

# Lawrence Berkeley National Laboratory

## Joint Genome Institute

### Title

Mycobacterium tuberculosis Whole Genome Sequences From Southern India Suggest Novel Resistance Mechanisms and the Need for Region-Specific Diagnostics

### Permalink

<https://escholarship.org/uc/item/3b91d48x>

### Journal

Clinical Infectious Diseases, 64(11)

### ISSN

1058-4838

### Authors

Manson, Abigail L

Abeel, Thomas

Galagan, James E

et al.

### Publication Date

2017-06-01

### DOI

10.1093/cid/cix169

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NonCommercial-NoDerivatives License, available at

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

Peer reviewed

# *Mycobacterium tuberculosis* Whole Genome Sequences From Southern India Suggest Novel Resistance Mechanisms and the Need for Region-Specific Diagnostics

Abigail L. Manson,<sup>1,a</sup> Thomas Abeel,<sup>1,2,a</sup> James E. Galagan,<sup>3,4,a</sup> Jagadish Chandrabose Sundaramurthi,<sup>5</sup> Alex Salazar,<sup>1,2</sup> Thies Gehrman,<sup>2</sup> Siva Kumar Shanmugam,<sup>5</sup> Kannan Palaniyandi,<sup>5</sup> Sujatha Narayanan,<sup>5</sup> Soumya Swaminathan,<sup>5,b</sup> and Ashlee M. Earl<sup>1,b</sup>

<sup>1</sup>Broad Institute of MIT and Harvard, Cambridge, Massachusetts; <sup>2</sup>Delft Bioinformatics Lab, Delft University of Technology, The Netherlands; <sup>3</sup>Department of Biomedical Engineering, and <sup>4</sup>National Emerging Infectious Diseases Laboratory, Boston University, Massachusetts; and <sup>5</sup>National Institute for Research in Tuberculosis, New Delhi, India

**Background.** India is home to 25% of all tuberculosis cases and the second highest number of multidrug resistant cases worldwide. However, little is known about the genetic diversity and resistance determinants of Indian *Mycobacterium tuberculosis*, particularly for the primary lineages found in India, lineages 1 and 3.

**Methods.** We whole genome sequenced 223 randomly selected *M. tuberculosis* strains from 196 patients within the Tiruvallur and Madurai districts of Tamil Nadu in Southern India. Using comparative genomics, we examined genetic diversity, transmission patterns, and evolution of resistance.

**Results.** Genomic analyses revealed (11) prevalence of strains from lineages 1 and 3, (11) recent transmission of strains among patients from the same treatment centers, (11) emergence of drug resistance within patients over time, (11) resistance gained in an order typical of strains from different lineages and geographies, (11) underperformance of known resistance-conferring mutations to explain phenotypic resistance in Indian strains relative to studies focused on other geographies, and (11) the possibility that resistance arose through mutations not previously implicated in resistance, or through infections with multiple strains that confound genotype-based prediction of resistance.

**Conclusions.** In addition to substantially expanding the genomic perspectives of lineages 1 and 3, sequencing and analysis of *M. tuberculosis* whole genomes from Southern India highlight challenges of infection control and rapid diagnosis of resistant tuberculosis using current technologies. Further studies are needed to fully explore the complement of diversity and resistance determinants within endemic *M. tuberculosis* populations.

**Keywords.** Indo-Oceanic lineage; EAI lineage; CAS lineage; India; drug resistance.

India has the largest number of tuberculosis patients worldwide, accounting for over 25% of cases [1]. Drug resistant forms of *Mycobacterium tuberculosis*, the causative agent of tuberculosis, are rapidly spreading in India. These contribute to India having the second highest number of multi-drug resistant cases of tuberculosis (MDR-TB) [1], resistant to both isoniazid and rifampicin. Despite the presence of over 2 million tuberculosis cases in India, little is known about the genetic diversity [2] and drug resistance determinants of Indian *M. tuberculosis* [3].

India stands out as being home to primarily *M. tuberculosis* lineages 1 (Indo-Oceanic or EAI lineage) and 3 (Central

Asian or CAS lineage) [4], which occur at substantially lower frequency elsewhere. Lineages 2 (East Asian or Beijing) and 4 (Euro-American) are most common in Europe, Africa, and many other parts of the world [4–6]. Within India, lineage 3 predominates in the North and Northwest, whereas lineage 1 is common in the South but is found at low frequency in other parts of the country [3, 7–9]. In contrast, lineage 2 has been reported at similar prevalence throughout India (17%), though it does predominate in some Northeast states [7].

Drug-resistant *M. tuberculosis* poses a threat because of low cure rates. In India, the rate of MDR-TB is 2–3% among new cases and 12–17% among reinfections [10] (similar to the global averages of 3.3% and 20%, respectively [1]). Further, about 4% of tuberculosis patients have XDR-TB [1], defined as MDR-TB with additional resistance to any fluoroquinolone and at least 1 of the second-line injectable drugs [11]. Rates of isoniazid resistance are much higher (10–15% among new and 30–40% among reinfections) and lead to poor treatment outcomes when associated with resistance to other first-line drugs. Thus, understanding of the evolution and determinants of drug resistance is essential to inform development of accurate and timely diagnostics [12].

Received 4 November 2016; editorial decision 26 January 2017; accepted 30 January 2017.

<sup>a</sup>These authors contributed equally to this work.

<sup>b</sup>These authors contributed equally to the manuscript

Correspondence: Dr. A. M. Earl, The Broad Institute of MIT and Harvard, 415 Main St, Cambridge, MA 02142 (aearl@broadinstitute.org).

