Deciphering Bedaquiline and Clofazimine Resistance in Tuberculosis: An Evolutionary Medicine Approach

Lindsay Sonnenkalb[+,1], Joshua Carter[+,2], Andrea Spitaleri[5,6], Zamin Iqbal[7], Martin Hunt[7,8], Kerri Malone[7], Christian Utpatel[1], Daniela Maria Cirillo[5], Camilla Rodrigues[9], Kayzad S. Nilgiriwala[10], the CRyPTIC Consortium[11], Philip W. Fowler[&, 8], Matthias Merker[&,1], Stefan Niemann[&,1,3,4]

Affiliations

[1]Molecular and Experimental Mycobacteriology, Research Center Borstel Leibniz Lung Center, Borstel Germany

[2]Medical Scientist Training Program, Stanford University, Stanford, USA

[3]National Reference Center, Research Center Borstel Leibniz Lung Center, Borstel Germany

[4]German Centre for Infection Research (DZIF), Partner site Hamburg-Lübeck-Borstel, Borstel, Germany

[5]Emerging Bacterial Pathogens Unit, Division of Immunology, Transplantation and Infectious Diseases, IRCCS San Raffaele Scientific Institute, Milan, Italy

[6]Center for Omics Sciences, IRCCS San Raffaele Scientific Institute, Milan, Italy

[7]European Bioinformatics Institute, Cambridge, UK.

[8]Experimental Medicine, Nuffield Department of Medicine, University of Oxford, UK

[9]Department of Microbiology, P. D. Hinduja National Hospital and Medical Research Centre, Mumbai, India

[10]The Foundation for Medical Research, Mumbai, India

[11]Comprehensive Resistance Prediction for Tuberculosis: an International Consortium. Author list in supplement.

[+]equally contributing first authors

[&]equally contributing last authors

*Corresponding author, email: sniemann@fz-borstel.de

Abstract

Bedaquiline (BDQ) and clofazimine (CFZ) are core drugs for treatment of multidrug resistant tuberculosis (MDR-TB), however, our understanding of the resistance mechanisms for these drugs is sparse which is hampering rapid molecular diagnostics. To address this, we employed a unique approach using experimental evolution, protein modelling, genome sequencing, and minimum inhibitory concentration data combined with genomes from a global strain collection of over 14,151 *Mycobacterium tuberculosis* complex isolates. Overall, 189 genomic variants causing elevated BDQ and/or CFZ MICs could be discerned, with 175 (95.1%) variants affecting the transcriptional repressor (Rv0678) of an efflux system (mmpS5-mmpL5). Structural modelling of Rv0678 suggests four major mechanisms that confer resistance: impairment of DNA binding, reduction in protein stability, disruption of protein dimerization, and alteration in affinity for its fatty acid ligand. These modelling and experimental techniques will improve personalized medicine in an impending drug resistant era.

40 **Introduction**

41 Multidrug-resistant (MDR: resistance to at least isoniazid [INH] and rifampicin [RMP]) Mycobacterium
42 tuberculosis complex (Mtbc) strains represent a serious challenge for global tuberculosis (TB) control [1,2].
43 The World Health Organization (WHO) estimates that close to half a million people worldwide were
44 infected in 2019 with an RMP-resistant (RR) Mtbc strain, of whom 78% were MDR [1].

45 Compared to patients with sensitive TB, treating MDR-TB treatment takes longer (years), the drugs are
46 less effective and more toxic, and cure rates are low (about 60% globally) [1,3]. As a consequence, ineffective
47 MDR-TB regimens in many high incidence settings have led to an expansion of drug resistant Mtbc strains.
48 The WHO has recommended new treatment regimens for MDR-TB, including new and repurposed drugs
49 such as bedaquiline (BDQ) and clofazimine (CFZ) [4–7]. Both drugs are now central to MDR-TB therapy, and
50 are also part of the recently WHO-endorsed shorter, all-oral MDR-TB regimen [7,8].

51 Selection experiments have found resistance to BDQ and or CFZ as mediated by mutations in the genes
52 *atpE*, *pepQ*, *Rv1979c* [9–11], and *Rv0678* - the latter of which seems to be the most clinically relevant and
53 strains with BDQ associated mutations in *Rv0678* and *atpE* have been selected in multiple patients[12–17].
54 Rv0678 is a marR-like transcriptional regulator of the mmpS5-mmpL5 efflux pump with a canonical N-
55 terminal winged helix-turn-helix (wHTH) DNA binding domain and a C-terminal helical dimerization
56 domain. The Rv0678 protein forms a homodimer which binds to the promotor region of the mmpS5-
57 mmpL5 operon [18–21]. This repression is reversed upon the conformational change of the Rv0678 molecule
58 in the presence of a fatty acid ligand [20]. Importantly, Mtbc strains with mutations in *Rv0678* have been
59 shown to confer resistance against both BDQ and CFZ [22,23], thus undermining the recently WHO-endorsed
60 treatment guidelines [8]. Moreover, mutations in *Rv0678* are the most common resistance mechanism to
61 CFZ and BDQ found in patient isolates today [16,17,24–29].

62 So far, mutations scattered across the full gene length of *Rv0678* (498bp) have been found to occur *in*
63 *vitro* and in patients isolates leading to variable shifts in BDQ and CFZ minimum inhibitory concentrations
64 (MICs) [13,29,30]. However, a comprehensive understanding of the associations between *Rv0678* mutations
65 and their resulting phenotype, as well as their structural effects on the transcriptional repressor Rv0678,
66 is lacking.

67 In this paper we present the results of an experimental model that evolves Mtbc strains under sub-lethal
68 drug concentrations. Such a weak selection pressure has been shown to select for diverse pheno- and
69 genotypes in other bacterial species and likely also reflects the physiological conditions in TB patients [31–
70 33]. Recent studies indicate the inability of most drugs to diffuse into the granuloma at therapeutic levels,
71 including CFZ and BDQ which strongly bind caseum macromolecules [34–38]. BDQ has further been shown to
72 accumulate in cellular compartments of various types of macrophages (including neutrophils), but was
73 unevenly distributed in intracellular bacteria [39,40]. Single resistant clones and total bacterial populations
74 from consecutive cultivation passages were selected at clinically relevant drug concentrations, analyzed
75 by whole genome sequencing (WGS), and compared with mutations identified in 14,151 patient isolates
76 collected by the CRyPTIC consortium. Selected mutations were further investigated computationally for
77 their putative effects on the Rv0678 protein structure and function, which provides a starting point for
78 design of computational algorithms to predict drug resistance. The data generated provide a
79 comprehensive resistance mutation catalogue for genotypic BDQ/CFZ drug susceptibility testing and will
80 improve MDR-TB treatment design and likely outcomes for patients receiving a BDQ and/or CFZ
81 containing therapies.

82 **Results**

83 **Experimental evolution under sub-lethal drug concentrations**

84 To better understand the resistance mechanisms of Mtbc strains against BDQ (which often includes
85 correlated resistance to CFZ), we evolved Mtbc strains under sub-lethal drug concentrations. The
86 experimental set-up was developed to allow for a high number of bacterial generations under a moderate
87 drug selection pressure, thus increasing the mutation diversity compared to classical mutant selection
88 experiments, i.e. exposing a culture at a late exponential phase to the critical drug concentrations (Figure
89 S1). Our H37Rv laboratory strain (ATCC 27294) MIC was 0.25 mg/L for BDQ and 0.5 mg/L for CFZ, in high
90 bacterial density liquid culture (10^7 CFU/mL inoculum) – we shall refer to these as the wild-type (WT)
91 MICs. During five culture passages spread over 20 days, we exposed H37Rv to either BDQ or CFZ at one of
92 four concentrations, ranging from one-half to one-sixteenth of the WT MIC (Figure S2), and selected
93 colonies on solid media (7H11 plates) supplemented with 0.12 mg/L (1:2 MIC) and 0.25 mg/L (MIC) for
94 BDQ and 0.25 mg/L (1:2 MIC) for CFZ. At passages 1, 3, and 5, single colonies as well as the whole bacterial

population were genotypically characterized, and single selected mutations were phenotypically defined (elevated MIC) (Tables 1, S1).

Strikingly, exposure to BDQ or CFZ concentrations as low as one-eighth of the WT MIC over 20 days (12 to 19 generations) led to the selection and enrichment of significantly more resistant cells as compared to growth in the absence of the drugs ($P=2.8 \times 10^{-10}$, and $P=2.4 \times 10^{-3}$, respectively, Figure S2). As expected, there was less enrichment of resistant populations at early timepoints, e.g. after 4 days.

Table 1: *In vitro* selected mutations conferring BDQ resistance. Genotype of each variant was determined by whole genome sequencing, and substitution annotated by coding position, "alternative allele" denotes base pair change. Minimum inhibitory concentration (MIC) was determined by resazurin microtiter plate assay and compared to wild type (WT) susceptible ancestor to describe MIC fold increase. The number of times a mutation was independently selected was included under "# of clones". Finally, from which experiment (BDQ or CFZ) the mutant was isolated from included as "isolate from".

