Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Shah NS, Auld SC, Brust JCM, et al. Transmission of extensively drug-resistant tuberculosis in South Africa. N Engl J Med 2017;376:243-53. DOI: 10.1056/NEJMoa1604544

SUPPLEMENTARY APPENDIX

Supplement to: Shah NS, Auld SC, Brust JCM, et al. Transmission of Drug-resistant Tuberculosis in South Africa. New England Journal of Medicine 16-04544

Table of Contents

Methods

The diagnostic XDR-TB isolate was obtained for all enrolled participants through collaboration with the provincial referral laboratory. Isolates were then re-cultured on Löwenstein-Jensen slants and were grown at 37°C for 3-5 weeks. Genomic DNA extraction and IS*6110* RFLP genotyping was done in accordance with standard methods.[1] All DNA underwent targeted sequencing of eight resistance-conferring regions for rifampin, isoniazid, pyrazinamide, fluoroquinolones and second-line injectable drugs: *rpoB, katG, inhA, pncA, gyrA, rpsL, rrs, gidB*. A subset of 298 isolates also underwent paired-end whole genome sequencing (WGS). Sequencing libraries were prepared using Nextera DNA kits (Illumina, San Diego, CA). Raw paired-end sequencing reads were generated on the Illumina (MiSeq) platform and aligned to the H37Rv reference genome (NC_000962.2) using the Burrows-Wheeler Aligner.[2] All isolates had reads covering >99% of the reference genome, and the lowest mean coverage depth for any isolate was 15X. Single nucleotide polymorphisms (SNPs) were detected using standard pairwise resequencing techniques (Samtools v0.1.19)[3] against the reference. SNPs were annotated as *syn* or *nonsyn*, and *genic* or *intergenic*. Concatenated alignments imputed from these SNP calls were employed for maximum likelihood phylogenetic reconstruction using RAxML[4] at quality threshold of 40 (i.e., Q40). Trees were estimated with and without sequence data from the highly variable PE/PPE gene class.

Maximum likelihood phylogenetic reconstruction

A maximum-likelihood phylogenetic tree was estimated using the general time-reversal-gamma distribution model of nucleotide substitution and ascertainment bias correction, as implemented in RAxML v8.[47] Node robustness was evaluated using 100 bootstrap pseudoreplicates. 298 study isolates were included for analysis, as were five publicly available sequences from three *Mtb* phylogeographic lineages: Percy 256 (Lineage 7, SRA ID: ERR181435), W4 (Lineage 2, SRA ID: ERR071083), W148 (Lineage 2, SRA ID: SRR849475), GM_1503 (Lineage 4, South Africa, SRA ID: SRR029156), and F11 (Lineage 4, South Africa, SRA ID: SRR974839).

Figure S1. Validation of genotypic clusters defined by RFLP and targeted gene sequencing.

(A) Distribution of pairwise SNP differences for isolates within the same cluster (grey) and between isolates in different clusters (black) using the study genotypic cluster definition (i.e., RFLP and gene sequencing data); (B) Distribution of pairwise SNP differences for isolates within the same cluster (grey) and between isolates from different clusters (black) using broad RFLP groups (i.e., strain family level).

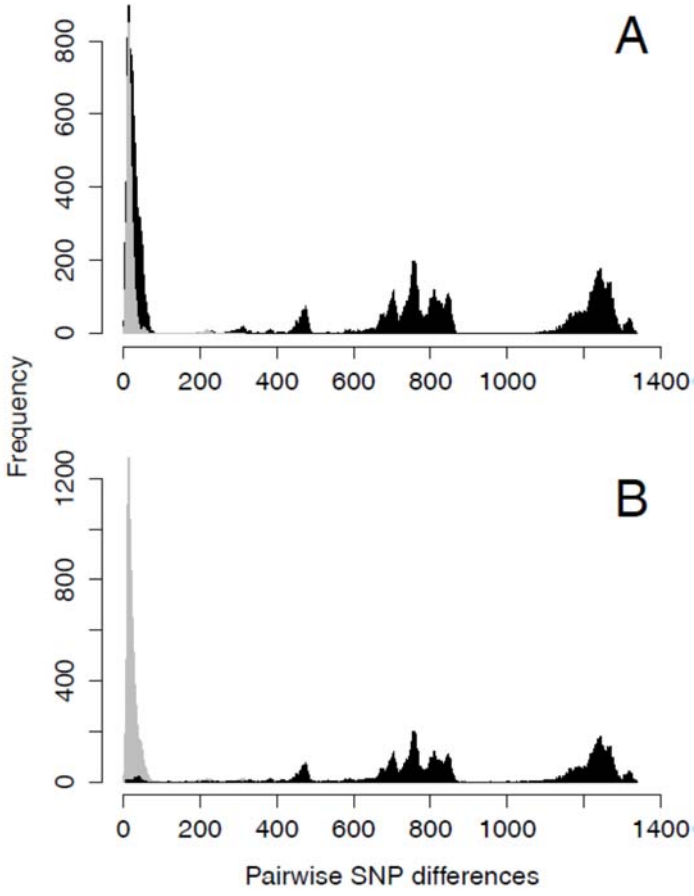


Figure S2. Acquired versus Transmitted Resistance among Patients with Extensively Drug-resistant Tuberculosis (XDR-TB) by Clinical versus Genotypic Case Definition.