Clinical Infectious Diseases® 2017;64(11):1494–501

© The Author 2017. Published by Oxford University Press for the Infectious Diseases Society of America. This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial reproduction and distribution of the work, in any medium, provided the original work is not altered or transformed in any way, and that the work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com). DOI: 10.1093/cid/cix169

Recent whole genome sequencing studies have brought us closer to the goal of defining the complete catalog of drug resistance in *M. tuberculosis* [13–15]. Although recent reports suggest that current mutation sets can account for the majority of phenotypic resistance in other parts of the world [16–19], such efforts have not yet included large-scale sequencing studies from India, where *M. tuberculosis* lineages are known to vary from other global regions. Previous work has shown that Indian *M. tuberculosis* strains contain unique drug resistance mutations [20, 21], and that drug resistance patterns can vary among different lineages in India. For example, lineage 2 strains are more frequently MDR than lineage 1 or 3 strains [7, 22–25], and lineage 3 strains have been reported as more resistant than lineage 1 strains [7, 8, 26].

In this study, we sequenced 223 randomly selected strains from 196 patients within the Tiruvallur and Madurai Districts in Southern India. Through this effort, we extensively expanded the sampled diversity of lineages 1 and 3, leading to a better understanding of their genetic composition, including the determinants of drug resistance. We also examined the contributions of transmission versus *de novo* acquisition of resistance mutations, as well as the order in which drug resistance mutations were acquired. Finally, we examined the ability of existing mutation sets to predict phenotypic drug resistance and identified cases of unexplained drug resistance.

## METHODS

Between 1999 and 2005, 27 *M. tuberculosis* strains were isolated for sequencing from unique patients in Madurai and 196 strains from 169 unique patients in Tiruvallur. The sample set included data from 154 males and 42 females with median age  $39 \pm 14.5$  years. For each strain, resistance to isoniazid, ethambutol, rifampicin, and streptomycin was determined. Sequencing data

**Table 1. Summary of strain information for the set of 201 strains, after removing highly similar strains found within the same patient**  
**A) Distribution of strains with resistant and susceptible phenotypes for each of the four drugs tested.**

Drug	No. of resistant strains	No. of susceptible strains
Isoniazid	36 (18%)	162 (82%)
Ethambutol	5 (3%)	193 (97%)
Rifampicin	12 (6%)	186 (94%)
Streptomycin <sup>a</sup>	16 (9%)	171 (91%)

<sup>a</sup>For streptomycin, an additional 11 strains did not have DST information for this drug

<sup>b</sup>198 of the 201 strains had DST information and are included in this table. The remaining 3 strains did not have associated metadata.

## B) Distribution of strains across the 4 lineages.

Lineage	No. of strains
LIN-1 (EAI)	141 (70%)
LIN-2 (Beijing)	22 (11%)
LIN-3 (CAS)	32 (16%)
LIN-4 (Euro-American)	6 (3%)

were submitted to the Sequence Read Archive under BioProject PRJNA235851. See Detailed Methods.

## RESULTS

In order to examine genetic features of *M. tuberculosis* strains from Southern India, we generated high-quality whole genome sequences for 223 isolates (Table 1; Supplementary Table 1). Within our data set, phenotypic drug resistance, including MDR-TB, was not significantly associated with relapse, gender, smoking or drinking (Fisher exact test,  $P > .05$ ), even after correcting for over-representation from related same-patient isolates (see Methods). HIV status was only known for strains from Madurai, all of which were HIV positive, and only 5 of which were drug resistant (Table S1). We did not detect a significant association between strains from HIV-positive hosts and MDR or drug resistance though sample size was small.

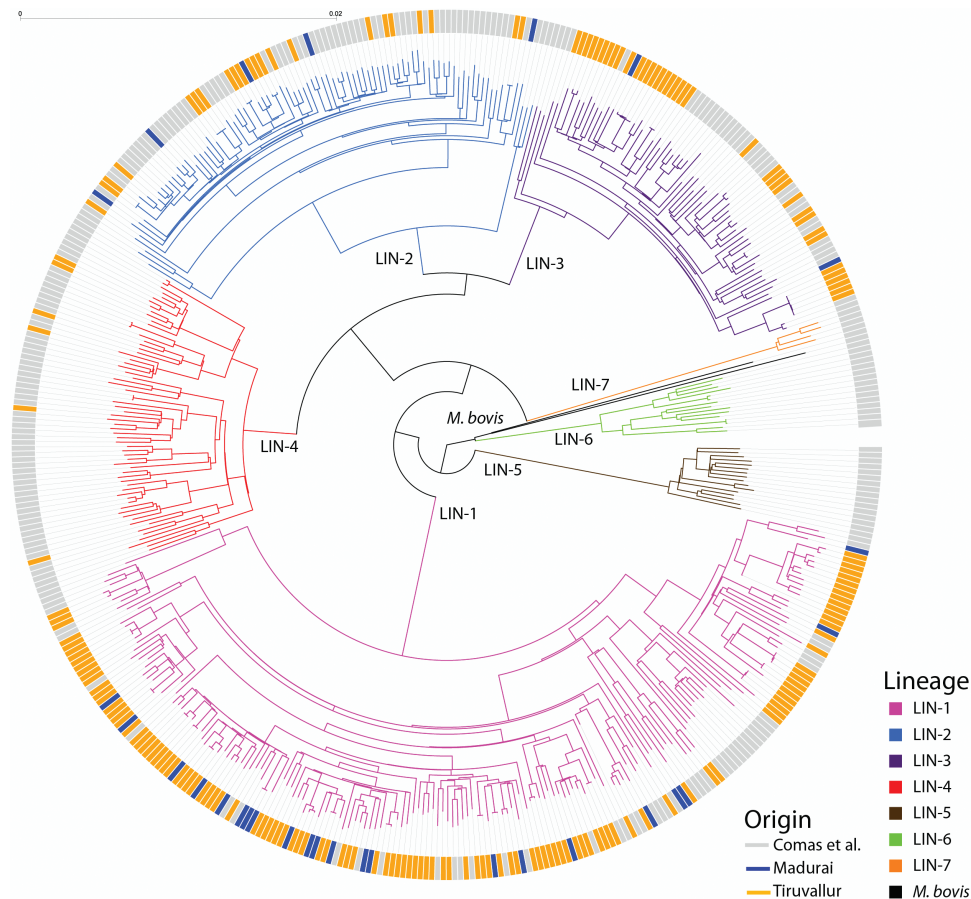
### Newly Sequenced Strains Expand the View of Genetic Diversity in Lineage 1

Using whole genome data, we constructed a phylogenetic tree and used digital spoligotyping (see Methods) to determine lineages and spoligotypes (Supplementary Figures 1–2). Lineage 1 strains dominated among newly sequenced isolates (Table 1), representing 141 (70%) of strains after correcting for over-representation from related same-patient isolates. The relative abundances of lineages 1 and 3 were in agreement with earlier epidemiological observations [2, 3].