Gene	Position	Substitu-tion	Type	Alt. allele	# of clones	BDQ MIC (mg/L)	BDQ MIC fold increase	CFZ MIC (mg/L)	CFZ MIC fold increase	Isolated from	Drug concentration of selection media (mg/L)
	WT ancestor					0.25-0.5		0.5-1			
Rv0678	778980	promotor	indel	-9 ins g	1	2	4-8	1	0-2	BDQ	0.12
Rv0678	779005	6fs	indel	16 del g	1	1	2-4	2	2-4	BDQ	0.12
Rv0678	779015	9fs	indel	26 del ag	1	2	4-8	NA	NA	BDQ	0.12
Rv0678	779029	14fs	indel	40 del c	1	2	4-8	2	2-4	BDQ	0.12
Rv0678	779038	17fs	indel	49 del a	1	2	4-8	2	2-4	CFZ	0.25
Rv0678	779051	21fs	indel	62 del a	1	2	4-8	1	0-2	BDQ	0.12
Rv0678	779060	G24V	SNP	T	3	2	4-8	2	2-4	CFZ	0.25
Rv0678	779062	G25S	SNP	A	1	2	4-8	NA	NA	BDQ	0.25
Rv0678	779078	30fs	indel	89 del g	1	2	4-8	2	2-4	CFZ	0.25

Rv0678	779101	R38*	SNP	T	4	2	4-8	2	2-4	CFZ	0.25
Rv0678	779102	R38P	SNP	C	1	1	2-4	2	2-4	CFZ	0.25
Rv0678	779108	L40S	SNP	C	1	1	2-4	2	2-4	CFZ	0.25
Rv0678	779109	40fs	indel	120 del g	1	1	2-4	2	2-4	CFZ	0.25
Rv0678	779117	L43P	SNP	C	5	1-2	2-8	2	2-4	BDQ/CFZ	0.12 / 0.25
Rv0678	779117	L43R	SNP	G	1	1	2-4	2	2-4	BDQ	0.12
Rv0678	779120	L44P	SNP	C	4	1	2-4	2	2-4	CFZ	0.25
Rv0678	779137	R50W	SNP	T	12	1	2-4	2	2-4	CFZ	0.25
Rv0678	779141	Q51R	SNP	G	2	2	4-8	4	4-8	CFZ	0.25
Rv0678	779149	E54*	SNP	T	1	2	4-8	2	2-4	CFZ	0.25
Rv0678	779152	E55*	SNP	T	1	1	2-4	2	2-4	CFZ	0.25
Rv0678	779159	A57E	SNP	A	1	2	4-8	2	2-4	CFZ	0.25
Rv0678	779174	A62D	SNP	A	1	2	4-8	2	2-4	CFZ	0.25
Rv0678	779176	S63G	SNP	G	5	2	4-8	2	2-4	CFZ	0.25
Rv0678	779181	64fs	indel	192 ins g	20	1-4	2-16	2-4	2-8	BDQ/CFZ	0.12 / 0.25
Rv0678	779182	65fs	indel	193 del g	12	1-2	2-8	2-4	2-8	BDQ/CFZ	0.12 & 0.25 / 0.25
Rv0678	779186	G66E	SNP	A	1	2	4-8	1	0-2	BDQ	0.12
Rv0678	779191	S68G	SNP	G	9	1-2	2-8	2	2-4	BDQ/CFZ	0.12 / 0.25
Rv0678	779203	R72W	SNP	T	2	1-2	2-8	2	2-4	CFZ	0.25
Rv0678	779204	R72L	SNP	T	2	2-4	4-16	4-6	4-12	BDQ	0.12
Rv0678	779210	L74P	SNP	C	8	1-2	2-8	2-4	2-8	BDQ/CFZ	0.12 / 0.25
Rv0678	779217	76fs	indel	228 ins t	1	2	4-8	NA	NA	BDQ	0.12
Rv0678	779224	F79V	SNP	G	1	1	2-4	2	2-4	CFZ	0.25
Rv0678	779225	F79C	SNP	G	2	1	2-4	2	2-4	BDQ	0.12
Rv0678	779240	A84V	SNP	T	2	1-2	2-8	1-2	0-4	BDQ	0.12
Rv0678	779252	88fs	indel	263 ins t	1	2	4-8	1	0-2	BDQ	0.12
Rv0678	779263	92fs	indel	274 ins a	3	1	2-4	2	2-4	CFZ	0.25
Rv0678	779270	94fs	indel	281 del g	1	2	4-8	2	2-4	CFZ	0.25
Rv0678	779280	97fs	indel	291 ins a	2	2	4-8	2	2-4	CFZ	0.25
Rv0678	779281	98fs	indel	292 del g	1	2	4-8	2	2-4	BDQ	0.12
Rv0678	779285	A99V	SNP	T	16	1-2	2-8	1-2	0-4	BDQ/CFZ	0.12 / 0.25
Rv0678	779293	A102T	SNP	A	4	1-4	2-16	2-4	2-8	BDQ	0.12

Rv0678	779296	G103R	SNP	C	1	1	2-4	2	2-4	BDQ	0.12
Rv0678	779330	L114P	SNP	C	1	2	4-8	2	2-4	CFZ	0.25
Rv0678	779332	Q115*	SNP	T	1	2	4-8	2	2-4	BDQ	0.12
Rv0678	779342	A118D	SNP	C	1	1	2-4	2	2-4	CFZ	0.25
Rv0678	779349	120fs	indel	360 del g	7	1-4	2-16	2-4	2-8	BDQ/CFZ	0.12 / 0.25
Rv0678	779351	G121E	SNP	A	3	2	4-8	2-4	2-8	CFZ	0.25
Rv0678	779354	L122P	SNP	C	1	2	4-8	2	2-4	CFZ	0.25
Rv0678	779363	125fs	indel	374 del t	1	2	4-8	2	2-4	BDQ	0.12
Rv0678	779363	L125P	SNP	C	1	2	4-8	2	2-4	CFZ	0.25
Rv0678	779384	R132P	SNP	C	1	2	4-8	2	2-4	CFZ	0.25
Rv0678	779392	R135W	SNP	T	1	2	4-8	2	2-4	BDQ	0.12
Rv0678	779396	L136P	SNP	C	2	2	4-8	2-4	2-8	BDQ	0.12
Rv0678	779406	139fs	indel	417 ins CGGG ATCTG TTGGC ATATA T	2	1-2	2-8	2	2-4	BDQ	0.12
Rv0678	779414	142fs	indel	425 del t	1	2	4-8	2	2-4	CFZ	0.25
Rv0678	779415	142fs	indel	426 ins TTGGC ATA	1	2	4-8	NA	NA	BDQ	0.12
Rv0678	779424	Y145*	SNP	G	6	1-2	2-8	2	2-4	BDQ/CFZ	0.12 & 0.25 / 0.25
Rv0678	779440	S151P	SNP	C	1	2	4-8	2	2-4	BDQ	0.12
Rv0678	779450	L154P	SNP	C	2	2	4-8	2	2-4	CFZ	0.25
Rv0678	779454	155fs	indel	465 ins c	2	1-2	2-8	2	2-4	CFZ	0.25
Rv0678	779455	R156*	SNP	T	3	2	4-8	2	2-4	BDQ/CFZ	0.12 & 0.25 / 0.25
Rv0678	-	-	inver sion	-	3	2	4-8	2	2-4	BDQ	0.12
Rv1305	1461127	D28G	SNP	G	1	≥10	20	NA	NA	BDQ	0.25
Rv1305	1461127	D28V	SNP	T	3	≥8	16	NA	NA	BDQ	0.25
Rv1305	1461227	E61D	SNP	C	5	2-≥10	4-40	1	0-2	BDQ	0.25
Rv1305	1461227	E61D	SNP	T	2	≥10	20	1-2	0-4	BDQ	0.12 & 0.25
Rv1305	1461231	A63P	SNP	C	7	≥10	20	1-2	0-4	BDQ	0.25

Alt- alternative, BDQ- bedaquiline, CFZ- clofazimine, del- deletion, fs- frameshift, ins- insertion, MIC- minimum inhibitory concentration, NA- not available, SNP- single nucleotide polymorphism, *-stop codon

109 **Phenotypic and genotypic characterization of BDQ resistant mutants**

110 In total, we randomly selected 270 single colonies from BDQ- and CFZ-supplemented agar plates that had
111 been inoculated with samples from passages 1 (4 days), 3 (12 days), and 5 (20 days), of two independent
112 experiments (Figure S1). In total, 203/270 mutants were successfully pheno- and genotyped (67 were
113 excluded due to DNA library preparation issues or inadequate growth). The single colonies were sub-
114 cultured and the MICs for BDQ and CFZ were determined for each clone using broth microtiter dilution.
115 All clones exhibited an elevated MIC at least 2-fold higher than the WT drug susceptible ancestor and we
116 identified in total 61 unique variants with one mutation in the *Rv0678* gene per clone. We additionally
117 selected five unique mutations in the *atpE* (*Rv1305*) gene (Table 1).

118 Clones which harbored a mutation in *Rv0678* had a 2-4-fold MIC increase for BDQ, whereas mutations in
119 *atpE*, exhibited 4 to over 10 times the MIC of the WT strain (Table 1). For 12 randomly selected clones
120 with seven different variants we verified the results obtained in broth microtiter dilution plates using a
121 Mycobacterium growth indicator tube (MGIT) assay. These experiments showed comparable MICs of the
122 mutant clones with those obtained from the previous method (Table S2).

123 The 61 unique mutations in *Rv0678* affected 54 different codons (Table 1), out of which 21 (34%) were
124 frameshift (fs) mutations, 34 (56%) had one non-synonymous SNP, and 6 (10%) a premature inserted stop
125 codon. The mutations were scattered over the entire sequence of *Rv0678* and no dominant cluster could
126 be identified (Figure 1A). However, the four mutations which were most frequently observed were: 192
127 ins g (64fs), A99V, 193 del g (65fs), and R50W; these mutations were detected in 19, 17, 13, and 12 single
128 selected clones, respectively (Table 1). Notably, nucleotides 192-198 were also identified as a hotspot for
129 frameshifting variation in a prior review, potentially due to the homopolymeric nature of this region[41].