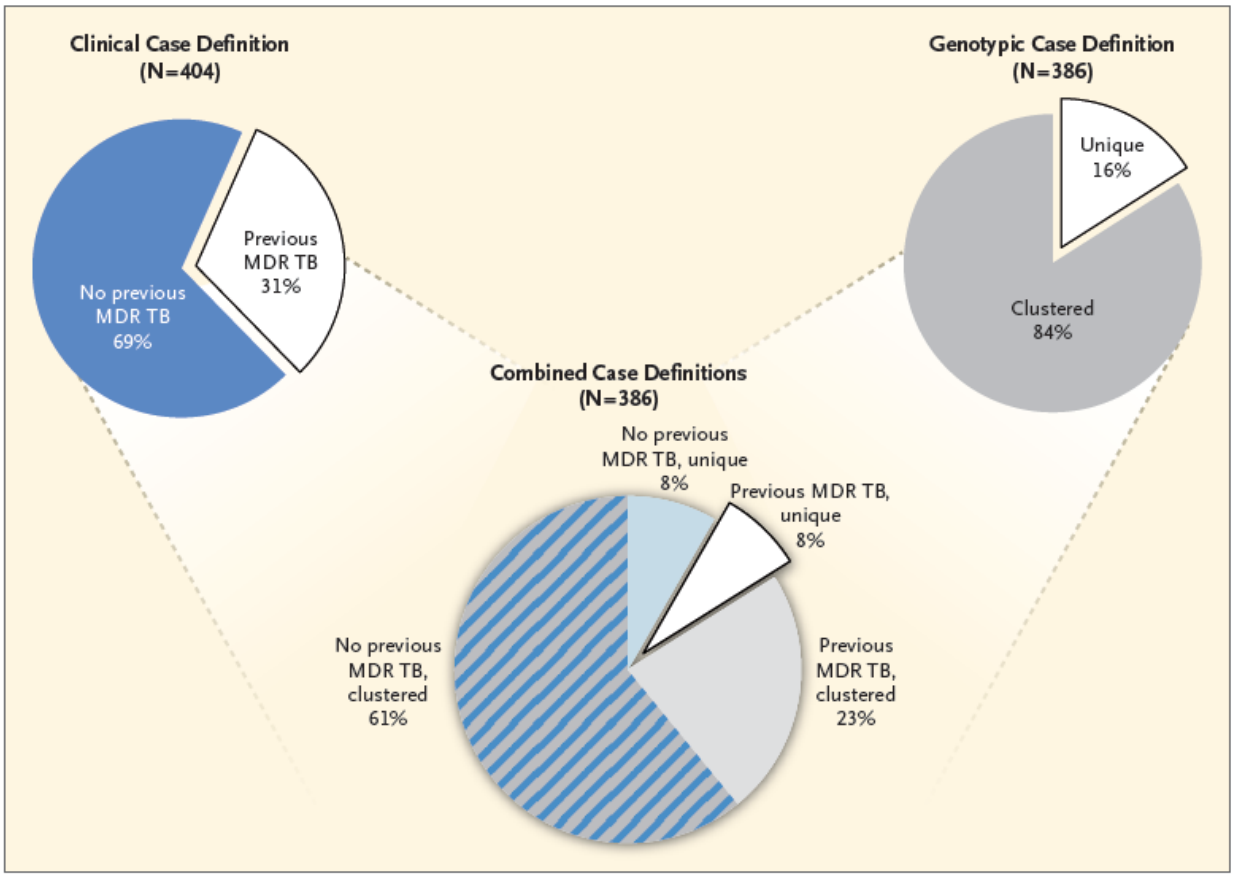
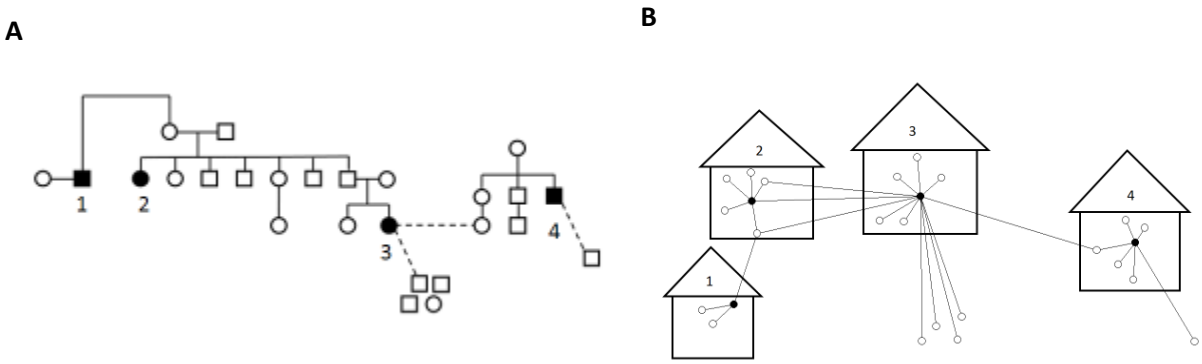


Figure S3. Social Networks in Homes and Communities.

