Epidemiology, genetic basis and outcomes of bedaquiline resistant tuberculosis in South Africa: 2015-2019

Supplemental Appendix

Table of Contents

1. Supplementary Methods

1.1. Whole Genome Sequencing methodology and bioinformatic analysis for variant calling, lineage assignment and phylogenetic analysis

Whole genome sequencing (WGS) was performed on all *Mycobacterium tuberculosis complex* isolated from liquid culture on the MiSeq (Illumina, UK) as previously described(1). Briefly, the Nextera-XT or Nextera DNA Flex library preparation kit (Illumina, UK) was used to perform library preparation followed by sequencing using the 2 x 300bp MiSeq cartridge v.3 (Illumina, UK) with a targeted 100x depth of coverage.

Resequencing analysis was carried out on CLC Genomics Workbench 11.0.2 (Qiagen, Venlo, Netherlands). Single nucleotide polymorphisms (SNPs) were identified by reference mapping of the paired-end reads against the reference genome H37Rv (NC000962.3), with quality-based and coverage filters applied to each SNP using the following parameters Q-score ≥ 20, a frequency of 10% or more provided at least 5 reads supported the snp in both sequence directions (forward and reverse) as evidence supports a ≥5% frequency to most likely be clinically relevant(2), we set 10% as to increase stringency. Resistance-associated SNPs were identified based on the TB Drug Resistance Mutation Database (TBDReaMDB)(3), TBProfiler(4) and PhyResSE(5) mutation catalogues. If a SNP did not exist in the any of these catalogues, a literature search was conducted to identify association with resistance.

The SNP barcode described by Coll et al (2014)(6) was used to assign the *M. tuberculosis* lineage to each isolate. In brief, variant files containing a list of SNPs for each genome were generated on CLC Genomics Workbench 11.0.2 (Qiagen, Venlo, Netherlands). The SNPs were concatenated to an alignment and phylogeny was inferred based on a comparison of SNP alignments of strains. Recent transmission events may were inferred between isolates with a <6 SNP difference(7), applying this cut-off resulted in a higher stringency to assume relationships based on a genetic association only. Maximum-likelihood trees were generated using IQ-TREE(8) based on the SNP alignments. Visualization and annotation of the trees was performed on ITOL v.4.4(9).

1.2. Datasets, variables and considerations for analysis

The three datasets (NICD surveillance, routine NHLS laboratory data and EDRWeb) were linked using a combination of deterministic (sample reference number, ID number) and probabilistic methods (name, surname, date of birth, sex, facility, folder number, treatment and sample dates,

and province) with manual review. HIV, antiretroviral status, specialised treatment centre, and DR-TB classification (Rif-R, MDR, Pre-XDR, or XDR) were extracted from EDRWeb using standard definitions. Age was categorised into 6 10-year age bands. Guideline period was classified into 3 categories according to date of dissemination of guidelines. Prior MDR-TB, bedaquiline and clofazimine data were obtained from the BDQ surveillance laboratory request forms and the EDRWeb database.

Observational studies are subject to various sources of bias including selection bias, information bias and differential misclassification bias. Missing data were not imputed in this study. As noted, the surveillance could not capture all patients started on bedaquiline-based treatment, particularly in the later years, and transmission is probably under-estimated, requiring further research. Reporting of prior BDQ/CFZ exposure may have been under reported, but findings are consistent with other data. Only a subset of bedaquiline-susceptible isolates had WGS and may have been biased towards those with poor treatment, e.g. those with at least a single post-baseline positive culture. We may thus have over-represented resistance emergence frequency, and the findings may represent a worst-case scenario. Furthermore, outcomes were not always interpretable, for example, patients who were transferred out. We were only able to analyse those with clearly classifiable outcomes, which may have led to further bias; however, this represents real-world and national data, which provides important insights not otherwise observable.

2. Supplementary Results

2.1. Children (<12 years) excluded in cross sectional analysis

Of the 32 children less than 12 years of age excluded, 22 children would have been excluded for other reasons including negative culture results. Of the 10 remaining children with BDQ resistance results, 9 were classified as BDQ-S and 1 was BDQ-R. It was not possible to draw any conclusions regarding acquisition of BDQ-R from these 10 children.

2.2. Risk factor analysis

Multivariable models where variable selection is not informed by directed acyclic graphs (DAGs) to assess the direct or indirect effect of a primary exposure are difficult to interpret (the table two fallacy)(10) and can introduce bias through for example introducing associations between variables which are independent or adjusting for variables which are on the causal pathway. We

have thus only included only a univariate analysis in the manuscript but for completeness we have included two multivariable models with the outputs shown in Tables S1 and S2 respectively. The first model includes: Age, sex, HIV status, province, specialized treatment centre, previous MDR, previous BDQ/CFZ exposure, guideline period, and drug resistance classification. The second, more parsimonious, model includes: Age, sex, HIV status, previous BDQ/CFZ exposure, and drug resistance classification. The measures of strength of effect of both models are very similar to the univariate results and do not change conclusions.