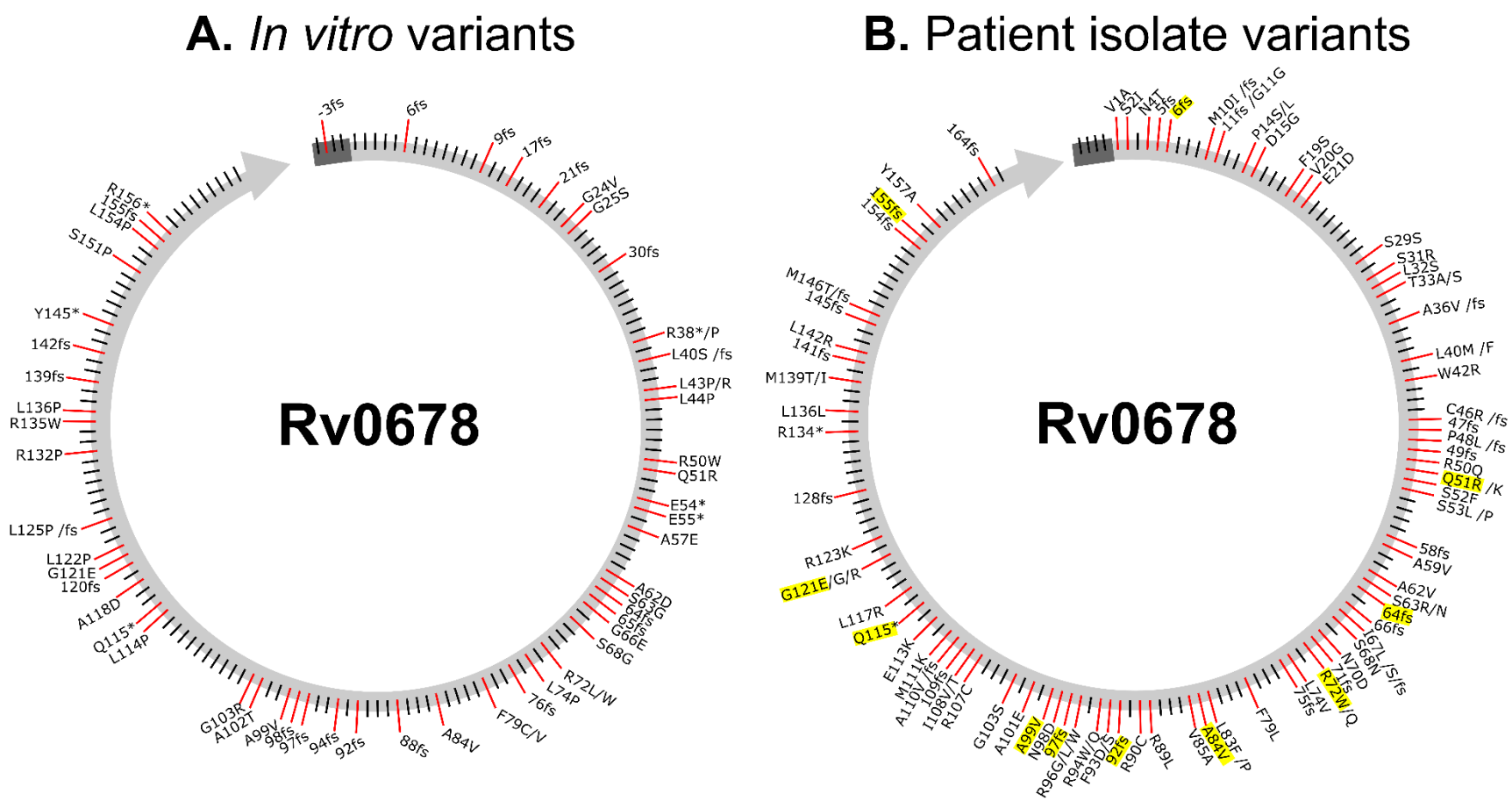


Figure 1: BDQ resistance associated variants mapped in *Rv0678* gene. Resistance associated mutations detected within patient dataset and *in vitro* selection were designated along Rv0678 gene at corresponding coding positions by non-synonymous single nucleotide polymorphisms, and by a protein disrupting mutation such as a stop codon insertion (*) or frameshift (fs) mutations. Data sets divided by (A) mutations which were selected in *in vitro* experiments with elevated minimum inhibitory concentrations verified to bedaquiline, and (B) mutations which were detected in patient isolates either collected by CRyPTIC or described in published literature. Overlapping mutations in observed in both patients and *in vitro* are highlighted in yellow.

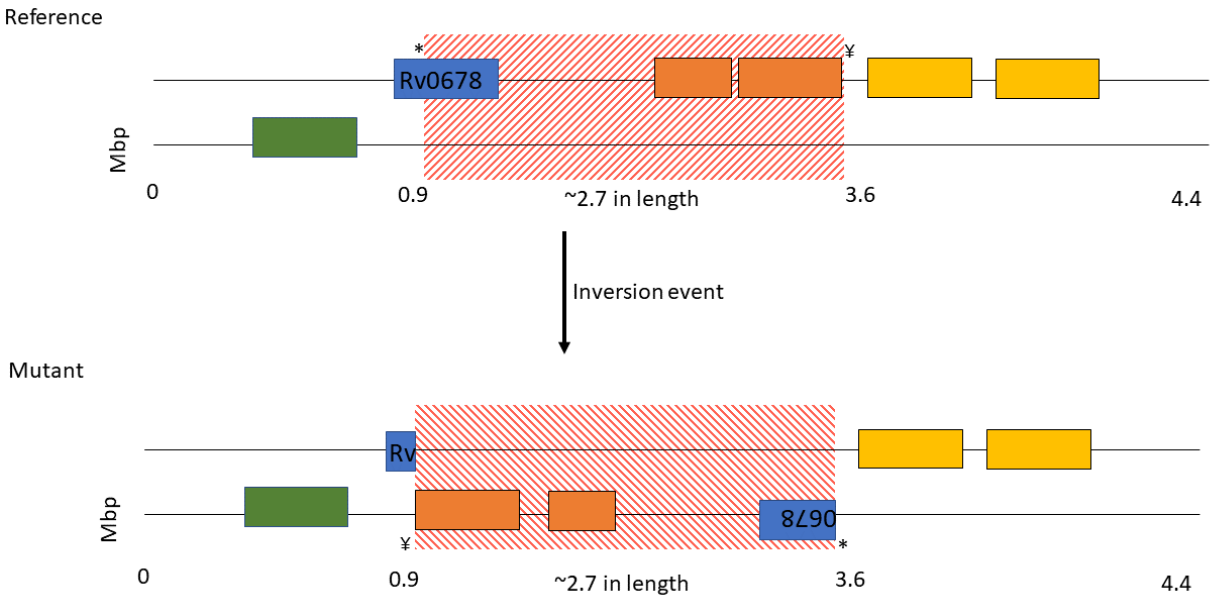
137 As mentioned before, no mutant clone had more than one resistance mediating variant in *Rv0678* or *atpE*,
138 however, 38 mutants harbored a second mutation in one of 14 genes that have not been previously
139 proposed to be associated with resistance against BDQ and/or CFZ (Table S3). The most common mutation
140 found in addition to *Rv0678* mutations was R119H in *Rv1890c* – this gene codes for a conserved
141 hypothetical protein [42], and the mutation was selected 11 times along with six different resistance
142 mutations. The second most common co-selected mutation was *130R in *Rv1871c* (a conserved
143 hypothetical membrane protein [43]) that was selected seven times along with four different resistance
144 mutations. Other secondary mutations occurred in genes which were involved in cell wall synthesis,
145 information pathways, metabolism/respiration, protein regulation, or lipid metabolism (Table S3).

146 In addition to single colony sequencing, we also employed a more unbiased approach based on population
147 sequencing of all the mutant colonies on a given plate, in order to elucidate all possible variants. On
148 average the genome wide coverage for a total of 81 samples was 308±50. We then reported mutations if
149 the position was covered by at least one read in both forward and reverse orientations and there were
150 two reads with a phred score over 20. This identified 45 additional mutations in *Rv0678* and one in *atpE*
151 (A63T), as well as four mutations in *pepQ* (*Rv2535c*; F97V, G96G, V92G, A87G) (Table S1). As in the single
152 colony analysis, we found the following mutations dominating in different independent evolutionary
153 experiments and selected across different drug concentrations: *Rv0678* 192 ins g (64fs), A99V, R50W, and
154 L43P.

155 **Resistance caused by huge inversion interruption in Rv0678**

156 Three mutants with elevated BDQ (2 mg/L) and CFZ (2 mg/L) MICs but lacking mutations in the BDQ/CFZ
157 resistance associated genes were further subjected to long read sequencing with the PacBio Sequel II
158 system. A *de novo* assembly employing the PacBio SMRT® Link software resulted in one closed Mtbc
159 genome, and two assemblies with 1 and 4 four contiguous sequences (contigs), respectively. All
160 assemblies covered more than 99.9% of the H37Rv reference genome (NC_000962.3). All three assemblies
161 showed a large-scale sequence re-arrangement with the borders at position 779,073 (Rv0678 coding
162 sequence) and 3,552,584 (intergenic region), flanked by a transposase open reading frame (IS*6110*). A
163 2.5Mb fragment was inverted and thus split the *Rv0678* gene into halves (Figure 2), which was not
164 detected by a classical reference mapping approach and revealed a further interesting resistance
165 mechanism. We scanned all BDQ resistant samples in the CRyPTIC strain collection for similar events using

166 local assembly with ARIBA [44]. In one isolate we assembled one contig which clearly represented a similar
167 inversion disrupting *Rv0678*, but since the depth of support was just 1 read (mean, across the contig) and
168 this was Illumina (short read) data, we considered this only circumstantial evidence.



169

170 **Figure 2: Large scale gene rearrangement in Rv0678 gene.** *De novo* assembly using PacBio SMRT® Link
171 software indicated a 2.7 Mb inversion (red) impacting *Rv0678* (blue), at positions 779,073 coding
172 sequence (amino acid 34 in Rv0678) and 3,552,584 intergenic region; flanked by transposase related
173 genes (orange and yellow).

174 **Mutations associated with bedaquiline and clofazimine resistance in clinical isolates**

175 Next, we aimed to comprehensively describe the BDQ and CFZ resistant variants occurring in patients, to
176 develop a catalogue of known resistance variants. We searched the literature for patient reports
177 describing BDQ resistance during MDR-TB therapy and analyzed pheno- and genotypes of 14,151 clinical
178 Mtbc strains collected by the CRyPTIC consortium (Table S4). In this patient dataset, we identified 98
179 mutations throughout *Rv0678*, 71 of which were not previously described in patients (Table 2). In addition,
180 seven variants in *Rv0678* harbored mutations found in different lineages (Table S4), pointing to
181 convergent evolution at these sites. Eleven mutations from the patient-derived set were also shared with
182 our *in vitro* derived datasets (selected variants and population sequencing) (Figure 1B & 3A).

