged outside the Tugela Ferry XDR Clone (Figure 2 and Table S6). Many of these mutations were acquired at
leaf nodes, which implied very recent gains of resistance. Including the Tugela Ferry ancestor, we calculated that genotypic MDR *sensu stricto*—defined as both isoniazid and rifampicin resistance-conferring mutations—indeed very recently arose a minimum of 13 times. Within LAM4, the Tugela Ferry XDR Clone represented the single and only evolutionary gain of genotypic XDR—as defined by acquisition of resistance-conferring mutations to the four XDR-defining drugs: isoniazid, rifampicin, ofloxacin and kanamycin. However, within LAM4 we also observed 10 independent gains of either a kanamycin or an ofloxacin resistance-conferring mutation in a background of genotypic MDR *sensu stricto*. As such, thirteen LAM4 strains identified in this study would be considered genotypic “pre-XDR” and only one SNP away from XDR-level resistance.

Beyond LAM4, we observed many other independent evolutionary emergences of MDR and XDR across this dataset. Twelve and seven spoligotypes contained strains with phenotypic MDR and XDR, respectively (Table S7), suggesting that these resistance patterns emerged no fewer than twelve and seven times. However, when we quantified the total number of independent evolutionary emergences of genotypic MDR and XDR across our entire dataset, we estimated that MDR *sensu stricto* and XDR evolved no less than 56 and 9 independent times, respectively (Table S7).

Remarkably, the first drug resistance acquisition in the Tugela Ferry XDR Clone was consistent with other emergences of MDR and XDR across the entire dataset. For the 214 strains with genotypic resistance to two or more of the MDR and XDR defining drugs, we quantified the number of evolutions in which a specific drug resistance mutation was gained before a second resistance mutation. We observed that isoniazid resistance via non-synonymous mutation at the *katG* S315 codon was gained before rifampicin resistance in 46 unique evolutionary events, whereas rifampicin resistance was never acquired before the *katG* S315 mutation (Table 2). When we repeated this for all pairwise comparisons, we found that isoniazid resistance, conferred by mutation of the *katG* S315 codon, preceded or co-occurred with resistance mutations to all other drugs in our dataset. Mutations other than the *katG* S315 mutations that confer isoniazid resistance (*i.e.* *inhA* promoter mutations or *katG* deletions) occurred before rifampicin resistance mutations in nine unique events, whereas we only observed the reverse ordering twice. These data indicate that, beyond the Tugela Ferry XDR Clone, isoniazid resistance, and in particular the S315 codon mutation in *katG*, has been the initial resistance-conferring mutation leading to poly-drug resistance, including MDR and XDR, among strains from KwaZulu-Natal.
<table>
<thead>
<tr>
<th>First resistance</th>
<th>Second resistance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>katG S315</td>
</tr>
<tr>
<td>katG S315</td>
<td>-</td>
</tr>
<tr>
<td>Other INH</td>
<td>0</td>
</tr>
<tr>
<td>Rifampicin</td>
<td>0</td>
</tr>
<tr>
<td>Kanamycin</td>
<td>0</td>
</tr>
<tr>
<td>Ofloxacin</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2: Isoniazid resistance is the first step towards drug resistance

Acquisition of katG S315 mutations precedes all other resistance mutations, including rifampicin, in all instances in which the order of acquisition can be disambiguated. For the 214 strains with genotypic resistance to two or more MDR or XDR defining drugs, and in which the order of acquisition of these mutations could be disambiguated, we quantified the number of evolutions in which resistance to one drug was gained before resistance to a second drug. Isoniazid resistance was divided into mutations conferred by the katGS315 codon vs. “Other INH” mutations (defined as loss of function mutations in katG that do not involve codon 315 or mutations in the inhA promoter). Reported numbers represent the number of independent evolutionary events (not the number of strains) in which the drug resistance indicated by the row labeled (“First resistance”) was acquired before the drug resistance indicated by the column labeled (“Second resistance”). The background color is shaded to indicate the fraction of unambiguous evolutionary events in which the “First resistance” was acquired before the “Second resistance” for that given drug pair.