2.3. Patient outcomes by phenotype and sequencing results

Outcomes by phenotype and sequencing results, which was only performed on a subset (15%;171/1140) of patient isolates having interpretable outcomes and WGS data completed are shown in Table S5. Of the phenotypically susceptible isolates, 4.6% (6/130) had a Rv0678 mutation, and successful outcomes were no different between those that were wild type or mutant (OR 1.0, 95%CI: 0.2-5.4). Of the phenotypically resistant isolates, the majority (83.3%; 30/36) had a Rv0678 mutation and successful outcomes were lower in those with a mutation compared with those without a mutation but not statistically significant (OR 0.4, 95%CI: 0.1-2.4).

3. Supplementary Tables and Figures

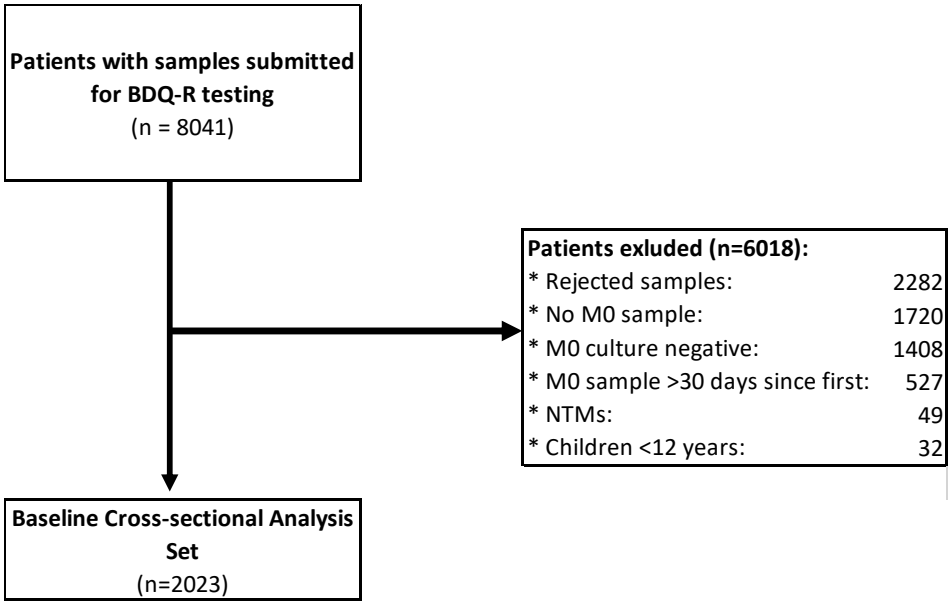

***Figure S1**:* Flow diagram of longitudinal analysis set

M0 = month 0 (baseline), NTM= non-tuberculosis mycobacteria

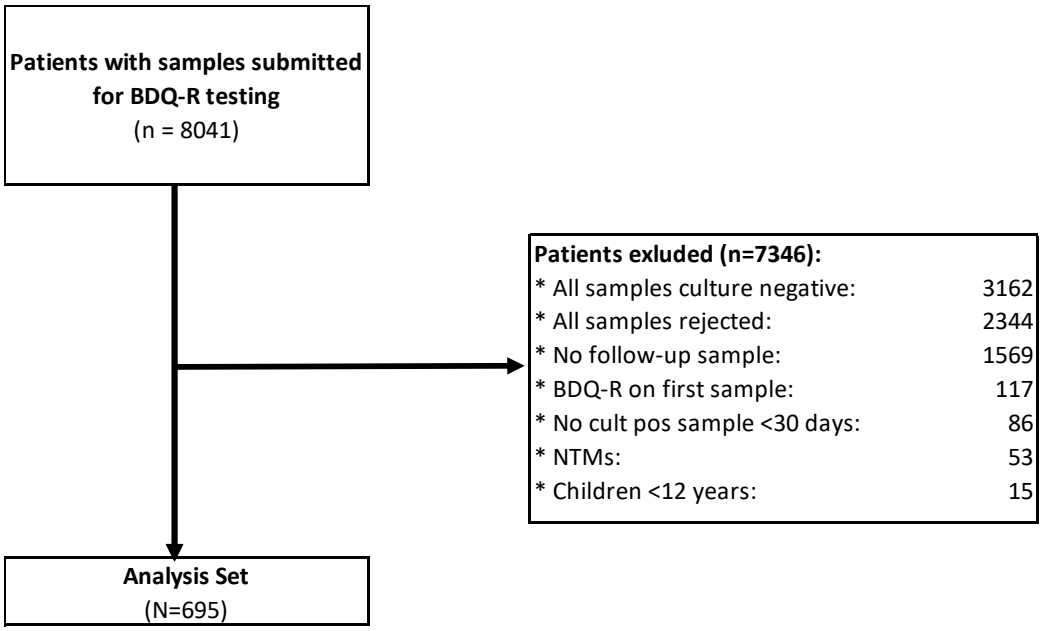

Figure S2: Flow diagram of longitudinal analysis set

M0 = month 0 (baseline), NTM= non-tuberculosis mycobacteria

Table S1: Odds ratios of risk factors associated with bedaquiline resistance (unadjusted and adjusted – model 1)