183 **Table 2: *Rv0678* mutations detected in patients (published literature and CRyPTIC collection).** Reference
184 catalog of *Rv0678* mutations which have been described in previously published patient isolates or
185 collected by the Comprehensive Resistance Prediction for Tuberculosis an International Consortium
186 (CRyPTIC) partners. Additionally, alternative (Alt.) allele information included, describes the variants at
187 the base pair position. Variants included in list if presented a borderline or resistant phenotype to at least
188 bedaquiline (BDQ)MIC determined by different drug susceptibility testing (DST) techniques depending on
189 the study, abbreviated as: [A] BD-MGIT™, [B] UKMYC5/UKMYC6 plates, [C] Alamar blue/resazurin assay,
190 [D] 7H10 plates, [E] 7H11 plates

191 Phenotype interpretation was based off of: [A] 1mg/L BDQ/CFZ[80]; [B] 0.12-0.25mg/L BDQ/CFZ[75], [C]
192 defined by study authors, [D] 0.5mg/L BDQ[81]/CFZ[82], [E] 0.25 mg/L BDQ[80] -CFZ determined by study authors

Gene	Substitu -tion	Alt Allele	Type	BDQ MIC (mg/L)	CFZ MIC (mg/L)	Phenotype interpretati on BDQ- R, B, S	Phenotype interpretati on CFZ- R, B, S	DST Method	Author
Rv0678	V1A	c	SNP	6.4/0.5	>4/NA	R/B	R	[A]	[24,25]
Rv0678	S2I	t	SNP	0.73	4	R	R	[C]	[28]
Rv0678	N4T	acc	SNP	0.06-0.12	0.12	S/B	B	[B]	CRyPTIC
Rv0678	5fs	15 del c	indel	0.5	0.25	R	R	[B]	CRyPTIC
Rv0678	6fs	18 del gg	indel	0.5	NA	R	NA	[E]	[68]
Rv0678	6fs	16 del gg	indel	0.5	NA	R	NA		[83]
Rv0678	10fs	28 del t	indel	0.12	≤0.06	B	S	[B]	CRyPTIC
Rv0678	10fs	29 del t	indel	0.25	0.12	R	B	[B]	CRyPTIC
Rv0678	M10I	atc	SNP	0.12	NA	B	NA	[B]	CRyPTIC
Rv0678	11fs	32 del g	indel	0.25	NA	R	NA	[E]	[48]
Rv0678	G11G	t	SNP	0.5-2.0	NA	R	NA	[C]	[84]
Rv0678	P14S	tcc	SNP	0.25	0.5	R	R	[B]	CRyPTIC
Rv0678	P14L	ctc	SNP	0.12	≤0.06	B	S	[B]	CRyPTIC
Rv0678	D15G	NA	SNP	0.25	NA	R	NA	[B]	CRyPTIC
Rv0678	F19S	tcc	SNP	0.5	NA	R	NA	[B]	CRyPTIC
Rv0678	V20G	NA	SNP	0.25/2/0.25/0.25	NA/2/NA/4	R/B	R	[E]/[A]/[A]/[E]	[69,69, 52, 29]
Rv0678	V20A	gcc	SNP	0.12	0.12	B	B	[B]	CRyPTIC
Rv0678	E21D	c	SNP	0.5-1.0	NA	R	NA	[C]	[84]
Rv0678	S29S	tct	SNP	0.12	0.12	B	B	[B]	CRyPTIC
Rv0678	S31R	c	SNP	0.5	NA	R	NA	[C]	[84]

Rv0678	L32S	tcg	SNP	0.12	0.12	B	B	[B]	CRyPTIC
Rv0678	T33A	g	SNP	0.5	2	R	R	[C]	[23]
Rv0678	T33S	g	SNP	2	NA	R	NA	[C]	[84]
Rv0678	A36V	t	SNP	0.5	4	R	R	[C]	[23]
Rv0678	36fs	107 ins g	indel	0.5	0.25	R	R	[B]	CRyPTIC
Rv0678	L40M	atg	SNP	0.5	NA	R	NA	[B]	CRyPTIC
Rv0678	L40F	ttc	SNP	0.5	NA	R	R	[B]	CRyPTIC
Rv0678	W42R	cgg	SNP	0.25/0.12	0.25/2	R/B	R	[B]/[D]	CRyPTIC, [14]
Rv0678	46fs	136 ins g	indel	1	NA	R	NA	[E]	[52]
Rv0678	46fs	137 ins g	indel	0.25	1	R	R	[B]	CRyPTIC
Rv0678	46fs	138 ins g	indel	0.25/0.25/2/>1/1	0.25/NA/4/NA/NA	R	R	[B]/[E]/[A]/[E]/[E]	CRyPTIC, [69],[69],[12],[52]
Rv0678	46fs	138 ins ga	indel	>1.0	NA	R	NA	[E]	[15]
Rv0678	C46R	cgt	SNP	0.25	NA	R	NA	[B]	CRyPTIC
Rv0678	47fs	139 ins g	indel	0.25	NA	R	NA	[E]	[48]
Rv0678	47fs	140 ins g	indel	0.25	NA	R	NA	[E]	[68]
Rv0678	47fs	141 ins c	indel	0.06-0.25/0.5/4/0.5	0.06-1/NA/4/NA	S-R	S-R	[B]/ [E]/ [A]/[E]	CRyPTIC, [69],[69],[29]
Rv0678	48fs	144 ins c	indel	0.25	NA	R	NA	[E]	[48]
Rv0678	P48L	NA	SNP	0.25-0.5	NA	R	B	[B]	CRyPTIC
Rv0678	E49fs	NA	indel	0.03-0.25	0.25-2	R	R	[E]	[29]
Rv0678	R50Q	NA	SNP	0.25	NA	R	NA	[B]	CRyPTIC
Rv0678	Q51K	NA	SNP	0.25	1	R	R	[B]	CRyPTIC
Rv0678	Q51R	NA	SNP	0.3	1	R	R	[B]	CRyPTIC
Rv0678	S52F	NA	SNP	0.48	NA	R	NA	[D]	[14]
Rv0678	S53L	NA	SNP	0.25-0.4	2	R	R	[C]	[28]
Rv0678	S53P	NA	SNP	0.5	2-4	R	R	[C]	[28]
Rv0678	58fs	172 ins IS6110	indel	0.5	1	R	R	[C]	[23]
Rv0678	A59V	NA	SNP	0.48	NA	R	NA	[D]	[14]
Rv0678	A62V	NA	SNP	0.48	NA	R	NA	[D]	[14]
Rv0678	A62T	acc	SNP	0.12	0.5	B	R	[B]	CRyPTIC
Rv0678	S63R	a	SNP	0.5	1.25	R	R	[C]	[22]
Rv0678	S63N	a	SNP	0.25	NA	R	NA	[C]	[85]

Rv0678	64fs	192 ins g	indel	0.03-0.25/0.5/4	0.12/NA/4	S-R	R	[B]/[E]/[A]	CRyPTIC,[69],[69]
Rv0678	64fs	192 ins c	indel	>1.0	NA	R	NA	[E]	[12]
Rv0678	66fs	198 ins g	indel	0.24-1	NA	R	NA	[D]	[14]
Rv0678	I67L	ctc	SNP	0.5	0.12	R	B	[B]	CRyPTIC
Rv0678	I67S	agc	SNP	0.5	0.25	R	R	[B]	CRyPTIC
Rv0678	S68N	aac	SNP	0.5	2	R	R	[B]	CRyPTIC
Rv0678	N70D	gat	SNP	0.25	NA	R	NA	[B]	CRyPTIC
Rv0678	71fs	212 del c	indel	0.5	1	R	R	[C]	[23]
Rv0678	R72W	NA	SNP	0.06-0.25	≤0.06->4	S-R	S-R	[B]	CRyPTIC
Rv0678	R72Q	a	SNP	0.757	2.77	R	R	[C]	[86]
Rv0678	L74V	gtg	SNP	0.25	0.12	R	B	[B]	CRyPTIC
Rv0678	75fs	224 ins a	indel	0.24	NA	R	NA	[D]	[14]
Rv0678	F79L	tta	SNP	0.12	0.12	B	B	[B]	CRyPTIC
Rv0678	L83F	NA	SNP	0.5	NA	R	NA	[B]	CRyPTIC
Rv0678	L83P	c	SNP	0.25	1	R	R	[B]	CRyPTIC
Rv0678	A84V	t	SNP	0.06	≤0.06	S/R	S/R	[B]	CRyPTIC
Rv0678	V85A	gcc	SNP	0.25/0.12	0.25	R/B	R	[B]/[E]	CRyPTIC,[48]
Rv0678	R89L	t	SNP	0.12	0.12	B	B	[B]	CRyPTIC
Rv0678	R90C	NA	SNP	0.3	0.16			[B]	CRyPTIC
Rv0678	92fs	274 ins a	indel	0.25/1	0.5/NA	R	NA	[B]/[D]	CRyPTIC,[14]
Rv0678	F93D	NA	SNP	0.25	NA	R	NA	[E]	[52]
Rv0678	F93S	NA	SNP	0.25	1	R	R	[E]	[29]
Rv0678	R94W	NA	SNP	2	4	R	R	[A]	[17]
Rv0678	R96W	NA	SNP	0.25	NA	R	NA		[83]
Rv0678	R96G	ggg	SNP	0.25	0.25	R	R	[B]	CRyPTIC
Rv0678	R96L	ctg	SNP	0.25	NA	R	NA	[B]	CRyPTIC
Rv0678	97fs	291 del c	indel	0.25	NA	R	NA	[E]	[48]
Rv0678	N98D	gac	SNP	0.25	≤0.06-0.12/NA/0.75-1/NA	R/S	S-R	[B]/[C]/[A]/[C]	CRyPTIC,[84],[30],[85]
Rv0678	A99V	t	SNP	0.25	NA	R	NA	[E]	[48]
Rv0678	A101E	gag	SNP	0.25	0.12	R	B	[B]	CRyPTIC
Rv0678	G103S	agc	SNP	0.03-0.12	0.5	B	R	[B]	CRyPTIC
Rv0678	R107C	NA	SNP	0.25	NA	R	NA	[B]	CRyPTIC