As in other organisms, in vitro studies have suggested that drug resistance in M. tb may be associated with a variable fitness cost that can be offset by compensatory evolution [41,42]. Non-synonymous mutations in the α and β′ subunits of RNA polymerase have been postulated to compensate for fitness costs associated with rifampicin resistance [42,43]. Among our 226 strains with phenotypic rifampicin resistance, 76 strains had mutations known to compensate for rifampicin resistance [17,43–45] (Table S8). Using the phylogenetic framework and parsimony, we determined that 23 of the 27 previously described compensatory mutations had an evolutionary pattern consistent with rifampicin compensation i.e., mutations that evolved only after or concurrent with mutations that conferred rifampicin resistance.

We also attempted to identify novel rifampicin compensatory mutations with this approach. In addition to the 27 previously described mutations, we detected an additional 38 non-synonymous polymorphisms in rpoA, rpoC, and the non-rifampicin resistance-determining regions (RRDR) regions of rpoB (Table S8). By parsimony analysis, we established the acquisition order of these rpoA, rpoC, and non-RRDR rpoB mutations in relation to genotypic rifampicin resistance. An additional 26 of these previously uncharacterized mutations also evolved in a pattern consistent with a role in compensation, which suggests that they may also function in this capacity. While there were 10 unique RRDR mutations with subsequent or concurrent evolutionary gain of a putative rifampicin compensatory polymorphism, the vast majority of putative compensatory mutations occurred in association with rpoB S450L (p = 0.000032) (Table S9). This pattern was
observed regardless of whether the compensatory mutation was previously known or uncharacterized.

Beyond rifampicin compensation, we also applied our combined phylogenetic and parsimony approach to known isoniazid and ethambutol compensatory mutations. With respect to isoniazid compensation, only a single evolution of the *ahpC* promoter mutation was observed in our dataset (Table S10). It was gained after genotypic isoniazid resistance, which supports its compensatory role for certain isoniazid resistance mechanisms [41]. Non-synonymous mutations in *ubiA* (Rv3086c) have previously been implicated in ethambutol resistance [46]. In our dataset, there were at least two occasions in which these mutations unambiguously arose prior to the acquisition of genotypic ethambutol resistance, suggesting that these are more likely to be stepping-stone mutations rather than compensatory (Table S10).

**DISCUSSION**

We report on the whole genome sequencing and comparative analysis of the largest collection of drug-resistant *M. tb* sequenced to date from South Africa. From analysis of these genomes, we determined the molecular antecedents of the Tugela Ferry XDR Clone and dated the emergence of genotypic resistance to eight drugs. We showed that the development of XDR in KwaZulu-Natal had its roots in first-line drug resistance that arose in the late 1950’s, and MDR that emerged in the 1980’s. Our dating analysis indicated that the Tugela Ferry XDR Clone took nearly four decades to evolve from its initial isoniazid and streptomycin resistances to full blown XDR. Although our data confirmed that the XDR outbreak in KwaZulu-Natal was indeed a clonal event, we showed that drug resistance in this region is driven by both the development of *de novo* drug resistance as well as clonal spread. We elucidated common evolutionary patterns of drug resistance acquisition and determined that isoniazid was overwhelmingly the first drug resistance to be acquired. Lastly, we validated that certain previously described rifampicin compensatory mutations do indeed evolve in a pattern consistent with compensation, and have identified 26 novel polymorphisms that may also function in this capacity. Collectively, these data have important implications for the public health control of TB in sub-Saharan Africa and elsewhere.

Using a combination of likelihood, parsimony and Bayesian computational approaches, we observed a decades-long evolutionary trajectory toward XDR-level drug resistance in LAM4 that mirrored the order and timing of introduction of antitubercular drugs into clinical practice [47]. Though parsimony-based approaches can interpret rapid independent evolution of an identical polymorphism in multiple strains as a single evolutionary event (occurring at a single node), our predictions indicated that resistance-conferring mutations evolved only after each drug’s clinical introduction and not before, as might be expected if homoplasy were a major contributor to pattern predictions (Table S11). In addition, one of the oldest acquired mutations toward XDR-level resistance was a specific 130 bp deletion in *gidB*, which is extremely unlikely to arise many times independently and supports accurate reconstruction in our evolutionary analysis. Furthermore, though our calculated mutation rate for LAM4 was slightly higher than previous reports [28,30,35–37], our estimate was within reported confidence intervals and was based on a larger fraction of the H37Rv genome than previous studies (99.9% versus <90% H37Rv mapping coverage) [28]. This was due to the inclusion of sequence data generated from both PCR-free short fragment and jumping libraries and analysis with improved bioinformatics tools.
that enabled us to examine SNPs within more variable and high GC content regions of the genome, including PE and PPE genes that have been reported to have a higher mutation rate [48]. Thus, because we are including data from more of the genome, our estimation of the M. tb mutation rate may more closely approximate the actual mutation rate of the organism as compared to previously published studies.