To place our strains within a global context, we constructed a phylogeny combining genomic data from newly sequenced strains with a previously published set of sequences from 243 globally diverse *M. tuberculosis* isolates [4] (Figure 1). Although lineage 1 and 3 isolates from the published Comas study were phylogenetically intermingled with newly sequenced strains, the average distance between a new strain and its closest relative from the Comas set was 313 single-nucleotide polymorphisms (SNPs) (range: 71–833), representing an average divergence of >500 years (range: 118–2276 years) (see Detailed Methods) [27]. In contrast, we observed deeply branching clusters composed solely of Indian strains, with many strains having few or no SNP differences (average 147 SNPs, range 0–810 SNPs, and representing an average divergence time of 245 years with times ranging from 0 to 2,700 years). Based on the number of unique SNPs, the new genomic data from lineage 1 and 3 isolates roughly doubled the known genetic diversity within these lineages (Supplementary Table 2).

### Recent Transmission of Regionally Localized Strains

In addition to revealing close relationships among isolates from this study, the phylogeny indicated that recent transmission of strains was occurring among patients from the same and not different regions. To examine this more systematically, we clustered all strains from this study into “clonal groups” having  $\leq 10$  SNP



**Figure 1.** Phylogeny of all 223 newly sequenced strains, together with 243 previously published strains from Comas et al. Branches are colored according to lineage, and outer ticks are colored according to dataset of origin and geographic location. For each strain, we performed variant detection relative to the *M. tuberculosis* H37Rv reference genome (see Detailed Methods) and identified a total of 67 722 variable SNP loci that were used to construct this phylogeny. .

differences (see Methods), a threshold previously used to define recent transmission [17, 28]. Six of the 22 clonal groups (D28, D15, D30, D1, D8, and M4) contained strains from more than 1 patient (Figure 2; Table 2), and none contained strains from both Madurai and Tiruvallur. Instead, strains from each of these clones were from patients reporting to the same treatment center, indicating recent transmission was highly localized. In all but 1 case, patients having the same clone were infected with a susceptible strain. In the exceptional clone, M4 (Figure 2), we observed that an isoniazid monoresistant strain from patient 54 belonged to the same clone as three isoniazid and rifampicin-resistant strains isolated from patient 97 more than 8 months later, indicating that a strain from this clonal group likely acquired rifampicin resistance during this time period; however, resistance could have occurred earlier in either patient or in an unsampled individual.

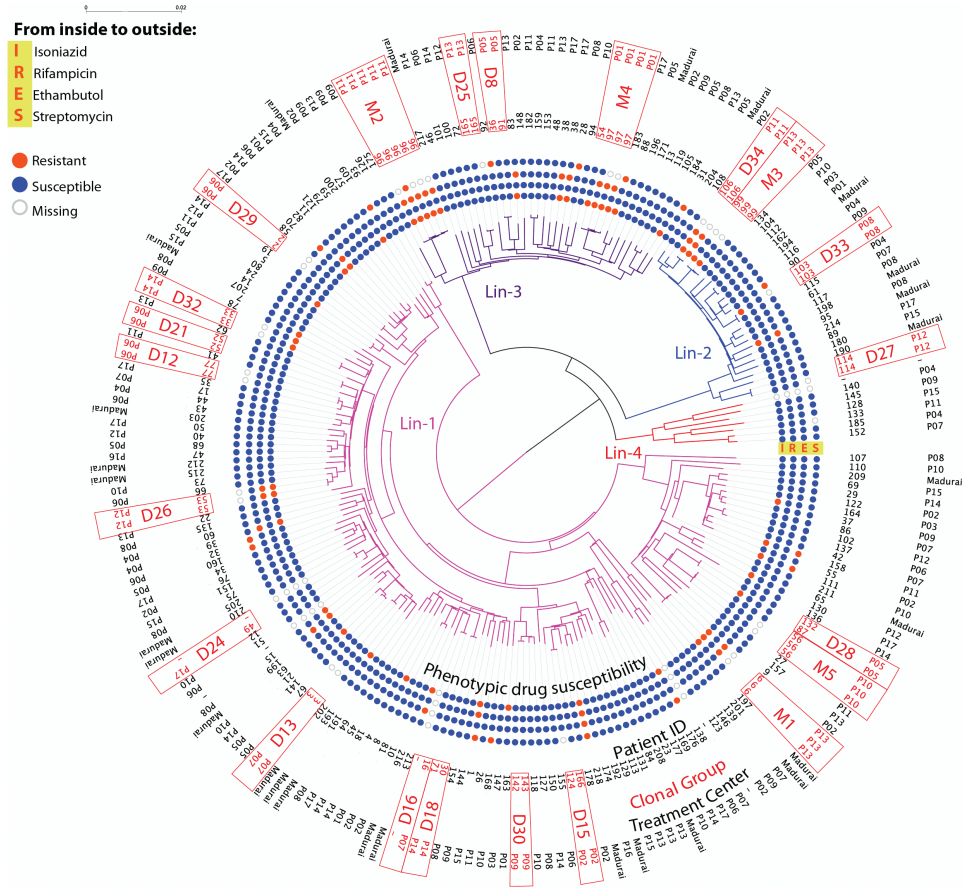
#### The Order in Which Drug Resistance Arises Among Southern Indian Tuberculosis is Similar to Other Global Regions

Our strain collection was enriched for drug resistant isolates, as compared to the overall incidence of drug-resistant

*M. tuberculosis* in India (see Methods). Isoniazid resistant strains were most common (36 strains; Table 1a), consistent with previous observations from other global regions [29, 30]. Using a parsimony-based analysis (see Methods), we determined that resistance to isoniazid arose 33 independent times across the phylogeny. For 4 isoniazid/rifampicin pairs of resistance arisals across the nodes of our phylogeny, we had sufficiently dense sampling to determine relative ordering. In all 4 cases, isoniazid arose first. Although there were only a small number of resistance pairs for which we could determine the order, the observation that isoniazid arose first in all 4 cases is in agreement with results from previously published genomic studies from South Africa [27] and South America [31].