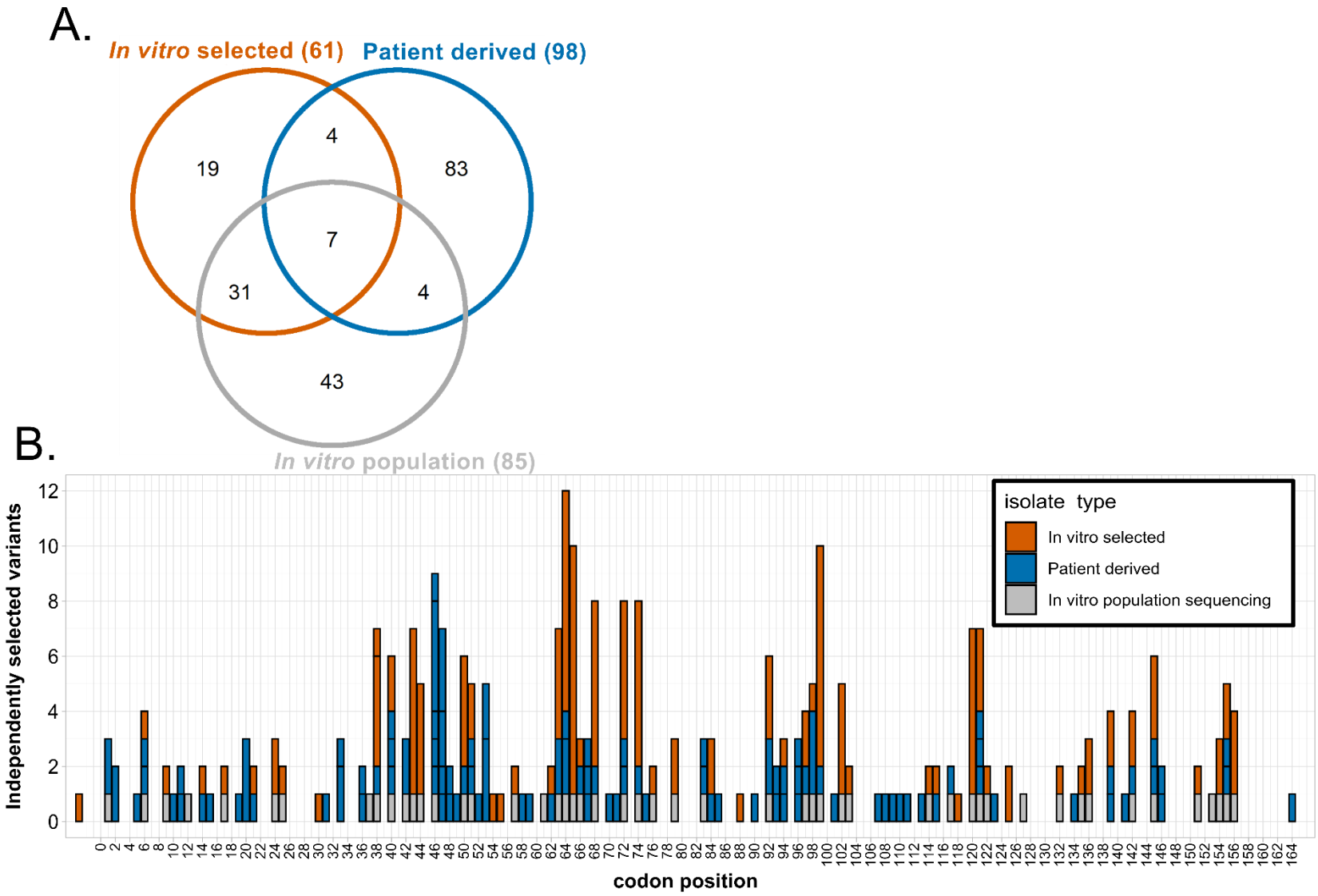
Rv0678	I108T	NA	SNP	0.25	0.25	R	B	[B]	CRyPTIC
Rv0678	I108V	gtc	SNP	0.12	0.25	B	R	[B]	CRyPTIC
Rv0678	109fs	325 ins g	indel	1	NA			[C]	CRyPTIC
Rv0678	A110V		SNP	0.5	0.5	B	R	[A]	[30]
Rv0678	110fs	329 del t	indel	0.12	≤0.06	B	S	[B]	CRyPTIC
Rv0678	M111K	NA	SNP	0.25	NA	R	NA		[83]
Rv0678	E113K	aaa	SNP	0.25	0.25	R	R	[B]	CRyPTIC
Rv0678	Q115*	tag	SNP	0.12	0.25	B	R	[B]	CRyPTIC
Rv0678	L117R	cgg	SNP	0.015-0.25/1.5	0.06-0.12/4.16	S-R	S-R	[B][C]	CRyPTIC, [28]
Rv0678	G121E	a	SNP	0.25	NA	R	R	[E]	[48]
Rv0678	G121R	agg	SNP	0.5/0.75	NA/2	R/B	R	[B]	CRyPTIC
Rv0678	G121G	gga	SNP	0.25	0.12	R	B	[B]	CRyPTIC
Rv0678	R123K	a	SNP	1	NA	R	NA	[C]	[84]
Rv0678	128fs	382 del t	indel	≤0.015	≤0.06	B	S	[B]	CRyPTIC
Rv0678	R134*	t	SNP	0.5	1.25	R	R	[C]	[22]
Rv0678	L136L	cta	SNP	0.12	NA	B	NA	[B]	CRyPTIC
Rv0678	M139T	NA	SNP	0.25	NA	R	NA	[E]	[68]
Rv0678	M139I	ata	SNP	0.25	0.25	R	R	[B]	CRyPTIC
Rv0678	141fs	421 ins g	indel	0.25	NA	R	NA	[E]	[48]
Rv0678	L142R	NA	SNP	0.25-1.0	NA	R	NA	[E]	[48]
Rv0678	145fs	435 del t	indel	0.25	NA	R	NA	[E]	[48]
Rv0678	146fs	438 ins t	indel	0.25	0.5			[E]	[52], [29]
Rv0678	M146T	acg	SNP	0.25-0.8	1.2	S-R	S-R	[B]/[C]/[A]	CRyPTIC, [28]
Rv0678	154fs	461 del t	indel	0.12	0.5	B	R	[B]	CRyPTIC
Rv0678	155fs	464 ins gc	indel	0.5	1			[B]	CRyPTIC
Rv0678	155fs	465 ins c	indel	0.12/1-2	0.25/2	B/R	R	[B]/[C]	CRyPTIC/ this study
Rv0678	Y157A		SNP	0.125	2	B	R	[C]	[87]
Rv0678	164fs	492 ins ga	indel	0.25	NA			[B]	CRyPTIC

Alt- alternative, BDQ- bedaquiline, CFZ- clofazimine, DST- drug susceptibility testing, del- deletion, fs- frameshift, ins- insertion, MIC- minimum inhibitory concentration, NA- not available, SNP- single nucleotide polymorphism, **-stop codon

196

197 Wild-type isolates presented an MIC95 of 0.12 mg/L BDQ/CFZ in the UKMYC5/6 plates[45]. Therefore, the
198 patient isolates in the study were catalogued as >0.12 mg/L BDQ/CFZ resistant, <0.12 susceptible, and
199 =0.12 borderline. The CRyPTIC dataset contained 178 patient isolates with a single *Rv0678* variant, 52 of
200 which were classified as BDQ resistant (27 cross-resistant with CFZ), 32 with borderline BDQ resistance,
201 and 94 BDQ susceptible (Table S4). To better explain this high number of BDQ susceptible strains with
202 *Rv0678* mutations we looked for additional mutations in the efflux pump genes *mmpL5* and *mmpS5*, as it
203 has been postulated that variants in these genes may reduce or abrogate acquired resistance [46]. However,
204 we found that 64/94 (68%) of susceptible isolates harboring *Rv0678* mutations had WT *mmpL5/mmpS5*
205 genes (Table S4). Likely functional/inactivating mutations in *mmpL5* (19/19 isolates with frameshifts and
206 3/5 with missense) were associated with susceptibility while synonymous mutations appeared
207 uncorrelated with susceptibility (7/18 isolates susceptible).

208 Plotting all 147 resistance and borderline resistance-associated *Rv0678* mutations on the gene sequence
209 (Figure 1), reveals that they occur throughout the entire length of the gene with no single hot spot region.
210 However, there are several regions where mutations are more likely to occur, i.e. codons 46-53 and 62-
211 68 (Figure 3B).


A.
In vitro selected (61)
Patient derived (98)
In vitro population (85)
19
4
83
7
31
4
43
B.
Independently selected variants
codon position
isolate type
In vitro selected
Patient derived
In vitro population sequencing

212

213 **Figure 3: Mutations and coding position overlap throughout *Rv0678, in vitro* and patient datasets.** Overlapping mutations which conferred BDQ
214 resistance were compared between three datasets: *in vitro* selected mutants (Table 1), *in vitro* population sequencing (Table S1), and patient derived
215 variants (Table 2). (A) Venn diagram describes number overlapping mutations in the three datasets, all mutations must have the same SNP or a
216 deletion/insertion at same coding position to overlap. (B) Data represents number of mutants selected at given coding position for *in vitro* selected
217 (orange) and patient derived variants (blue), indicated with gray box if detected in *in vitro* population sequencing.

218 Next, we searched the CRyPTIC dataset for resistance conferring mutations (BDQ MIC >0.12 mg/L) in *atpE*,
219 *pepQ* (*Rv2535c*) and *Rv1979c*. After filtering out phylogenetic SNPs and strains with preexisting *Rv0678*
220 mutations, we found 35 isolates with an MIC >0.12 mg/L for BDQ harboring *Rv1979c* mutations and only
221 3 isolates with mutations in *atpE* [47]. There were no BDQ resistant isolates with *pepQ* mutations, but 19
222 strains had borderline BDQ resistance with an MIC of 0.12 mg/L. In total, we defined four *Rv1979c*-
223 mediated resistant mutations, and 7 possible borderline-resistance conferring mutations in *pepQ*
224 (*Rv2535c*).

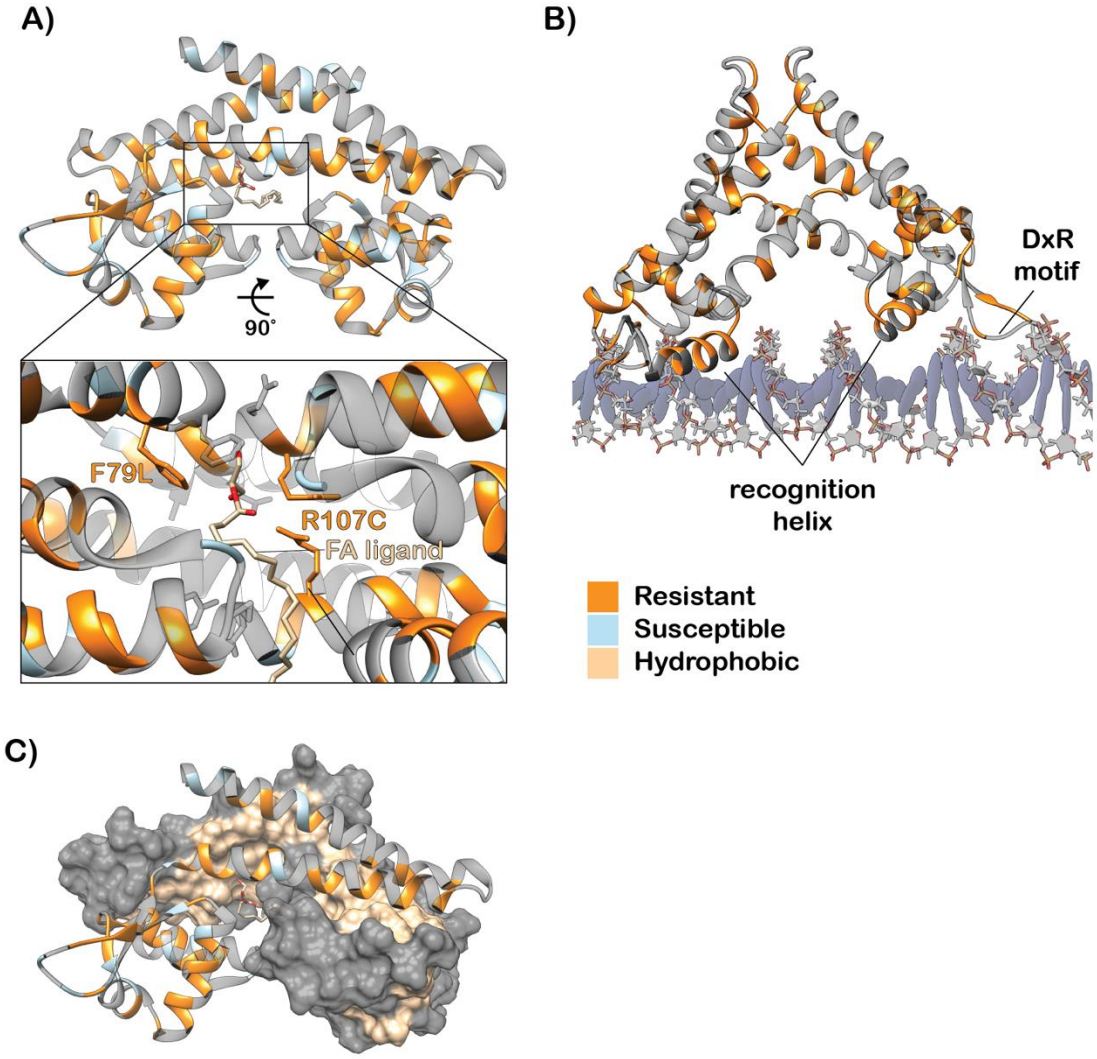
225 Finally, after an in-depth literature search including *in vitro*, *in vivo*, and patient derived strains, we
226 complied a comprehensive catalogue of variants in BDQ resistance associated genes (Table S5). Due to
227 insufficient data for BDQ and CFZ critical concentration determination for many phenotypic assays, and
228 the high overlap of MIC distributions between WT and mutant clones, resistance interpretation of
229 mutations was based on the ascribed phenotype in each publication or by WHO critical concentration
230 when available.