		BDQ-S (n=1947)	BDQ-R (n=76)	Total (N=2023)	OR (95%CI)	p-value	adjusted OR (95%CI)	p-value
Sex (n=2022)								
Female		794 (96%)	33 (4%)	827	1		1	
Male		1152 (96%)	43 (4%)	1195	0.9 (0.6 - 1.4)	0.65	0.9 (0.6 - 1.5)	0.72
Age (IQR)		37 (30-37)	34 (28-45)	37(30-47)		0.91		0.18
HIV status (n=1917)								
	Negative	632 (95%)	30 (5%)	662	1		1	
	Positive	1211 (96%)	44 (4%)	1255	0.8 (0.5 - 1.2)	0.51	0.6 (0.4 - 1.1)	0.10
Started on ART (n=1221)								
	No	14 (100%)	0 (0%)	14	1			
	Yes	1164 (96%)	43 (4%)	1207	0.8 (0.5 - 1.4)			
Province (n=2009)								
	Eastern Cape	458 (95%)	23 (5%)	481	1		1	
	Free State	0 (0%)	0 (0%)	0	NA		NA	
	Gauteng	579 (96%)	24 (4%)	603	0.8 (0.4 - 1.4)		1.3 (0.7 - 2.6)	
	KwaZulu-Natal	232 (95%)	13 (5%)	245	1.3 (0.6 - 2.7)	0.32	1.7 (0.8 - 3.7)	0.60
	Limpopo	11 (100%)	0 (0%)	11	NA		NA	
	Mpumalanga	66 (99%)	1 (1%)	67	0.6 (0.1 - 2.6)		1.3 (0.1 - 13.8)	
	North West	65 (98%)	1 (2%)	66	0.3 (0.1 - 2.4)		0.6 (0.1 - 4.5)	

	Northern Cape	144 (97%)	5 (3%)	149	0.6 (0.2 - 1.7)		0.9 (0.3 - 2.4)	
	Western Cape	378 (98%)	9 (2%)	387	0.5 (0.2 - 1.1)		0.7 (0.3 - 1.7)	
Centre of Excellence (n=2023)								
	No	379 (99%)	4 (1%)	383	1		1	
	Yes	1568 (96%)	72 (4%)	1640	4.35 (1.6 - 11.9)	0.002	3.6 (1.0- 12.1)	0.04
Previous MDR (n=2006)								
	No	1681 (96%)	65 (4%)	1746	1		1	
	Yes	249 (96%)	11 (4%)	260	1.1 (0.6 - 2.2)	0.5	1.1 (0.5 - 2.3)	0.78
Previous BDQ exposure (n=2006)								
	No	1922 (96%)	75 (4%)	1997	1			
	Yes	8 (89%)	1 (11%)	9	3.2 (0.4 - 25.9)	0.25		
Previous CFZ exposure (n=2006)								
	No	1916 (96%)	72 (4%)	1988	1			
	Yes	14 (78%)	4 (22%)	18	7.6 (2.4 - 23.7)	<0.001		
Previous BDQ/CFZ exposure (n=2006)								
	No	1915 (96%)	72 (4%)	1987	1		1	
	Yes	15 (79%)	4 (21%)	19	7.1 (2.3 - 21.9)	<0.001	5.2 (1.5 - 17.2)	<0.001
Guideline period								
	XDR/pre-XDR/AE	1074 (96%)	36 (4%)	1110	1		1	
	decentralisation	523 (95%)	25 (5%)	548	1.4 (0.8 - 2.4)	0.29	1.3 (0.8 - 2.3)	0.63
	all RR patients	350 (96%)	15 (4%)	365	1.3 (0.7 - 2.4)		1.2 (0.6 - 2.3)	
EDR DR Classification (n=1924)								
	Rif-R	433 (98%)	7 (2%)	440	1	<0.001	1	0.004

	MDR	608 (97%)	18 (3%)	626	1.8 (0.8 - 4.2)	2.1 (0.8 - 5.1)
	Pre-XDR	406 (95%)	23 (5%)	429	3.5 (1.5 - 8.3)	4.0 (1.6 - 9.5)
	XDR	403 (94%)	26 (6%)	429	4.0 (1.7- 9.3)	4.4 (1.8 - 10.9)
Individual Drug resistance profiles at baseline testing						
Fluoroquinolone Resistance (n=1515)						
	Sensitive	887 (98%)	14 (2%)	901	1	
	Resistant	571 (93%)	43 (7%)	614	4.8 (2.6 - 8.8)	<0.001
Second line inhjectables (n=1513)						
	Sensitive	1016 (97%)	32 (3%)	1048	1	
	Resistant	440 (95%)	25 (5%)	465	1.8 (1.1 - 3.1)	<0.001
Linezolid (n=1093)						
	Sensitive	1038 (96%)	41 (4%)	1079	1	
	Resistant	11 (79%)	3 (21%)	14	6.9 (1.9 - 25.7)	0.001
Clofazamine (n=1092)						
	Sensitive	942 (99%)	13 (1%)	955	1	
	Resistant	106 (77%)	31 (23%)	137	21.2 (7.3 - 41.7)	<0.001
Rifabutin (n=438)						
	Sensitive	209 (99%)	3 (1%)	212	1	
	Resistant	213 (94%)	13 (4%)	226	4.3 (1.2 - 15.1)	0.02