Importantly, from the pattern of drug resistance evolution within LAM4 it is clear that the precursors to XDR evolved well before the explosive South African HIV epidemic of the 1990’s, indicating that the selection of transmissible XDR strains can occur in low prevalence HIV settings. While recent failures in TB and infection control and the current high HIV prevalence rates, combined, undoubtedly contributed to the spread of XDR, they were not the sole causes of XDR in this setting. Indeed, strains that evolved first-line drug resistance soon after the introduction of chemotherapy were a critical entry point to today’s drug-resistant epidemic.

Drug-resistant strains that emerged from the mid 20th Century were evidently maintained within the population of M. tb, presenting the opportunity for the acquisition of successive resistance and compensatory mutations that culminated in transmissible XDR and the Tugela Ferry outbreak. Drug-resistant strains may have been maintained within a population over time either by ongoing cycles of infection and transmission or through reactivation of latent disease. It is unclear which of these may have been the most important in this setting, but it suggests that fitness costs due to first-line drug resistance may not be severe.

Beyond LAM4, and as has been shown in other studies [10,14] drug resistance emerged de novo repeatedly in KwaZulu-Natal as we identified numerous independent evolutionary events of MDR and XDR across multiple lineages and spoligotypes. Of particular note was the detection of multiple independent evolutions of MDR to pre-XDR within LAM4, which may herald a new wave of XDR in the near future. Thus, the repeated emergence of de novo high-level drug resistance underscores the reality that, even in middle-income sub-Saharan African countries, the current approach to TB control is failing to stem the ongoing emergence of drug resistance. In fact, results from our analyses suggest this was not due to infrequent poor adherence to TB drugs, but instead decades of inadequate TB control that has driven resistance development in a stepwise fashion, multiple times over. Given that our estimates of resistance evolution were based on identification of known resistance-conferring mutations and that the majority of sequenced strains were single colony purified, our calculations are likely an underestimation due to incomplete understanding of all mutations that confer drug resistance and the possibility of mixed infections, respectively. Thus, the state of drug resistance emergence is likely more dire than we have described.

Recent studies from KwaZulu-Natal have emphasized transmission of a limited number of strains as a driving force behind the emergence of drug resistance [10,14]. Our data also confirm that once drug resistance develops, clonal spread of resistant strains can and does occur in this context. We found that recent person-to-person spread of resistant strains is apparent in KwaZulu-Natal, as evidenced by identification of multiple drug-resistant clones. Importantly, in contrast to initial reports from Tugela Ferry where nearly all XDR cases were TB/HIV co-infected [7,9], eight patients in our study in whom the Tugela Ferry XDR Clone was identified were HIV negative. This reemphasizes that even XDR drug-resistant strains are sufficiently fit to transmit person-to-person and cause morbidity in both immunocompetent and
immunosuppressed persons. Improved infection control and rapid case finding will be necessary to prevent further spread of drug-resistant strains and to detect such cases in the community as well as in hospital settings [49].

Our genomic analysis uncovered a common initial pattern of drug resistance that is not optimally detected by current diagnostic algorithms. Isoniazid resistance was overwhelmingly the first drug resistance to occur along the pathway to multiple drug resistances. However, current TB control strategies in South Africa focus on early detection of rifampicin resistance as a surrogate marker of MDR, and do not include the detection of isoniazid resistance. Clinical diagnostic policies that rely on Xpert MTB/RIF (a WHO-endorsed and widely deployed molecular diagnostic) [50] without more extensive drug resistance testing allow isoniazid resistance to go undetected and unchecked. Moreover, under current short course treatment guidelines which utilize four months of isoniazid and rifampicin in the continuation phase [51], failure to recognize isoniazid mono-resistance is tantamount to provision of unopposed rifampicin therapy and may rapidly select for rifampicin resistance. This phenomenon may be underappreciated and incompletely accounted for in mathematical models that recommend continued use of screening tools that identify only rifampicin resistance [52]. Our ordering of drug resistance acquisition provides strong evidence that isoniazid mono-resistance is a common pathway toward development of MDR and highlights the importance of prompt identification and treatment of isoniazid mono-resistance. Failure to do so would be recapitulating the scenario that led to the current XDR problem.