Panel A & B. Family and community linkages for a selected network of a multigenerational family affected by XDR-TB and their community-based linkage to another family with XDR-TB. Black shading represents study participants (circle = female, square = male). In Panel B, solid lines represent direct family relations; dotted lines represent community relations (e.g., workplace, school). Patient 1 is the 32 year-old cousin to 30 year-old patient 2, who is aunt to 12 year-old patient 3. While these patients were living separately at the time of their XDR-TB diagnoses (as represented by the houses in Panel C), patients 2 and 3 had previously lived together. Patient 3 was schoolmates with a child who lived in the same home as patient 4. Patients 1, 2 and 3 shared the same RFLP genotype, while patient 4 had a 1-band difference in RFLP from the others. There was 1 SNP difference between patients 1 and 2; 8 SNP differences between patients 2 and 3; and 30 SNP differences between patients 3 and 4.



Panel C. Hospital overlap for a subset of 11 participants admitted to the same hospital before or after their XDR-TB diagnosis. Black circles indicate sputum collection date for XDR-TB diagnosis. Grey bars indicate duration of hospitalization. The vulnerable periods, defined as >1 month prior to XDR-TB sputum collection, are shown with a black outline around the hospital admission periods. Patients are presumed to be infectious for 30 days prior to and for the duration of hospitalization following sputum collection. Patients A and C overlapped and had matching genotype; patients D, F, G and H overlapped and had matching genotype.

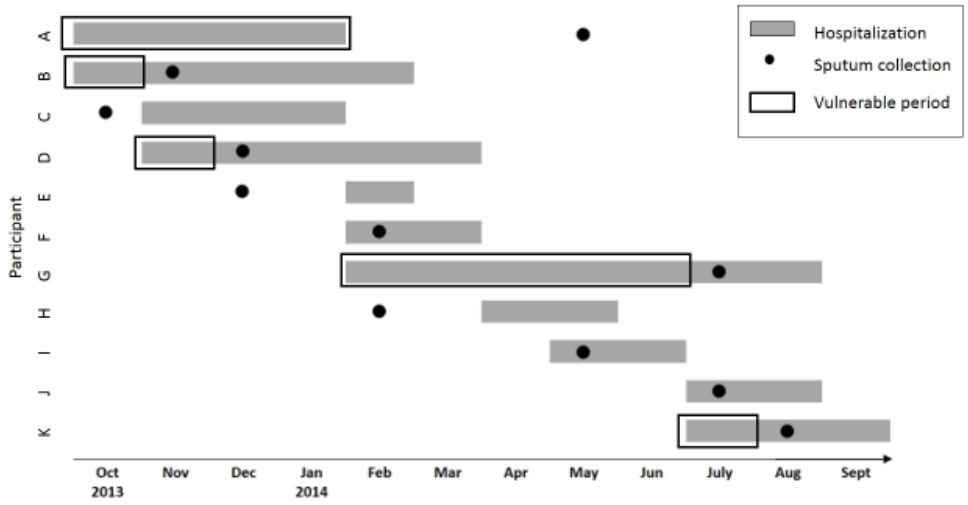


Table S1. Pairwise SNP distances between isolates within the same RFLP group (AH-W) and within the same cluster group (1001-2300). Cluster groups containing fewer than three isolates are excluded. Median pairwise distance is shown for the closest participant within a cluster and for the overall cluster.

					SNP distance	
Cluster	Family	Lineage	# Isolates	# WGS	Median closest	Median overall
10(AH)	X	4	15	8		14
1003			10	6	6	11.5
11(BF)	T	4	3	1	.	.
12(BH)	S	4	6	4		13
13(BM)	S	4	2	1	.	.
14(BW)	Haarlem	4	9	5		18.5
1401			6	3	6	24
15(CC)	LAM	4	7	4		6.5
16(GD)	CAS	3	2	0	.	.
17(GY)	T	4	15	11		20
1701			4	3	10	20
1702			4	4	9	16
18(HP)	LAM	4	285	230		18
1801			4	4	7	23
1802			3	3	23	23
1805			4	4	19.5	23.5

1810			14	13	8	21.5
1817			5	5	10	15.5
1818			208	163	5	16
1823			8	8	10.5	17.5
19(KO)	S	4	1	1	.	.
20(KR)	T	4	1	1	.	.
21(M)	T	4	1	1	.	.
22(MH)	Beijing	2	28	20		18
2201			5	3	14	21
2207			4	3	6	10
23(W)	Beijing	2	7	4		211.5

WGS: Whole genome sequencing

References

1. van Embden JD, Cave MD, Crawford JT, et al. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406-9.

2. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-60.

3. Li H, Handsaker B, Wysoker A, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078-9.

4. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30:1312-3.