Table S2: Odds ratios of risk factors associated with bedaquiline resistance (unadjusted and adjusted – model 2)

	BDQ-S (n=1947)	BDQ-R (n=76)	Total (N=2023)	OR (95%CI)	p-value	adjusted OR (95%CI) (N=1,876)	p-value	adjusted OR (95%CI) (N=1,894)	p-value
Sex (n=2022)									
Female	794 (96%)	33 (4%)	827	1		1		1	
Male	1152 (96%)	43 (4%)	1195	0.9 (0.6 - 1.4)	0.65	0.9 (0.6 - 1.5)	0.78	0.9 (0.6 - 1.5)	0.75
Age category (n=2007)									
12-19	75 (95%)	4 (5%)	79	1.2 (0.4 - 3.6)		1.1 (0.3 - 3.3)		1.2 (0.4 - 3.7)	
20-29	379 (95%)	19 (5%)	398	1.2 (0.6 - 2.1)		1.1 (0.6 - 2.1)		1.1 (0.6 - 2.1)	
30-39	653 (96%)	28 (4%)	681	1		1		1	
40-49	442 (97%)	23 (3%)	455	0.7 (0.4 - 1.3)	0.57	0.7 (0.3 - 1.4)	0.60	0.7 (0.3 - 1.3)	0.58
50-59	288 (97%)	8 (3%)	296	0.6 (0.3 - 1.4)		0.6 (0.3 - 1.5)		0.7 (0.3 - 1.6)	
60+	95 (97%)	3 (3%)	98	0.7 (0.2 - 2.5)		0.5 (0.1 - 2.0)		0.5 (0.1 - 2.0)	
HIV status (n=1917)									
Negative	632 (95%)	30 (5%)	662	1		1		1	
Positive	1211 (96%)	44 (4%)	1255	0.8 (0.5 - 1.2)	0.51	0.6 (0.4 - 1.1)	0.10	0.8 (0.5 - 1.3)	0.32
Started on ART (n=1221)									
No	14 (100%)	0 (0%)	14	1					
Yes	1164 (96%)	43 (4%)	1207	0.8 (0.5 - 1.4)					
Province (n=2009)									
Eastern Cape	458 (95%)	23 (5%)	481	1	0.32	1	0.60		

	Free State	0 (0%)	0 (0%)	0	NA		NA	
	Gauteng	579 (96%)	24 (4%)	603	0.8 (0.4 - 1.4)		1.3 (0.7 - 2.6)	
	KwaZulu-Natal	232 (95%)	13 (5%)	245	1.3 (0.6 - 2.7)		1.7 (0.8 - 3.7)	
	Limpopo	11 (100%)	0 (0%)	11	NA		NA	
	Mpumalanga	66 (99%)	1 (1%)	67	0.6 (0.1 - 2.6)		1.3 (0.1 - 13.8)	
	North West	65 (98%)	1 (2%)	66	0.3 (0.1 - 2.4)		0.6 (0.1 - 4.5)	
	Northern Cape	144 (97%)	5 (3%)	149	0.6 (0.2 - 1.7)		0.9 (0.3 - 2.4)	
	Western Cape	378 (98%)	9 (2%)	387	0.5 (0.2 - 1.1)		0.7 (0.3 - 1.7)	
Specialised treatment centre (n=2023)								
	No	379 (99%)	4 (1%)	383	1		1	
	Yes	1568 (96%)	72 (4%)	1640	4.35 (1.6 - 11.9)	0.002	3.5 (1.0-12.1)	0.04
Previous MDR (n=2006)								
	No	1681 (96%)	65 (4%)	1746	1	0.5	1	0.81
	Yes	249 (96%)	11 (4%)	260	1.1 (0.6 - 2.2)		1.1 (0.5 - 2.3)	
Previous BDQ exposure (n=2006)								
	No	1922 (96%)	75 (4%)	1997	1	0.25		
	Yes	8 (89%)	1 (11%)	9	3.2 (0.4 - 25.9)			
Previous CFZ exposure (n=2006)								
	No	1916 (96%)	72 (4%)	1988	1	<0.001		