#### Novel Predicted Drug Resistance Mutations in India

Recent sequencing studies suggest that known mutations can account for the majority of phenotypic resistance [16–19] but have been far from exhaustive in representing global *M. tuberculosis* diversity. To assess how well-known mutations explain drug resistance in India, we calculated how well 2 previously published



**Figure 2.** Clonal groups overlaid onto a phylogeny of all 223 newly sequenced strains from Southern India. Tree branches are colored by lineage, as in Figure 1. The central rings of dots indicate susceptibility phenotypes for isoniazid (11), rifampicin (11), ethambutol (11), and streptomycin (11) (from inside to outside, marked as “IRES”). A gray “Missing” mark indicates that no conclusive phenotype was available. Moving outward, the next ring is a numeric patient identifier, followed by clonal group ID (in red) and treatment center. Clonal groups are boxed in red. Treatment centers P01-P14 are in Tiruvallur, whereas all Madurai strains are from the Gov. Rajaji Hospital. Clonal groups D16 and D24 contained a strain that lacked patient information and were excluded from further analysis.

**Table 2. Summary table of clonal groups of strains occurring in multiple patients.**

Clonal group	Treatment center	Patient ID	Strain ID	Date	Phenotype <sup>a</sup>				Genotype <sup>b</sup>				Notes
					INH	EMB	RIF	STR	INH	EMB	RIF	STR	
D28	P05	87	M1560	10/8/2002	R	S	S	S	S	S	S	S	Phenotypic DR change only
		132	M2010	4/22/2003	S	S	S	S	S	S	S	S	
D15	P02	124	M2343	9/17/2003	S	S	S	-	S	S	S	S	
		166	M2342	9/17/2003	S	S	S	S	S	S	S	S	
D30	P09	142	M2129	7/3/2003	S	S	S	S	S	S	S	S	
		143	M2128	7/1/2003	S	S	S	S	S	S	S	S	
D1	P14	30	M1272	11/9/2001	S	S	S	S	S	S	S	S	
		71	M1425	7/9/2002	S	S	S	S	S	S	S	S	
D8	P05	36	M1294	3/6/2002	S	S	S	S	S	S	S	S	
		91	M1638	11/8/2002	S	S	S	S	S	S	S	S	
M4	P01	54	M1393	5/21/2002	R	S	S	S	R	S	S	S	Genotypic evolution, rpoB S450L
		97	M1780	1/10/2003	R	S	R	S	R	S	R	S	
			M2323	8/21/2003	R	S	R	-	R	S	R	S	
			M2324	8/21/2003	R	S	R	-	R	S	R	S	

Abbreviation: DR, drug resistance.

<sup>a</sup>S, susceptible; R, resistant; -: missing information. Indicates drug susceptibility for isoniazid (INH), ethambutol (EMB), rifampicin (RIF) and streptomycin (STR)

<sup>b</sup>Genotypic resistance is determined based on the Cohen et al. set

**Table 3. Sensitivity and specificity for predicting phenotypic resistance using the Coll et al. and Cohen et al. lists of mutations**

Drug	Mutation set	True Positives <sup>a</sup>	False Negatives	False Positives	True Negatives	Ambiguous <sup>a</sup>	Sensitivity	Specificity
Isoniazid	Coll	24	9	9	155	3	<b>72.7%</b>	94.5%
	Cohen	25	9	5	159	2	<b>73.5%</b>	97.0%
	Coll+Cohen	26	8	9	155	3	<b>76.5%</b>	94.5%
Rifampicin	Coll	8	3	2	186	1	<b>72.7%</b>	98.9%
	Cohen	8	3	1	187	1	<b>72.7%</b>	99.5%
	Coll+Cohen	8	3	2	186	1	<b>72.7%</b>	98.9%
Streptomycin	Coll	4	11	3	180	1	<b>26.7%</b>	98.4%
	Cohen	9	6	7	177	1	<b>60.0%</b>	96.2%
	Coll+Cohen	9	6	7	176	1	<b>60.0%</b>	96.2%
Ethambutol	Coll	4	1	137 <sup>b</sup>	57	2	<b>80.0%</b>	29.4%
	Cohen	2	2	1	194	2	<b>50.0%</b>	99.5%
	Coll+Cohen	4	1	137	57	3	<b>80.0%</b>	29.4%

<sup>a</sup>Ambiguous single-nucleotide polymorphism calls, with reads supporting a drug sensitive and a drug resistant genotype. These positions were excluded from the calculations for sensitivity and specificity.

<sup>b</sup>The large number of false positives for streptomycin using the Coll dataset is due to the presence of the *embB* E378A mutation, which other publications suggest does not cause ethambutol resistance. This mutation has been previously reported as a phylogenetic marker of lineage 1 of the ancestral MTBC, and this mutation is indeed present in most lineage 1 strains in this collection.