231 **Structural variation in Rv0678**

232 For better mechanistic understanding of how *Rv0678* mutations result in BDQ/CFZ resistance, we mapped
233 missense mutations from our catalogue to the previously determined Rv0678 crystal structure (107
234 mutations mapped to resolved structure positions PDB ID 4NB5, Table S5) [20]. As mutations in different
235 parts of the protein structure are likely to have different effects, we investigated four different
236 mechanisms for conferring resistance: reducing folded protein/protein stability, impairing DNA binding,
237 altering protein dimerization, and reducing affinity for the fatty acid ligand of Rv0678.

238 The numerous introductions of premature stop codons and frameshift mutations reported here and
239 elsewhere suggest loss of functional Rv0678 protein is a common mechanism of resistance[22,41,48].
240 Consistent with this, if we compare the missense *Rv0678* mutations in resistant isolates with those in
241 susceptible isolates, we find the former are more commonly predicted to destabilize the Rv0678 wHTH
242 domain required for DNA binding (destabilization defined as $\Delta\Delta G <= -1$ kcal/mol as measured by mCSM
243 protein stability, 34/38 R, 5/10 S, OR 7.96, Fisher exact p=0.012), suggesting that disruption of protein
244 stability is also a resistance mechanism. These findings are also compatible with the observation that
245 missense mutations from resistant isolates are predicted to have a more deleterious effect by SNAP2
246 score compared with those from susceptible isolates (Wilcoxon p=0.018) [49].

247 To understand how mutations may impact DNA binding, we structurally aligned the wHTH DNA binding
248 domain of Rv0678 to several other Mtbc marR-family transcriptional regulators whose DNA-bound
249 structures have been determined (Figure 4B, S3; average RMSD 0.81 Å)[50]. We found that mutations
250 clustering in regions 62-68 and 88-92 correlate with the recognition α-helix and conserved DxR motif
251 respectively, which directly contact the DNA and have previously been shown to critical for the binding
252 affinity of other marR-family regulators (Figure 4B)[50,51]. However, further experimental verification of
253 these mutations' effects on DNA binding is required.



254

255 We also investigated the possibility that mutations alter the protein-protein interactions of the Rv0678
256 dimer. Mutations associated with resistance are enriched at the dimer interface (27/29 interface
257 mutations R), suggesting they may impact dimer formation and/or stability (Figure 4C). However,

258 computational prediction of interface mutations' effects on non-DNA bound dimer stability revealed no
259 significant difference in destabilization as defined by mCSM PPI ΔΔG <= -1 kcal/mol (16/27 R, 1/2 S).
260 Interpretation of the DNA-bound dimer interface was not feasible due to the significant conformational
261 rearrangement of the dimer alpha-helices upon DNA binding. Finally, all isolates where an amino acid that
262 contacts the fatty acid ligand 2-stearoylglycerol was mutated were resistant, suggesting that disrupting
263 the binding of the ligand confers resistance, possibly through perturbing non-DNA bound homodimer
264 formation (Figure 4A).

265 **Molecular dynamics simulations**

266 Rv0678 must also undergo a conformational rearrangement to bind DNA which is not captured by the
267 static structures in the previous section. Therefore, we employed molecular dynamics simulations to
268 understand how particular mutations in the "hinge" region (connecting the DNA binding and dimerization
269 domains) might disrupt this motion and lead to resistance. To do this, we performed 100ns simulations
270 on the WT Rv0678 homodimer as well as Rv0678 homodimers containing either L40F and A101E
271 mutations (which are associated with resistance but do not clearly interact with the DNA or disrupt protein
272 folding) or the L40V mutation (consistently phenotypically susceptible). These simulations revealed that
273 mutations from resistant isolates (L40F and A101E) resulted in changes in the conformational dynamics
274 and pocket crosstalk of the individual Rv0678 monomers distinct from WT or the L40V mutation (Figure
275 5). In particular, A101E formed a new stable salt-bridge interaction in the "hinge" domain, which could
276 disrupt the conformational flexibility required by this region to transition between the DNA- and ligand-
277 bound homodimer states (Figure S4).

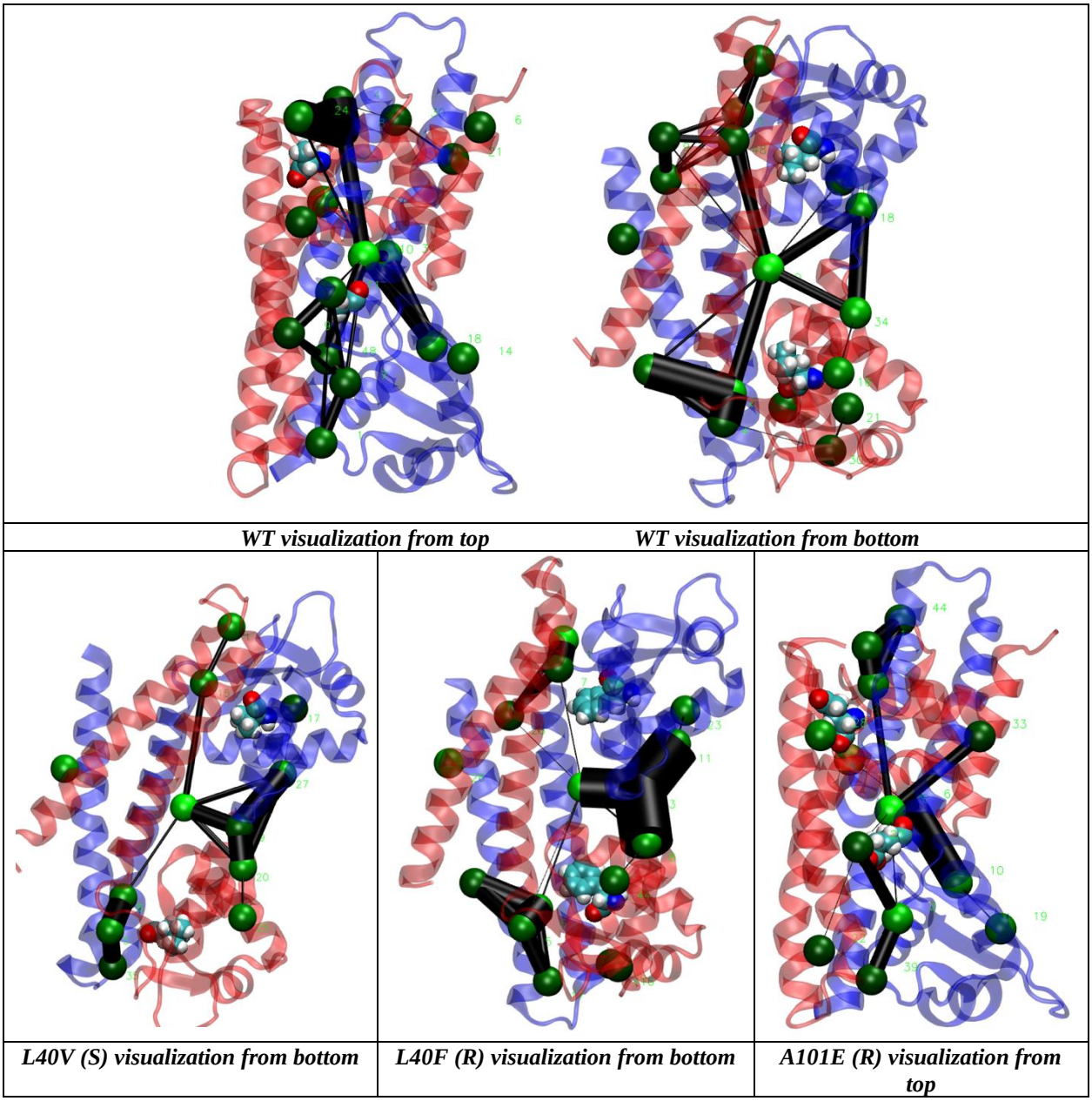


Figure 5: Networks of the most persistent pockets found in the wild-type (WT) and mutated Rv0678. The pocket crosstalk analysis on simulated systems identifies allosteric signaling. All systems shared a large common pocket located at the interface of the two monomers (Figure S5). This large pocket communicates with other smaller pockets at different zones of the homodimer by the network edges (black lines). Comparative analysis showed that the resistance-conferring mutations, L40F and A101E displayed different network edges during the 100ns of simulations with respect to the WT. R- resistant, S-susceptible

285 **Discussion**

286 In this study, we identify new bedaquiline and clofazimine resistance determinants by combining *in vitro*
287 evolution experiments with WGS and MIC data, including the analysis of 14,151 clinical Mtbc strains
288 provided by the CRyPTIC consortium (http://www.crypticproject.org/). The results have allowed us to
289 define a comprehensive set of genomic variants associated with resistance to BDQ (& CFZ), which are
290 central to treating MDR, pre-XDR and XDR-TB [8]. This project also demonstrates the potential of *in vitro*
291 evolution experiments employing selection under sub-lethal drug concentrations to generate clones with
292 a wide spectrum of mutations, including clinically relevant mutations. Collectively, this work helps extend
293 the use of genomics in providing personalized medicine to patients with MDR, pre-XDR, and XDR-TB,
294 through bolstering genotypic catalogues with phenotypically verified variant lists.