	Yes	14 (78%)	4 (22%)	18	7.6 (2.4 - 23.7)				
Previous BDQ/CFZ exposure (n=2006)									
	No	1915 (96%)	72 (4%)	1987	1	<0.001	1	<0.001	1	<0.001
	Yes	15 (79%)	4 (21%)	19	7.1 (2.3 - 21.9)		5.2 (1.5 - 17.2)		6.3 (1.9 - 21.0)	
Guideline period									
	XDR/pre-XDR/AE	1074 (96%)	36 (4%)	1110	1		1			
	decentralisation	523 (95%)	25 (5%)	548	1.4 (0.8 - 2.4)	0.29	1.3 (0.8 - 2.3)	0.63		
	all RR patients	350 (96%)	15 (4%)	365	1.3 (0.7 - 2.4)		1.2 (0.6 - 2.3)			
EDR DR Classification (n=1924)									
	Rif-R	433 (98%)	7 (2%)	440	1		1		1	
	MDR	608 (97%)	18 (3%)	626	1.8 (0.8 - 4.2)		2.1 (0.8 - 5.0)		1.8 (0.8 - 4.4)	
	Pre-XDR	406 (95%)	23 (5%)	429	3.5 (1.5 - 8.3)	<0.001	3.9 (1.6 - 9.5)	0.004	3.4 (1.4 - 8.1)	0.004
	XDR	403 (94%)	26 (6%)	429	4.0 (1.7- 9.3)		3.7 (1.6 - 8.8)		4.3 (1.7 - 10.7)	
Individual Drug resistance profiles at baseline testing										
Fluoroquinolone Resistance (n=1515)										
	Sensitive	887 (98%)	14 (2%)	901	1					
	Resistant	571 (93%)	43 (7%)	614	4.8 (2.6 - 8.8)	<0.001				
Second line inhjectables (n=1513)										
	Sensitive	1016 (97%)	32 (3%)	1048	1	<0.001				

Resistant	440 (95%)	25 (5%)	465	1.8 (1.1 - 3.1)			
Linezolid (n=1093)							
Sensitive	1038 (96%)	41 (4%)	1079	1	0.001		
Resistant	11 (79%)	3 (21%)	14	6.9 (1.9 - 25.7)			
Clofazamine (n=1092)							
Sensitive	942 (99%)	13 (1%)	955	1			
Resistant	106 (77%)	31 (23%)	137	21.2 (7.3 - 41.7)	<0.001		
Rifabutin (n=438)							
Sensitive	209 (99%)	3 (1%)	212	1			
Resistant	213 (94%)	13 (4%)	226	4.3 (1.2 - 15.1)	0.02		

Table S3: Mutations associated with baseline bedaquiline-resistant patient isolates

UID	Gene	Reference Position	Variant Type	Length	Reference	Allele	Count	Coverage	Frequency	Forward Reverse balance	Average quality	Nucleotide	Amino acid
100274	Rv0678	779131	Insertion	1		C	64	79	81.01	0.40	32.67	144dup	Glu49fs
50272	Rv0678	779419	Insertion	2		GC	74	83	89.16	0.49	34.77	430_431dup	Tyr145fs
70019	Rv0678	779350	SNV	1	G	A	97	98	98.98	0.48	34.13	G361A	Gly121Arg
2490127	Rv0678	779197	SNV	1	A	G	59	61	96.72	0.48	33.56	A208G	Asn70Asp
2290165	Rv0678	779190	SNV	1	C	G	18	102	17.65	0.46	35.00	C201G	Ile67Met
2290165	Rv0678	779159	SNV	1	C	A	31	102	30.39	0.49	32.61	C170A	Ala57Glu
2690197	Rv0678		Insertion						100.00			ins_IS6110	
3090146	Rv0678	779454	Insertion	2		GC	61	67	91.04	0.49	33.45	466_467dup	Tyr157fs
60	Rv0678	779125	SNV	1	T	C	89	89	100.00	0.44	33.72	T136C	Cys46Arg
30153	Rv0678	779126	Insertion	1		G	115	124	92.74	0.48	34.23	137dup	Cys46fs
10061	Rv0678	779045	SNV	1	T	C	126	130	96.92	0.47	31.44	T56C	Phe19Ser
4290194	Rv0678	779131	Insertion	1		C	92	99	92.93	0.45	31.71	144dup	Glu49fs
10245	Rv0678	779197	SNV	1	A	G	62	62	100.00	0.49	36.71	A208G	Asn70Asp
1990113	Rv0678	779128	Insertion	1		G	101	105	96.19	0.48	34.04	139dup	Asp47fs
1490134	Rv0678	779275	SNV	1	C	T	120	120	100.00	0.50	32.11	C286T	Arg96Trp
590206	Rv0678	779130	Insertion	2		TC	26	110	23.64	0.46	30.22	141_142dup	Pro48fs
30308	Rv0678	779111	SNV	1	G	T	153	154	99.35	0.49	32.11	G122T	Gly41Val