<sup>c</sup>Analysis performed for 201 strains, after removing highly similar strains found within the same patient.

mutation lists explained phenotypic resistance in these isolates. These lists included (i) 1325 mutations explaining resistance to 15 drugs [32] (Coll set) and (ii) a curated list of polymorphisms explaining resistance to eight drugs [27] (Cohen set; Table 3). The sensitivity to predict isoniazid resistance among strains from Southern India was 73% for the Coll set and 74% for the Cohen set. Overall, these sensitivities were significantly lower than those reported from an analysis of isolates from the United Kingdom, Sierra Leone, and South Africa (>85%) [17] but were similar to an earlier study from Southern India in which isoniazid resistant strains had a lower frequency of known *katG* mutations [33]. The sensitivities for detecting rifampicin, streptomycin, and ethambutol resistance (73%, 27–60%, and 50–80%, respectively) were also lower than those reported for other geographic regions by Walker et al. (92%, 82%, and 82%, respectively) [17] and Desjardins et al. (>90%) [19], indicating that novel mechanisms may drive some drug resistance in India. Though novel *rpoB* and *gidB* mutations previously identified in Indian strains were not included in either list, neither mutation was present within this set of strains [20, 21]. Notably, 27% of rifampicin resistant strains lacked mutations in the resistance-determining region of *rpoB*, the basis for a widely used MDR-TB diagnostic [34, 35]. We confirmed that our results extended to other published mutation lists and that overall sensitivity was only marginally improved by including any mutation that would impact the resulting protein in genes associated with resistance (see Supplementary Note and Supplementary Figure 3).

To further examine cases of unexplained resistance, we first searched for potential novel mutations within genes previously implicated in resistance but not found in our mutation lists. Of the 9 strains with unexplained isoniazid resistance (Table 4; Supplementary Figure 4), one encoded the *katG* N138S mutation, previously implicated in resistance but not present on

either list [36]. Three strains contained mutations in 1 or more well-recognized resistance-determining targets, *katG* (A290P or L427P), *fadE24* (R454S) or *fabD* (A159T), that had not previously been implicated in resistance but that were specific to resistant strains in this study (Table 4; Supplementary Figure 4). The remaining 4 strains lacked mutations in genes previously

**Table 4. Overview of phenotypic drug resistance not explained by either the mutations in the Coll et al. or Cohen et al. lists, or by mixed infections.**

Drug	Strain identifier <sup>b</sup>	Category <sup>a</sup>	Description <sup>c</sup>
Isoniazid	M2139	Unexplained	
	M1560	Unexplained	
	M2259	Putative novel mutation	<i>fadE24</i> R454S
	M1292	Putative novel mutation	<i>fabD</i> A159T
	M2360, M2361	Unexplained	
	M1948	Putative novel mutation	<i>katG</i> A290P
	M2131	Putative novel mutation	<i>katG</i> L427P
M1545	Published mutation	<i>katG</i> N138S [36]	
Rifampicin	M1292	Unexplained	
	M1324, M1325	Unexplained	
	H2438	Unexplained	
Ethambutol	M2084, M2085	Unexplained	
Streptomycin	M1270	Unexplained	
	M1762	Unexplained	
	M2116	Unexplained	
	M2206	Unexplained	
	M0013	Published mutation	<i>gidB</i> R137W [17]
	H3367	Published mutation	<i>rrs</i> 877 [50]

<sup>a</sup>“Putative novel mutation” is defined by a non-synonymous mutation in a gene previously associated with drug resistance that was confined to phenotypically resistant isolates. “Unexplained” is defined when no new mutations were identified within genes previously associated with drug resistance. “Published mutation” is defined by mutations not present in either the Coll et al. or Cohen et al. mutation sets, but were previously associated with drug resistance.

<sup>b</sup>Multiple entries in the “strain identifier” column correspond to same-patient strains with high identity

<sup>c</sup>Previously published mutations, which were not included in either the Coll et al. or Cohen et al. datasets.

associated with isoniazid resistance (Table 4; Supplementary Figure 4). For streptomycin, we identified mutations in *gidB* and *rrs* (*gidB* R137W and *rrs* 877) that may explain resistance for 2 of these strains. However, for all remaining isolates with unexplained resistance (4 rifampicin-, 2 ethambutol-, and 2 streptomycin-resistant isolates), we could identify no additional candidate resistance-conferring mutations.

#### Presence of Mixed Infections Affects Ability to Predict Phenotype Based on Genotype

Recent studies suggest that tuberculosis patients can be infected with more than 1 *M. tuberculosis* strain, including strains with differing resistance profiles [37–40]. We searched for evidence of mixed infections by, first, comparing repeat isolates from single patients and, second, examining variant data to identify evidence for conflicting base calls at sites implicated in resistance. Sixteen patients were sampled multiple times: 8 patients were sampled at different time points, and 14 were sampled multiple times on the same day. Consistent with rates from other studies [41, 42], none of the isolates collected from same-day, same-patient samplings belonged to different clonal groups (Supplementary Table 3), differing by only 0–6 SNPs, and indicating that same-day, within-patient *M. tuberculosis* sequence variation is low. With one exception, for longitudinally sampled patients, the first and last isolate(11) differed by only 2–8 SNPs, also supporting limited diversity. The exception (patient 38) was represented by 2 isolates differing by 301 SNPs, collected 6 months apart. Given slow mutation rates [17, 27] and lack of recombination [43, 44] in *M. tuberculosis*, these 2 isolates were either transmitted to this patient at the initial point of infection or were acquired at different times.

Though we observed very consistent genotypes within the same patient, 3 longitudinally sampled patients exhibited changes in drug resistance phenotypes over the study (Supplementary Table 3). However, we observed no accompanying genotypic change that could explain this change, even when we expanded our analysis to include any changes in known resistance targets. Because *M. tuberculosis* cells are known to bundle [45], making it difficult to isolate single genotypes in culture, we hypothesized that unexplained changes in phenotype may be due to mixed communities of genotypically distinct strains that we could detect through sequencing. However, we could find no evidence for ambiguity in any gene previously associated with drug resistance among these isolates.