Beyond detection, identification of the initial drivers of isoniazid mono-resistance is also critical to the prevention of successive resistances. Isoniazid preventive therapy (IPT) has previously been implicated as a potential source of isoniazid mono-resistance [53,54]. Our work highlights the need to understand the true risks of mass IPT implementation [55] in high-burden settings.

We were able to verify that the evolutionary patterns of select previously described rifampicin and isoniazid compensatory mutations do indeed appear to be consistent with compensation to their respective drug. Similarly, ubiA was observed to evolve in a stepping-stone pattern rather than a compensatory pattern with respect to ethambutol resistance [46]. Furthermore, we have identified novel putative rifampicin compensatory mutations that may have acted to restore bacterial fitness and facilitate transmission of drug-resistant strains. While the majority of the previously described rifampicin compensatory mutations had an evolutionary pattern consistent with this role, four polymorphisms previously associated with rifampicin compensation (\(rpoB\) I491F, \(rpoC\) G594E and N826K, and \(rpoA\) E319K) were not observed to evolve concurrently or subsequent to genotypic rifampicin resistance (Table S8). These mutations may i) not be compensatory mutations in the classic sense (i.e., mutations that evolve following gain of genotypic drug resistance to mitigate a fitness cost) but instead serve as stepping-stone mutations; ii) evolve in concert with non-RRDR genotypic rifampicin resistance; or iii) have no association with rifampicin resistance. We have proposed 26 novel mutations whose evolutionary patterns are consistent with rifampicin compensation, and these should be investigated in future studies.

The most commonly observed genotypic rifampicin resistance mutation among our sequenced strains was \(rpoB\) S450L (often referred to as S531L using the \(E. coli\) codon numbering scheme), which is known to be the most prevalent RRDR mutation. Laboratory-derived strains carrying
the S450L were previously shown to have relatively high fitness in *in vitro* growth assays [42], supporting the hypothesis that high prevalence of the S450L mutation among clinical strains was due to it imparting few fitness consequences. However, as shown in our study and in several others [17,44], *rpoB* S450L was the most likely RRDR polymorphism to evolve putative compensatory mutations, which calls into question the low fitness cost of S450L *in vivo*. Song *et al.* assessed rifampicin fitness by transcriptional efficiency (rather than growth) and showed that the S450L mutation has half the transcriptional efficiency of WT *rpoB* [44], which is likely to impart fitness consequences if not compensated.

Here, we present the largest WGS study conducted to date of drug-resistant clinical isolates of *M. tb* from South Africa. Our dating analysis highlights the dire repercussions of failure to control first-line drug resistance. As acquisition of isoniazid resistance is the key initiation event for progression to MDR and beyond, TB control efforts that focus on the identification of isoniazid as well as rifampicin resistance will result in earlier detection of drug-resistant TB cases. Prudent antibiotic stewardship during the introduction of new antitubercular drugs will be critical to prevent the early fixation of resistance and protect the lifespan of novel agents.

**Supporting Information**

**Supplemental Methods**

**Figure S1.** Bootstrap values for phylogenetic tree of 337 strains (1000 bootstrap replicates).

**Figure S2.** *Defining clones with varying SNP thresholds.* The numbered columns to the right of the phylogenetic tree represent varying SNP thresholds used to define a clone. Strains that would be considered clonal by the SNP threshold listed in the column header are indicated by a unique three-letter code. By the 10 SNP threshold, the Tugela Ferry XDR Clone (labeled 10-AAI, shaded in gray) contains 50 members.

**Figure S3.** Drug-resistant clones are distributed widely across the phylogenetic tree. Columns to the right of the phylogenetic tree represent phenotypic DST (as indicated by the colored square), clones defined at 10 SNP threshold as shown in Table S5 and Figure S2, and HIV status of sampled patient.

**Figure S4:** *Wide geographic distribution of diverse strains across KwaZulu-Natal.* Pie charts indicate the fraction of sequenced *M. tb* belonging to computationally predicted spoligotypes in each of the 11 districts within KwaZulu-Natal. The size of the pie chart indicates the relative number of strains sequenced from each of the 11 districts. Tugela Ferry, in the uMzinyathi district, is indicated in red.