20009	Rv0678	779454	Insertion	2		GC	188	210	89.52	0.38	36.46	466_467dup	Tyr157fs
4090162	Rv0678	779302	SNV	1	C	T	138	141	97.87	0.47	32.17	C313T	Arg105Cys
2990191	Rv0678	779405	SNV	1	T	C	81	83	97.59	0.46	30.10	T416C	Met139Thr
40195	Rv0678	779128	Insertion	1		G	88	93	94.62	0.49	33.68	139dup	Asp47fs
10243	Rv0678	779189	SNV	1	T	G	127	127	100.00	0.49	33.71	T200G	Ile67Ser
3990247	Rv0678	779131	Insertion	1		C	146	153	95.42	0.48	34.11	144dup	Glu49fs
120201	Rv0678	779131	Insertion	1		C	137	147	93.20	0.49	34.23	144dup	Glu49fs
1790217	Rv0678	779131	Insertion	1		C	21	160	13.13	0.39	35.24	144dup	Glu49fs
1790217	Rv0678	779130	Insertion	2		TC	43	161	26.71	0.41	34.97	141_142dup	Pro48fs
1790217	Rv0678	779450	SNV	1	T	G	69	163	42.33	0.44	33.77	T461G	Leu154Arg
1590205	Rv0678	779270	SNV	1	G	A	13	114	11.40	0.47	33.23	G281A	Arg94Gln
1590205	Rv0678	779313	Deletion	1	C		16	112	14.29	0.50	30.63	325delC	Arg109fs
1590205	Rv0678	779296	SNV	1	G	A	20	115	17.39	0.28	32.25	G307A	Gly103Ser
190230	Rv0678	779131	Insertion	1		C	129	135	95.56	0.46	35.52	144dup	Glu49fs
140085	Rv0678	779276	SNV	1	G	T	163	164	99.39	0.49	35.21	G287T	Arg96Leu
2790220	Rv0678	779128	Insertion	4		GATC	77	96	80.21	0.49	35.62	139_142dup	Pro48fs
20149	Rv0678	779132	SNV	1	C	A	45	56	80.36	0.50	32.64	C143A	Pro48His
2690222	Rv0678	779128	Insertion	1		G	145	155	93.55	0.47	34.09	139dup	Asp47fs
2290170	Rv0678	779128	Insertion	1		G	150	159	94.34	0.46	33.85	139dup	Asp47fs

2890142	Rv0678	779015	SNV	1	A	T	12	93	12.90	0.21	37.00	A26T	Gln9Leu
2890142	Rv0678	779008	SNV	1	G	T	12	91	13.19	0.38	33.17	G19T	Val7Phe
2890142	Rv0678	779011	SNV	1	G	A	14	92	15.22	0.47	35.36	G22A	Asp8Asn
2890142	Rv0678	779125	SNV	1	T	C	54	96	56.25	0.49	35.54	T136C	Cys46Arg
3890165	Rv0678	779399	SNV	1	G	C	14	125	11.20	0.48	34.29	G410C	Arg137Pro
3890165	Rv0678	779191	SNV	1	A	C	30	137	21.90	0.49	33.13	A202C	Ser68Arg
3890165	Rv0678	779383	SNV	1	C	T	32	130	24.62	0.43	34.47	C394T	Arg132*
10209	Rv0678	779131	Insertion	1		C	6	61	10.00	0.44	37.83	144dup	Glu49fs
220299	Rv0678	779082	Deletion	1	T		110	112	98.21	0.49	34.28	95delT	Leu32fs
90151	Rv0678	779302	SNV	1	C	T	153	158	96.84	0.47	32.16	C313T	Arg105Cys
20250	Rv0678	779131	Insertion	1		C	195	208	93.75	0.49	34.39	144dup	Glu49fs
3390140	Rv0678	779389	SNV	1	C	T	135	148	91.22	0.46	33.73	C400T	Arg134*
2590159	Rv0678	779131	Insertion	1		C	156	169	92.31	0.47	35.20	144dup	Glu49fs
590196	Rv0678	779131	Insertion	1		C	56	137	40.88	0.48	34.95	144dup	Glu49fs
590196	Rv0678	779128	Insertion	1		G	71	137	51.82	0.37	32.72	139dup	Asp47fs
2790184	Rv0678	779138	Insertion	1		G	113	115	98.26	0.49	34.75	150dup	Gln51fs
20262	Rv0678	779182	Insertion	1		G	91	100	91.00	0.48	35.33	198dup	Ile67fs
60231	Rv0678	779048	SNV	1	T	C	20	73	27.40	0.45	30.30	T59C	Val20Ala
60231	Rv0678	779053	SNV	1	C	T	48	71	67.61	0.50	32.77	C64T	Gln22*