Using the same strategy, we systematically examined all strains for examples of mixed infections that could help explain cases of unexplained resistance (Table 4). Fifteen strains had ambiguous variant calls within genes known to be involved in resistance (Supplementary Table 4), including 7 strains with ambiguous calls at sites found within the Coll and Cohen sets (Table 3; Supplementary Table 4). In these cases, we observed additional ambiguous positions at >10 positions with similar allele ratios at other non-resistance-associated sites, indicating mixed infections with nonclonal

strains. One of these strains, M1961, contained a call of ‘ambiguous’ at 2 known resistance sites found in the Coll and Cohen sets, as well as at 3 other sites within genes known to be involved in resistance. At each site, the major allele was observed 52–58% of the time (Supplementary Table 4). Although we detected strains having ambiguity at sites within drug resistance genes not present within the Cohen and Coll sets, none of these ambiguities could explain additional resistance for isolates in Table 4. Despite this, the fact that at least 7% of our strains had ambiguity within drug resistance genes, and that ambiguous sites were not taken into account when we calculated sensitivity and specificity of known mutations to explain resistance, underscores mixed infections as a confounder of genotype-based predictions of phenotypes.

## DISCUSSION

Our study, focused on *M. tuberculosis* from Southern India, expands the catalog of genetic diversity for the *M. tuberculosis* complex, particularly contributing to our understanding of lineage 1 and 3 strains that predominate in India, a region for which whole genome sequencing-based analyses of *M. tuberculosis* have been limited [46, 47]. In addition, our study provides insights into *M. tuberculosis* transmission patterns in Southern India, how these strains evolve drug resistance and provides a community resource for exploration of global *M. tuberculosis* diversity, including the unique characteristics of lineages 1 and 3.

We observed highly localized cases of patient-to-patient transmission of strains ( $\leq 10$  SNP differences) within individual treatment centers. Consistent with previous studies in India, this could indicate a need for focus on prevention of nosocomial transmission [48]. However, we lacked sufficient metadata to determine whether transmission occurred within a hospital or community setting.

Though our strains were enriched for resistant isolates (56) compared to the overall incidence in India [10], including 21 MDR-TB, we were only able to disambiguate the relative ordering of acquisition of resistance for a small number of pairs. Though further studies will be needed to confirm our results, they were consistent with previous work showing that isoniazid resistance arises first and serves as a precursor for MDR-TB in other world regions [27, 31].

A substantial fraction of phenotypic drug resistance in our sample could be explained by known mechanisms, consistent with studies in other lineages [16–19]. However, we observed that lists of known mutations performed less well in predicting phenotypic drug resistance for Indian strains than for other regions [17, 33]. That we could not explain the cause of resistance for a quarter (14 of 56) of strains may not be surprising given that current catalogs of drug resistance mutations rely predominantly upon data from other lineages. Though enriched for drug resistance, our data set contained only a small number of resistant isolates—particularly for ethambutol, rifampicin, and

streptomycin—which may impact the reliability of our sensitivity and specificity calculations for larger collections and also makes identifying novel mechanisms outside of known targets difficult.

Our effort did reveal several possible explanations for unexplained resistance within known targets. First, we identified mutations—*katG* A290P, *katG* L427P, *fadE24* R454S, and *fabD* A159T—that should be experimentally prioritized for assessing their role in resistance. These add to the list of mutations previously identified as special to Indian strains [20, 21]. We did not find previously identified special resistance mutations among isolates from this collection, which could be explained by the fact that our study was small relative to the total number of resistant cases in India, and that it was also geographically constrained.

Second, careful examination of ambiguous base calls revealed that some patients harbored mixed infections of sensitive and resistant genotypes, impacting 7% of the strains we examined. However, even after accounting for the possibility of mixed infections at additional positions within genes involved in drug resistance, there were still 14 strains with unexplained resistance. Of the patients harboring these strains, 4 defaulted on treatment, which could have contributed to the development of resistance, and four were classified as “treatment failure.” Further studies are needed to examine whether these discrepancies are due to novel resistance mechanisms, the existence of mixed communities undetectable with our approach, phenotyping error, or alternative intrinsic mechanisms of resistance, such as the permeability of the cell wall and efflux pumps, which can vary between strains due to factors such as variable expression of genes [49].

Ultimately, the identification of putative novel resistance conferring mutations, paired with the higher rate of unexplained resistance, highlights the challenges for diagnosing drug resistance in India. Our findings raise the possibility that molecular diagnostics for tuberculosis drug resistance may need to be tailored for India. For instance, GeneXpert MTB/RIF [34], the front-line diagnostic used worldwide, relies on rifampicin resistance mutations to detect drug resistance [34], whereas in our study, we observed cases of rifampicin resistance of unknown cause. In addition, other commercially available diagnostics, such as Hain MTBDR*plus* (<http://www.hain-lifescience.de/en/products/microbiology/mycobacteria/tuberculosis/genotype-mtbdplus.html>), and Hain MTBDR*sl* (<http://www.hain-lifescience.de/en/products/microbiology/mycobacteria/tuberculosis/genotype-mtbdsl.html>) would also not have detected resistance for strains in Table 4. Larger whole-genome sequencing studies of strains from India are needed to establish patterns of lineage 1 and 3 specific mutations and to determine the utility of developing novel diagnostics specific for strains circulating in India.

### Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted

materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

### Notes

**Acknowledgments.** The samples were collected from Model DOTS project funded in part by a grant from the United States Agency of International Development provided through the World Health Organization. We thank D. Srinivasaraju for his technical help in DNA fingerprinting. We acknowledge the support from the staff of Departments of Clinical Research, Epidemiology, Immunology, and Bacteriology; Dr P.G. Gopi and Mr R. Subramani and their staff for data management. We acknowledge Drs. C. Kolappan and K. Sadacharam for coordination of field activities. We also thank Dr Keira Cohen for helpful comments and reading the manuscript.

**Disclaimer.** The contents of this publication are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

**Financial support.** This project also received funding from the Director's Fund, Boston University National Emerging Infectious Diseases Laboratories (JEG). This project has been funded in part with Federal funds from the National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health, Department of Health and Human Services [Contract no.:HHSN272200900018C, Grant no. U19AI110818 to the Broad Institute]. This research used infrastructure resources from the Broad Institute, the Delft University of Technology, Internet2, and SURFnet (the Dutch research and education network), supported under the Enlighten Your Research Global program.