295 Although BDQ/CFZ are rolled out worldwide as core drugs in treatment regimens for MDR-TB patients [8],
296 antibiotic susceptibility testing for these compounds is rarely available and patients are often treated
297 empirically. Indeed, recent studies report treatment failure and/or poor outcome of MDR-TB patients as
298 mutations in *Rv0678* emerge that are associated with BDQ/CFZ resistance [12,15,17,26,52]. Additionally,
299 misdiagnosis can promote the selection and subsequent transmission of specific MDR Mtbc strains. For
300 example, it was recently reported in Eswatini and South Africa, that patients infected with an MDR
301 outbreak strain harboring the RMP mutation I491F, which is not detectable by conventional phenotypic
302 or genotypic testing, probably continued to receive RMP although it was ineffective [30]. These strains now
303 represent more than 60% of MDR strains in Eswatini and additionally more than half of these developed
304 enhanced BDQ/CFZ MICs via acquired *Rv0678* mutations [1,29,30,53]. This scenario underlines the urgent need
305 for approaches to rapidly detect resistance, including to BDQ/CFZ, with baseline diagnostics and during
306 therapy (e.g. via WGS or targeted genome sequencing) to avoid treatment failure, further resistance
307 development and ongoing transmission of specific MDR strains.

308 Genome sequencing techniques have been demonstrated to accurately identify genomic variants in all
309 target regions involved in resistance development in a single analysis [3,54]. However, several major
310 challenges exist—including linking genotype and phenotype to distinguish resistance-mediating from
311 benign mutations and interpreting how multiple mutations interact when the resistance phenotype is
312 polygenic, as is the case with BDQ and CFZ[47]. This work catalogues the phenotype of 260 unique genomic
313 variants across six genes. Importantly, not all variation leads to resistance, as 38 of these mutations in

314 *Rv0678* are not associated with resistance to BDQ/CFZ (Table S5), which facilitates the prediction of
315 BDQ/CFZ susceptibility. Furthermore, 30/94 BDQ susceptible isolates with *Rv0678* mutations harbored an
316 additional mutation in the *mmpL5* or *mmpS5* efflux pump genes, which may abrogate the efflux pump
317 mechanism and reconstitute the susceptible phenotype[46]. Here, we provide preliminary evidence that
318 isolates with inactivating (frameshift, nonsense, or missense variants) *mmpL5* mutations can override
319 *Rv0678* resistance-mediating mutations, resulting in hyper-susceptibility to BDQ. Further work to
320 understand the frequency, distribution and origin of these inactivating mutations is necessary to
321 understand their clinical relevance worldwide.

322 While genome sequencing of clinical Mtbc strains has been deployed in several countries, its
323 implementation is constrained not just by technical and data analysis considerations, but also because
324 working directly from clinical samples (e.g. sputum) is difficult [3]. We recently demonstrated that targeted
325 genome sequencing using the commercially available Deeplex®-MycTB has a high sensitivity and
326 specificity for first- and second-line drug resistance prediction, including BDQ/CFZ resistance mutations
327 that have been found in patients [53,55,56]. In combination with our comprehensive set of mutations that
328 confer resistance to BDQ/CFZ, Deeplex®-MycTB and other targeted sequencing approaches allow for rapid
329 detection of BDQ/CFZ resistance from clinical samples in few days using small portable genome
330 sequencers such as the MinION [57].

331 Importantly, *in vitro* laboratory evolution experiments identified 50 resistance-conferring mutations that
332 have not yet appeared in clinical strains, thus allowing us to characterize and identify a significant
333 proportion of resistance-conferring mutations before they have even occurred in a patient. These results
334 highlight the value of *in vitro* evolution experiments as a complementary method for prospective
335 identification of resistance, as compared to traditional surveillance of patient isolates. Furthermore, this
336 method provides a large strain collection of *de novo* resistant clones from a comparable wild-type
337 susceptible ancestor. Finally, by using PacBio sequencing to establish the full genome sequence of mutant
338 clones without any mutations in the relevant targets, we were able to identify a large genomic inversion
339 that disrupts *Rv0678,* which should be considered as a novel BDQ/CFZ resistance mechanism.

340 Another method for prospectively identifying mutations that confer resistance is structure-based
341 modeling, which has recently been combined with machine learning methods to design algorithms for
342 predicting resistance to rifampicin, isoniazid, and pyrazinamide [58–62]. These approaches rely on the fact

343 that resistance-conferring mutations are often located in drug binding pockets or active sites and have
344 distinct biophysical consequences as compared to their susceptible counterparts [63]. This study identifies
345 several resistance mechanisms in *Rv0678* with quantifiable structural effects (protein stability, dimer
346 interactions, SNAP2 scores, and interaction with the DNA), suggesting that a structure-based machine
347 learning approach could also be successful for predicting BDQ/CFZ resistance.

348 As we have shown, any approach to predict BDQ resistance must consider not only mutations present in
349 *Rv0678*, but also potential resistance/susceptibility determinants in other putative resistance-conferring
350 genes e.g. *atpE*, *pepQ*, *Rv1979c,* and *mmpL/S5*. It has been attempted to predict BDQ resistance from *atpE*
351 mutations [64], so one feasible option would be a meta-approach that flags if any individual gene-level
352 algorithms predict resistance/susceptibility. In addition, CRyPTIC and other studies are producing large
353 MIC datasets, which will enable much more nuanced modeling of resistance, which is particularly
354 important for drugs with efflux-mediated resistance that often exhibit small sub-threshold changes in MIC.
355 Although structure-based modeling is helpful in predicting resistance, these structural hypotheses must
356 be confirmed experimentally for final validation.

357 As indicated by the overlap of *in vitro* selected mutations with 11 resistance conferring mutations found
358 in patients, sub-MIC drug concentration can select for clinically relevant resistance variants. This modeling
359 might also better imitate the intra-patient situation, since the complex pathology of TB disease prevents
360 certain drugs (i.e. bactericidal and bacteriostatic drugs), from reaching all bacteria at a therapeutic level
361 (for example structures like pulmonary cavities and granuloma lesions) [33,35,36,66]. BDQ and CFZ in particular
362 do not penetrate the central caseous lesion (or necrotic center) of the granuloma effectively [34]. Although
363 this study provides a large collection of *in vitro* selected Rv0678-mediated resistant clones, the efflux
364 pump modulation by these various mutations was not experimentally investigated and should be explored
365 in the future.

366 In addition to clear-cut resistance-associated mutations, additional "secondary" mutations were co-
367 selected with resistance-associated variants *in vitro*. These secondary mutations affected several genes,
368 mainly associated with cellular processes such as cell wall synthesis, metabolic processes, and protein
369 regulation; with mutations also in genes of unknown function (Table S3). It is possible that these mutations
370 bring about an additional phenotypic affect, or they may be compensatory mutations [5,67]. However, until

371 further experiments are conducted, such as competitive fitness assays (or even transcriptomic analysis),
372 it cannot be ruled out that these are merely hitchhiking mutations or culture-selected variants.

373 In this study, we focused on structural mechanisms of *Rv0678* mediated resistance, as these mutations
374 are the most clinically relevant and correlated with BDQ treatment failures. Other genes which have been
375 linked to BDQ or CFZ resistance, such as *atpE*, *pepQ* and *Rv1979c*, but their clinical relevance has yet to
376 be established [10,29,68,69]. In the CRyPTIC dataset we identified four *Rv1979c* variants which presented a
377 resistant phenotype to BDQ with variable CFZ cross resistance, and seven mutations in *pepQ* garnering
378 borderline BDQ resistance and variable CFZ resistance. Finally, mutations in *atpE* (coding for the target of
379 BDQ) were rare among the *in vitro* selection experiments and the CRyPTIC dataset, suggesting that *atpE*
380 mutations might have a high fitness cost.

381 In conclusion, our work advances our understanding of how resistance can arise to BDQ and CFZ. This
382 information provides an updated variant catalogue, comprising of mutations and structural variations
383 associated with BDQ/CFZ resistance as well as benign variants and mutations implicated with hyper-
384 susceptibility. Employing this information in DNA sequencing-based diagnostic approaches will allow for
385 the first time a rational design of BDQ and CFZ -containing MDR-TB therapies.

386 Materials and Methods

387 *In vitro* evolution experiments

388 *In vitro* experiments were carried out with H37Rv strain (ATCC 27294). Bedaquiline was purchased from
389 Janssen-Cilag GmbH and clofazimine was ordered from Sigma (C8895-1G) both were reconstituted from
390 powder in DMSO and stored at -20°C.

391 Evolutionary experiments were conducted with the *Mycobacterium tuberculosis* complex (Mtbc) lab strain
392 H37Rv. The bacteria were cultured from frozen stocks and (pre)cultured at 37° in 7H9 medium,
393 supplemented with 0.2% glycerol, 10% oleic albumin dextrose catalase (OADC), and 0.05% Tween^{80}
394 (termed culture medium). At exponential growth phase (between 0.3 and 0.6 optical density (OD) of
395 600nm), bacteria were transferred to new culture medium with a final OD of 0.05 in 50 mL, additionally
396 supplemented with either bedaquiline (BDQ) or clofazimine (CFZ). Four concentrations below the
397 minimum inhibitory concentration (MIC) were included for each drug at 1:2, 1:4, 1:8, and 1:16 below the

398 strain MIC (and an antibiotic-free control). The MIC of the susceptible WT ancestor was determined to be
399 0.12 mg/L BDQ and 0.5 mg/L CFZ in this culturing system.

400 These experiments were carried out over 20 days with five culture passages total. Culture passages were
401 conducted every four days with a final culture OD of 0.05 transferred. After passages 1, 3, and 5; bacteria
402 were plated on selective agar plates of 7H11, supplemented with 0.5% glycerol, 10% OADC and either
403 0.12 (1:2 MIC) or 0.25 (MIC) mg/L of BDQ, or 0.25 (1:2 MIC) mg/L CFZ. After 21-28 days of growth on
404 selective agar plates, single colonies were transferred to culture medium, grown for 2-4 weeks, then two
405 1 mL aliquots were frozen at -80, and samples were taken for whole genome sequencing (WGS). All
406 remaining colonies (collected by experimental condition) were collected for WGS, i.e. total population
407 sequencing.