1290213	Rv0678	779236	SNV	1	C	T	31	102	30.39	0.35	34.81	C247T	Leu83Phe
1290213	Rv0678	779095	SNV	1	G	A	63	93	67.74	0.49	34.06	G106A	Ala36Thr
190256	Rv0678	779138	Insertion	1		G	103	106	97.17	0.50	35.72	150dup	Gln51fs
200194	Rv0678	779257	SNV	1	C	T	71	72	98.61	0.50	31.79	C268T	Arg90Cys
990157	Rv0678	779131	Insertion	1		C	135	141	95.74	0.45	34.33	144dup	Glu49fs
4090191	Rv0678	779190	SNV	1	C	G	186	186	100.00	0.45	33.80	C201G	Ile67Met
890226	Rv0678	779059	SNV	1	G	C	12	126	10.00	0.44	34.67	G70C	Gly24Arg
890226	Rv0678	779372	Deletion	1	C		66	129	51.16	0.48	29.83	386delC	Pro129fs
1390198	Rv0678	779131	Insertion	1		C	22	35	62.86	0.49	34.91	144dup	Glu49fs
2690213	Rv0678	779339	SNV	1	T	G	102	128	79.69	0.46	33.63	T350G	Leu117Arg
20276	Rv0678	779057	Insertion	1		T	15	149	10.07	0.47	33.80	68dup	Met23fs
20276	Rv0678	779182	Insertion	1		G	25	159	15.72	0.38	34.60	198dup	Ile67fs
20276	Rv0678	779131	Insertion	1		C	26	145	17.93	0.47	32.77	144dup	Glu49fs
20276	Rv0678	779188	Insertion	1		A	56	158	35.44	0.48	34.52	199dup	Ile67fs
2890198	Rv0678	779115	SNV	1	G	A	41	142	28.87	0.39	34.27	G126A	Trp42*
2890198	Rv0678	779460	SNV	1	C	G	120	178	67.42	0.41	36.18	C471G	Tyr157*
1990103	Rv0678	779182	Insertion	1		G	9	68	13.24	0.38	33.78	198dup	Ile67fs
1990103	Rv0678	779209	SNV	1	C	G	18	76	23.68	0.36	29.44	C220G	Leu74Val
1990103	Rv0678	779414	SNV	1	T	G	21	83	25.30	0.48	31.76	T425G	Leu142Arg

* - stop codon

Table S4: Mutations associated with emergent bedaquiline-resistant patient isolates

UID	Gene	Reference Position	Type	Length	Reference	Allele	Count	Coverage	Frequency	Forward reverse balance	Average quality	Nucleotide	Amino acid
590189	atpE	1461227	SNV	1	G	C	224	225	99.56	0.38	35.19	G183C	Glu61Asp
220052	Rv0678	779128	Insertion	1		G	21	113	18.58	0.41	36.00	139dup	Asp47fs
150267	Rv0678	779123	SNV	1	T	G	14	76	18.42	0.14	22.86	T134G	Val45Gly
150267	Rv0678	779123	SNV	1	T	G	16	89	17.98	0.06	23.25	T134G	Val45Gly
220052	Rv0678	779126	Insertion	1		G	20	112	17.86	0.42	35.15	137dup	Cys46fs
150267	Rv0678	779117	SNV	1	T	G	14	91	15.38	0.14	21.50	T128G	Leu43Arg
220052	Rv0678	779117	Deletion	10	TGCTGGTGTG		18	117	15.38	0.43	34.95	130_139delCTGGTGTGTG	Leu44fs
20178	Rv0678	779120	SNV	1	T	G	15	98	15.31	0.07	21.87	T131G	Leu44Arg
150267	Rv0678	779363	SNV	1	T	G	14	92	15.22	0.07	24.36	T374G	Leu125Arg
141	Rv0678	779354	SNV	1	T	G	19	126	15.08	0.11	20.63	T365G	Leu122Arg
150267	Rv0678	779438	SNV	1	T	G	16	107	14.95	0.06	22.19	T449G	Val150Gly
150267	Rv0678	779125	SNV	1	T	G	14	95	14.74	0.07	20.71	T136G	Cys46Gly
150267	Rv0678	779300	SNV	1	A	C	16	111	14.41	0.35	22.88	A311C	Glu104Ala
150267	Rv0678	779348	SNV	1	T	G	13	95	13.68	0.15	23.77	T359G	Val120Gly