**Ethics Statement.** Informed consent was obtained. The study protocol was approved by the Massachusetts Institute of Technology Committee on Use of Humans as Experimental Subjects (Protocol # 0504001179) and the Institutional Ethics Committee of NIRT (Protocol # 2013010). All guidelines were followed in the conduct of research.

**Potential conflicts of interest.** All authors certify no potential conflicts of interest. All author have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

### References

1. WHO. Global Tuberculosis Report 2016. World Health Organization 2016.
2. Gutierrez MC, Ahmed N, Willery E, et al. Predominance of ancestral lineages of *Mycobacterium tuberculosis* in India. *Emerg Infect Dis* 2006; 12:1367–74.
3. Narayanan S, Gagneux S, Hari L, et al. Genomic interrogation of ancestral *Mycobacterium tuberculosis* from south India. *Infect Genet Evol* 2008; 8:474–83.
4. Comas I, Coscolla M, Luo T, et al. Out-of-Africa migration and neolithic coexpansion of *Mycobacterium tuberculosis* with modern humans. *Nat Genet* 2013; 45:1176–82.
5. Merker M, Blin C, Mona S, et al. Evolutionary history and global spread of the *Mycobacterium tuberculosis* Beijing lineage. *Nat Genet* 2015; 47:242–9.
6. Reed MB, Pichler VK, McIntosh F, et al. Major *Mycobacterium tuberculosis* lineages associate with patient country of origin. *J Clin Microbiol* 2009; 47:1119–28.
7. Singh J, Sankar MM, Kumar P, Couvin D, Rastogi N, Singh S; Indian TB Diagnostics Network. Genetic diversity and drug susceptibility profile of *Mycobacterium tuberculosis* isolated from different regions of India. *J Infect* 2015; 71:207–19.
8. Shanmugam S, Selvakumar N, Narayanan S. Drug resistance among different genotypes of *Mycobacterium tuberculosis* isolated from patients from Tiruvallur, South India. *Infect Genet Evol* 2011; 11:980–6.
9. Kandhakumari G, Stephen S, Sivakumar S, Narayanan S. Spoligotype patterns of *Mycobacterium tuberculosis* isolated from extra pulmonary tuberculosis patients in Puducherry, India. *Indian J Med Microbiol* 2015; 33:267–70.
10. Facing the Reality of Drug-Resistant Tuberculosis in India: Challenges and Potential Solutions: Summary of a Joint Workshop by the Institute of Medicine, the Indian National Science Academy, and the Indian Council of Medical Research. Washington, DC: National Academies Press, 2012.
11. Selvakumar N, Kumar V, Balaji S, et al. High rates of ofloxacin resistance in *Mycobacterium tuberculosis* among both new and previously treated patients in Tamil Nadu, South India. *PLoS One* 2015; 10:e0117421.
12. Jacobson KR, Tierney DB, Jeon CY, Mitnick CD, Murray MB. Treatment outcomes among patients with extensively drug-resistant tuberculosis: systematic review and meta-analysis. *Clin Infect Dis* 2010; 51:6–14.