**Figure S5.** Bootstrap values for phylogenetic tree of lam4 isolates shown in Figure 2 (1000 bootstrap replicates)

**Table S1: Cohorts**

337 clinical isolates derived from five cohorts of strains that were both prospectively and retrospectively collected from all 11 districts of KwaZulu-Natal from 2008 to 2013.[49,56]

**Table S2: Drug Susceptibility Testing**

Culture and DST for each cohort are described below.

**Table S3: Critical Concentrations**

DST was performed by the critical concentration method, using the WHO recommended critical concentrations (in µg/mL) listed below.[57] Pyrazinamide resistance testing was performed using PZA MGIT concentration of 100.0µg/mL or Nicotinamide critical concentration assessment at 500.0µg/mL.
Table S4: Participant Metadata

Each participant was assigned a strain number with the header Tuberculosis KwaZulu-Natal K-RITH (TKK). Metadata for each participant included age, sex, HIV status (if known), date of specimen collection, geographic district (where participant lived or where specimen was collected), specimen type and smear status (if known). DNA isolation technique via single colony isolation or non-single colony selection is denoted. Drug susceptibility testing results are reported for each tested drug as susceptible (S), resistant (R), or untested, (U). Genomic spoligotyping and lineage that were derived from the sequencing data are listed. Lastly, genotypic drug susceptibility prediction and membership in the Tugela Ferry Clone are reported.

Table S5: Identification of drug-resistant clones indicates recent person-to-person transmission of drug-resistant TB

A linkage analysis identified 11 drug-resistant clones in the entire dataset. The largest clone contained 50 members of the LAM4 spoligotype; this spoligotype was subsequently identified as the Tugela Ferry XDR clone. Within the LAM4 spoligotype, there were three additional clones identified, and clones were also identified in five other spoligotypes. All clone members were noted to be drug-resistant, indicating recent person-to-person transmission of drug-resistant TB. See Supplemental Methods for definition of a clone.

Table S6: Drug resistance mutations identified within the LAM4 spoligotype

Drug resistance mutations were identified within the LAM4 strains as a first step in determining their order of acquisition by parsimony analysis in Figure 2. Isolates that were found to contain a resistance mutation that conferred resistance to a given drug are indicated in color. Wild type loci are indicated with a grey dot. Ordering of strains is according to their position on the phylogenetic tree in Figure 2. M. tb codon numbering was utilized for all polymorphisms, including rpoB.

Table S7: Diverse drug-resistant strains and frequent de novo development of drug resistance

Drug-resistant strains belonged to many distinct spoligotypes, which highlights the diversity of the drug resistance epidemic in this region. With a parsimony-based analysis we quantified the independent evolutionary gains of genotypic MDR and XDR in our 340-strain dataset.

Table S8: Putative rifampicin compensatory mutations were identified in rpoA, rpoC and non-RRDR regions of rpoB. Polymorphisms were deemed consistent with compensatory mutations when they evolved after or concurrent to genotypic rifampicin resistance. Many previously described putative compensatory mutations occurred in this evolutionary pattern, and 26 novel polymorphisms were newly described.

Table S9: Distribution of putative rifampicin compensatory mutations across the RRDR

The vast majority of rpoA, rpoC, and non-RRDR rpoB mutations that evolved with an evolutionary pattern consistent with rifampicin compensation evolved in association with rpoB S450L.

Table S10: Ordering of acquisition of polymorphisms with respect to genotypic resistance.

ahpC and ubiA mutations were ordered with respect to genotypic isoniazid and ethambutol resistance, respectively.

Table S11: Drug resistance emergence in the region mirrored the dates of drug discovery

Dating analysis within the LAM4/F15/KZN spoligotype consistently assigned drug resistance gains after the dates of drug discovery.[58]

References


57. WHO Updated interim critical concentration for first-line and second-line DST [Internet]. Available: http://www.stoptb.org/wg/gli/assets/documents/Updated critical concentration table_1st and 2nd line drugs.pdf
Figure S1
Click here to download Supporting Information: S1_figure.pdf
Figure S5
Click here to download Supporting Information: S5_figure.pdf
Table S3
Click here to download Supporting Information: S3_Table.pdf
Click here to download Supporting Information: S4_Table.pdf
Table S5

Click here to download Supporting Information: S5_Table.pdf
Table S6
Click here to download Supporting Information: S6_Table.pdf
Table S8

Click here to download Supporting Information: S8_Table.xlsx
Table S9
Click here to download Supporting Information: S9_Table.xlsx
Table S11
Click here to download Supporting Information: S11_Table.pdf