141	Rv0678	779161	SNV	1	A	C	15	110	13.64	0.13	21.73	A172C	Thr58Pro
150267	Rv0678	779369	SNV	1	A	C	13	97	13.40	0.40	20.31	A380C	Asp127Ala
150267	Rv0678	779450	SNV	1	T	G	14	112	12.50	0.21	21.71	T461G	Leu154Arg
20178	Rv0678	779003	SNV	1	A	C	13	105	12.38	0.08	22.31	A14C	Asp5Ala
20178	Rv0678	779146	SNV	1	T	G	13	105	12.38	0.08	22.92	T157G	Ser53Ala
20178	Rv0678	779135	SNV	1	A	C	12	99	12.12	0.33	20.92	A146C	Glu49Ala
90092	Rv0678	779120	SNV	1	T	G	30	253	11.86	0.16	21.37	T131G	Leu44Arg
141	Rv0678	779135	SNV	1	A	C	13	110	11.82	0.38	20.92	A146C	Glu49Ala
160078	Rv0678	779369	SNV	1	A	C	29	250	11.60	0.21	21.03	A380C	Asp127Ala
150267	Rv0678	779336	SNV	1	A	C	11	95	11.58	0.18	21.91	A347C	Asp116Ala
150267	Rv0678	779435	SNV	1	T	G	12	104	11.54	0.17	21.08	T446G	Val149Gly
90092	Rv0678	779125	SNV	1	T	G	29	252	11.51	0.13	20.03	T136G	Cys46Gly
220052	Rv0678	779131	Insertion	1		C	13	114	11.40	0.47	33.62	144dup	Glu49fs
150267	Rv0678	779444	SNV	1	A	C	12	107	11.21	0.08	23.08	A455C	Asp152Ala
141	Rv0678	779369	SNV	1	A	C	14	126	11.11	0.19	24.00	A380C	Asp127Ala
20178	Rv0678	779461	SNV	1	A	C	13	117	11.11	0.21	20.08	A472C	Ser158Arg
150267	Rv0678	779243	SNV	1	T	G	9	81	11.11	0.11	20.44	T254G	Val85Gly
150267	Rv0678	779450	SNV	1	T	G	12	111	10.81	0.33	21.17	T461G	Leu154Arg
80143	Rv0678	779027	SNV	1	A	C	25	249	10.04	0.16	20.04	A38C	Glu13Ala
141	Rv0678	779117	SNV	1	T	G	10	100	10.00	0.27	20.10	T128G	Leu43Arg

Table S5: Treatment outcomes among a subset of patient isolates with whole genome sequencing stratified by Rv0678 status

Outcome	Rv0678 wild type	Rv0678 mutant	OR (95%CI)
Successfully outcome among phenotypically bedaquiline-susceptible patients (N=130)			
	124	**6**	
No	40 (32.3%)	2 (33.3%)	1
Yes	84 (67.7%)	4 (66.7%)	1.0 (0.2 – 5.4)
Successfully outcome among phenotypically bedaquiline resistant patients (N=36)			
	6	30	
No	2 (33.3%)	17 (56.7%)	1
Yes	4 (66.7%)	13 (43.3%)	0.4 (0.1 – 2.4)

OR= odds ratio, 95%CI = 95% confidence interval.

4. STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	3
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	5-6
Objectives	3	State specific objectives, including any prespecified hypotheses	6
Methods			
Study design	4	Present key elements of study design early in the paper	7-8
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	7-9
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	8 and appendix
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	NA
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	8-9 and Appendix
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	8-9 and Appendix

Bias	9	Describe any efforts to address potential sources of bias	15-16 and Appendix
Study size	10	Explain how the study size was arrived at	NA
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	Appendix
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8
		(*b*) Describe any methods used to examine subgroups and interactions	NA
		(*c*) Explain how missing data were addressed	Appendix
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	8
		(*e*) Describe any sensitivity analyses	NA

5. References

1. Ismail NA, Ismail F, Joseph L, Govender N, Blows L, Kaniga K, et al. Epidemiological cut-offs for Sensititre susceptibility testing of Mycobacterium tuberculosis: interpretive criteria cross validated with whole genome sequencing. Sci Rep. 2020;10(1):1013.
2. Nimmo C, Brien K, Millard J, Grant AD, Padayatchi N, Pym AS, et al. Dynamics of within-host Mycobacterium tuberculosis diversity and heteroresistance during treatment. EBioMedicine. 2020;55:102747.
3. Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB. Tuberculosis drug resistance mutation database. PLoS medicine. 2009;6(2):e2.
4. Coll F, McNerney R, Preston MD, Guerra-Assuncao JA, Warry A, Hill-Cawthorne G, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. Genome Med. 2015;7(1):51.
5. Feuerriegel S, Schleusener V, Beckert P, Kohl TA, Miotto P, Cirillo DM, et al. PhyResSE: a Web Tool Delineating Mycobacterium tuberculosis Antibiotic Resistance and Lineage from Whole-Genome Sequencing Data. J Clin Microbiol. 2015;53(6):1908-14.
6. Coll F, McNerney R, Guerra-Assuncao JA, Glynn JR, Perdigao J, Viveiros M, et al. A robust SNP barcode for typing Mycobacterium tuberculosis complex strains. Nat Commun. 2014;5:4812.
7. Nikolayevskyy V, Kranzer K, Niemann S, Drobniewski F. Whole genome sequencing of Mycobacterium tuberculosis for detection of recent transmission and tracing outbreaks: A systematic review. Tuberculosis (Edinb). 2016;98:77-85.

8. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32(1):268-74.
9. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 2019;47(W1):W256-W9.
10. Westreich D, Greenland S. The Table 2 Fallacy: Presenting and Interpreting Confounder and Modifier Coefficients. American Journal of Epidemiology. 2013;177(4):292-8.