13. Ali A, Hasan Z, McNeerney R, et al. Whole genome sequencing based characterization of extensively drug-resistant mycobacterium tuberculosis isolates from Pakistan. *PLoS ONE* **2015**; 10: 1–17.
14. Anderson LF, Tamme S, Brown T, et al. Transmission of multidrug-resistant tuberculosis in the UK: a cross-sectional molecular and epidemiological study of clustering and contact tracing. *Lancet Infect Dis* **2014**; 14:406–15.
15. Casali N, Nikolayevskyy V, Balabanova Y, et al. Evolution and transmission of drug-resistant tuberculosis in a Russian population. *Nat Genet* **2014**; 46:279–86.
16. Farhat MR, Shapiro BJ, Kieser KJ, et al. Genomic analysis identifies targets of convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. *Nat Genet* **2013**: 1–9.
17. Walker TM, Kohl Ta, Omar SV, et al. Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: a retrospective cohort study. *Lancet Infect Dis* **2015**; 3099: 1–10.
18. Zhang H, Li D, Zhao L, et al. Genome sequencing of 161 *Mycobacterium tuberculosis* isolates from China identifies genes and intergenic regions associated with drug resistance. *Nat Genet* **2013**; 45:1255–60.
19. Desjardins CA, Cohen KA, Munsamy V, et al. Genomic and functional analyses of *Mycobacterium tuberculosis* strains implicate *ald* in D-cycloserine resistance. *Nat Genet* **2016**; 48:544–51.
20. Verma JS, Gupta Y, Nair D, et al. Evaluation of *gidB* alterations responsible for streptomycin resistance in *Mycobacterium tuberculosis*. *J Antimicrob Chemother* **2014**; 69:2935–41.
21. Mani C, Selvakumar N, Narayanan S, Narayanan PR. Mutations in the *rpoB* gene of multidrug-resistant *Mycobacterium tuberculosis* clinical isolates from India. *J Clin Microbiol* **2001**; 39:2987–90.
22. Chatterjee A, D'Souza D, Vira T, et al. Strains of *Mycobacterium tuberculosis* from western Maharashtra, India, exhibit a high degree of diversity and strain-specific associations with drug resistance, cavitory disease, and treatment failure. *J Clin Microbiol* **2010**; 48:3593–9.
23. Gupta A, Kulkarni S, Rastogi N, Anupurba S. A study of *Mycobacterium tuberculosis* genotypic diversity and drug resistance mutations in Varanasi, north India. *Indian J Med Res* **2014**; 139:892–902.
24. Purwar S, Chaudhari S, Katoch VM, et al. Determination of drug susceptibility patterns and genotypes of *Mycobacterium tuberculosis* isolates from Kanpur district, North India. *Infect Genet Evol* **2011**; 11:469–75.
25. Sankar MM, Singh J, Diana SC, Singh S. Molecular characterization of *Mycobacterium tuberculosis* isolates from North Indian patients with extrapulmonary tuberculosis. *Tuberculosis (Edinb)* **2013**; 93:75–83.
26. Arora J, Singh UB, Suresh N, et al. Characterization of predominant *Mycobacterium tuberculosis* strains from different subpopulations of India. *Infect Genet Evol* **2009**; 9:832–9.
27. Cohen KA, Abeel T, Manson McGuire A, 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* **2015**; 12:e1001880.
28. Williams OM, Abeel T, Casali N, et al. Fatal nosocomial MDR TB identified through routine genetic analysis and whole-genome sequencing. *Emerg Infect Dis* **2015**; 21:1082–4.
29. Jenkins HE, Zignol M, Cohen T. Quantifying the burden and trends of isoniazid resistant tuberculosis, 1994–2009. *PLoS One* **2011**; 6: e22927.
30. WHO. WHO monitoring of Xpert MTB/RIF roll-out. **2012**.
31. Eldholm V, Monteserin J, Rieux A, et al. Four decades of transmission of a multidrug-resistant *Mycobacterium tuberculosis* outbreak strain. *Nat Commun* **2015**; 6:7119.
32. Coll F, McNeerney R, Preston MD, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. *Genome Med* **2015**; 7:51.
33. Unissa AN, Selvakumar N, Narayanan S, Suganthi C, Hanna LE. Investigation of Ser315 substitutions within *katG* gene in isoniazid-resistant clinical isolates of *Mycobacterium tuberculosis* from south India. *Biomed Res Int* **2015**; 2015:257983.
34. Boehme CC, Nabeta P, Hillemann D, et al. Rapid molecular detection of tuberculosis and rifampin resistance. *N Engl J Med* **2010**; 363:1005–15.
35. Telenti A, Imboden P, Marchesi F, et al. Detection of rifampicin-resistance mutations in *Mycobacterium tuberculosis*. *Lancet* **1993**; 341:647–50.
36. Cade CE, Dlouhy AC, Medzihradsky KF, Salas-Castillo SP, Ghiladi RA. Isoniazid-resistance conferring mutations in *Mycobacterium tuberculosis* *KatG*: catalase, peroxidase, and INH-NADH adduct formation activities. *Protein Sci* **2010**; 19:458–74.
37. van Rie A, Victor TC, Richardson M, et al. Reinfection and mixed infection cause changing *Mycobacterium tuberculosis* drug-resistance patterns. *Am J Respir Crit Care Med* **2005**; 172:636–42.
38. Warren RM, Victor TC, Streicher EM, et al. Patients with active tuberculosis often have different strains in the same sputum specimen. *Am J Respir Crit Care Med* **2004**; 169:610–4.
39. Baldeviano-Vidalón GC, Quispe-Torres N, Bonilla-Asalde C, Gastiaburú-Rodríguez D, Pro-Cuba JE, Llanos-Zavalaga F. Multiple infection with resistant and sensitive *M. tuberculosis* strains during treatment of pulmonary tuberculosis patients. *Int J Tuberc Lung Dis* **2005**; 9:1155–60.
40. Fang R, Li X, Li J, et al. Mixed infections of *Mycobacterium tuberculosis* in tuberculosis patients in Shanghai, China. *Tuberculosis (Edinb)* **2008**; 88:469–73.
41. Wollenberg KR, Desjardins CA, Zalutskaya A, et al. Whole genome sequencing of *Mycobacterium tuberculosis* provides insight into the evolution and genetic composition of drug-resistant tuberculosis in Belarus. *J Clin Microbiol* **2017**; 55:457–69.
42. Eldholm V, Norheim G, von der Lippe B, et al. Evolution of extensively drug-resistant *Mycobacterium tuberculosis* from a susceptible ancestor in a single patient. *Genome Biol* **2014**; 15:490.
43. Gutacker MM, Mathema B, Soini H, et al. Single-nucleotide polymorphism-based population genetic analysis of *Mycobacterium tuberculosis* strains from 4 geographic sites. *J Infect Dis* **2006**; 193:121–8.
44. Gutacker MM, Smoot JC, Migliaccio CA, et al. Genome-wide analysis of synodnymous single nucleotide polymorphisms in *Mycobacterium tuberculosis* complex organisms: resolution of genetic relationships among closely related microbial strains. *Genetics* **2002**; 162:1533–43.
45. Velayati AA, Farnia P. Morphological Characterization of *Mycobacterium tuberculosis*. *Understanding Tuberculosis: Deciphering the Secret Life of the Bacilli: InTech*. **2012**.
46. Das S, Roychowdhury T, Kumar P, et al. Genetic heterogeneity revealed by sequence analysis of *Mycobacterium tuberculosis* isolates from extra-pulmonary tuberculosis patients. *BMC Genomics* **2013**; 14:404.
47. Periwal V, Patowary A, Vellarikkal SK, et al. Comparative whole-genome analysis of clinical isolates reveals characteristic architecture of *Mycobacterium tuberculosis* pangenome. *PLoS One* **2015**; 10:e0122979.
48. Pai M, Kalantri S, Aggarwal AN, Menzies D, Blumberg HM. Nosocomial tuberculosis in India. *Emerg Infect Dis* **2006**; 12:1311–8.
49. Chatterjee A, Saranath D, Bhattar P, Mistry N. Global transcriptional profiling of longitudinal clinical isolates of *Mycobacterium tuberculosis* exhibiting rapid accumulation of drug resistance. *PLoS One* **2013**; 8:e54717.
50. Sreevatsan S, Pan X, Stockbauer KE, Williams DL, Kreiswirth BN, Musser JM. Characterization of *rpsL* and *rrs* mutations in streptomycin-resistant *Mycobacterium tuberculosis* isolates from diverse geographic localities. *Antimicrob Agents Chemother* **1996**; 40:1024–6.