408 **Whole genome sequencing**

409 Isolates from *in vitro* evolutionary experiments underwent DNA isolation by CTAB method [70]. Paired-end
410 DNA library preparations and sequencing was performed with Illumina technology (Nextera-XT and
411 NextSeq500) according to the manufacturer's instructions and with a minimum average genome coverage
412 of 50x. Fastq files (raw sequencing data) were mapped to the *M. tuberculosis* H37Rv reference genome
413 (GenBank ID: NC_000962.3) using the MTBseq pipeline [71].

414 Briefly, for the analysis of single mutants we considered single nucleotide polymorphisms (SNPs) which
415 were covered by a minimum of four reads in both forward and reverse orientation, four reads calling the
416 allele with at least a phred score of 20, and an allele frequency of 75%. For the detection of insertions and
417 deletion (InDels) we allowed a frequency over 50%.

418 Deep sequencing was conducted for population diversity analysis with an average genome wide coverage
419 of 200 – 420; SNPs and InDels were called with at least one forward and one reverse read and a phred
420 score >20 for at least 2 reads.

421 Three isolates underwent sequencing using the PacBio Sequel II System (Pacific Biosciences). Libraries
422 were prepared with the SMRTbell® Express Template Prep Kit 2.0 according to the manufacturer's
423 instructions. Barcoded overhang adapters for multiplexing were ordered at IDT (Integrated DNA
424 Technologies). During demultiplexing barcodes were filtered for a minimum quality of 50t, which yielded

425 long read sequencing data of an average of 4.4 GB and mean subread length of 9 KB. Long read sequences
426 were *de novo* assembled using PacBio SMRT® Link software version 9 and the "Microbial Assembly"
427 workflow with a set genome size of 4.5 GB with default parameters.

428 **Screening of Rv0678 mutations in clinical samples via the CRyPTIC strain collection**

429 Patient derived Mtbc isolates were collected by CRyPTIC partners throughout 27 countries and analyzed
430 in 14 different laboratories. Strains were selected from the CRyPTIC database for this study if they had
431 matched genotype and phenotype data, a high quality BDQ phenotype, and a mutation in BDQ resistance
432 associated gene (either *atpE*, *Rv1979c*, *pepQ*, *mmpL/S5*, or *Rv0678*). This led to the curation of 179 strains
433 in total.

434 The full analysis pipeline for CRyPTIC is documented in [paper in preparation], but we outline the key steps
435 here. Sequencing reads were deposited at the European Bioinformatics Institute and run through a
436 bespoke bioinformatics pipeline (publicly available here: https://github.com/iqbal-lab-org/clockwork). In
437 short, reads were filtered against human and other microbial species reads before being mapped to the
438 H37Rv reference genome. Two parallel variant callers (SamTools and Cortex) were used [72,73], one of which
439 makes high sensitivity SNP calls (SamTools) and one which makes high specificity SNP and indel calls
440 (Cortex). A graph-based adjudication tool (Minos, https://github.com/iqbal-lab-org/minos) was then used
441 to combine these results and create a final set of variants for downstream bioinformatics analysis. All
442 strains were re-genotyped at positions that were variant in at least one CRyPTIC sample, creating a final
443 variant call file with a call for each variant position in the *M. tuberculosis* H37Rv genome.

444 **Phenotyping**

445 *In vitro* single selected mutants were further analyzed for phenotypic drug susceptibility to both BDQ and
446 CFZ by broth microtiter dilution as a resazurin assay. All mutants were grown in antibiotic free culture
447 media to exponential growth phase (optical density of 0.3-0.8), then diluted to the McFarland standard of
448 1, with an additional 1:10 dilution, 100 µl of which was seeded in 96-well flat bottom plates (about $1x10^5$
449 CFU per well). Next 100 µL of antibiotic was added to each well with final concentrations of BDQ as follows:
450 8, 4, 2, 1, 0.5, 0.25, 0.12, and 0 mg/L; and CFZ: 16, 8, 4, 2, 1, 0.5, 0.25, 0.12, and 0 mg/L. Plates were sealed
451 with permeable tape, and incubated at 37° standing, in sealed plastic boxes. After nine days incubation at
452 37°, 30µl of resazurin was added to each well. After overnight incubation, fluorescence and absorbance

453 were measure in Biotek plate reader Synergy 2. MIC was determined as the highest concentration which
454 no bacterial growth was detected, either visually or by fluorescence measurement.

455 For MGIT susceptibility testing, 100µl of frozen bacterial stocks were transferred to Löwenstein-Jensen
456 agar slants and incubated at 37° for three weeks. BACTEC™ MGIT™ 960 SIRE Kit was used and test was
457 carried out according to manufacture instructions. Saline 0.83% solution was used for adjusting bacterial
458 suspension concentration instead of Middlebrook 7H9 broth. Drug concentrations tested were 0.5, 1.0,
459 and 2.0 mg/L for BDQ; 0.5, 1.0, and 2.0 mg/L for CFZ; 0.03, 0.06, and 0.12 mg/L for delamanid; and 0.5,
460 1.0, and 2.0 mg/L for linezolid. All MGIT tubes which were positive (growth units reached 400) before the
461 antibiotic free growth control were considered resistant. An H37Rv WT strain was not included in this
462 test.CRyPTIC strains were phenotyped using either the UKMYC5 plate or the updated UKMYC6 plate [74].
463 Plates were sealed, incubated at 37°C, and read at 14 days. In addition to manual plate readings, all plate
464 images underwent an automated reading using AMyGDA software [75]. Plates without essential agreement
465 for a drug MIC were marked as low quality and sent to a citizen science project (BashTheBug,
466 http://bashthebug.net) for additional verification. Plates with exact agreement between at least two
467 phenotyping methods were marked as high quality. Low (no method agrees) and medium (two methods
468 with essential agreement) quality phenotypes were excluded from this analysis.

469 All strains included in this study had at least phenotypic data for BDQ. Some strains do not include CFZ
470 data due to missing information, low quality analysis, or removal due to experimental error.

471 ***Rv0678* variant literature search**

472 An extensive search was performed to include and summarize previously published BDQ and/or CFZ
473 resistant associated mutations from *in vitro*, *in vivo*, and patient derived isolates. We used PubMed,
474 Google, and Google Scholar to search literature published from 2014 to June 2021. Search criteria included
475 the key "TB", "Mycobacterium tuberculosis", "MTB", "bedaquiline", "clofazimine", "treatment", "clinical
476 report", "patient", "MDR-TB", "XDR-TB", "diarylquinoline", and "drug resistance".

477 Mutations in all resistance associated genes: *Rv0678*, *atpE*, *Rv1979c*, and *pepQ* were included in our final
478 analysis, all variants with low MICs or multiple mutations in the same gene were excluded (Table 2, Table
479 S5).

480 **Structural modelling**

481 The crystal structure of Rv0678 (PDB ID: 4NB5) was visualized using UCSF Chimera [76]. Structural alignment
482 of the wHTH domain was performed using the MatchMaker tool in Chimera with a Needleman-Wunsch
483 algorithm using a BLOSUM-62 matrix and was iteratively pruned until no long atom-pair was > 2 Å resulting
484 in a final average RMSD of 0.81 Å over the 5 guide structures (PDB IDs: 5HSO, 5HSM, 4FX0, 4FX4, 4YIF). All
485 protein stability, protein-protein and protein-DNA interactions were modeled using established mCSM
486 methods with either ligand-bound or DNA-bound Rv0678 [77,78]. Mutations that presented both resistant
487 and susceptible phenotypes were treated as resistant during statistical calculations. Screening our
488 resistance catalogue for missense mutations that were resolved in the structure with phenotypes yielded
489 107 unique missense mutations. mCSM tools are available at http://biosig.unimelb.edu.au/biosig.

490 **Molecular dynamics simulations**

491 All the systems simulated in the present work, including Rv0678-WT, Rv0678-A101E, Rv0678-L40V and
492 Rv0678-L40F, were prepared using BiKi Life Sciences Software Suite version 1.3.5 of BiKi Technologies s.r.l
493 [79]. Each simulated system consisted of Rv0678 homodimer unit X-ray structure (PDB ID 4NB5) and the
494 mutations were generated using UCSF Chimera software [76]. The Amber14 force field was used in all
495 molecular dynamic simulations. TIP3P waters were added to make an orthorhombic box. Adding a suitable
496 number of counter-ions neutralized the overall system. Then, the energy of the whole system was
497 minimized. Four consecutive equilibration steps were then performed: 1) 100 ps in the NVT ensemble at
498 100K with the protein backbone restrained (k=1000 kJ/mol nm^2), 2) 100 ps in the NVT ensemble at 200K
499 with the protein backbone similarly restrained, 3) 100 ps in the NVT ensemble at 300K with the protein
500 backbone restrained, and 4) 1000 ps in NPT ensemble at 300K with no restraints. For atoms less than
501 1.1nm apart, electrostatic forces were calculated directly; for atoms further apart electrostatics were
502 calculated using the Particle Mesh Ewald. Van der Waals forces were only calculated for atoms within 1.1
503 nm of one another. The temperature was held constant using the velocity rescale thermostat, which is a
504 modification of the Berendsen's coupling algorithm. Finally, simulations 100 ns long in the NPT ensemble
505 at 300K were performed for each system. To detect allosteric signal transmission networks across the
506 protein surface, defined as interconnected pocket motions, we carried out the allosteric communication
507 network analysis using the Pocketron module in BiKi Life Sciences Suite version 1.3.5 [79].

508 **CRyPTIC Ethics statements as at 22 Feb 2021.** Approval for CRyPTIC study was obtained by Taiwan Centers
509 for Disease Control IRB No. 106209, University of KwaZulu Natal Biomedical Research Ethics Committee
510 (UKZN BREC) (reference BE022/13) and University of Liverpool Central University Research Ethics
511 Committees (reference 2286), Institutional Research Ethics Committee (IREC) of The Foundation for
512 Medical Research, Mumbai (Ref nos. FMR/IEC/TB/01a/2015 and FMR/IEC/TB/01b/2015), Institutional
513 Review Board of P.D. Hinduja Hospital and Medical Research Centre, Mumbai (Ref no. 915-15-CR [MRC]),
514 scientific committee of the Adolfo Lutz Institute (CTC-IAL 47-J / 2017) and in the Ethics Committee
515 (CAAE: 81452517.1.0000.0059) and Ethics Committee review by Universidad Peruana Cayetano
516 Heredia (Lima, Peru) and LSHTM (London, UK).

517 **Authors' contributions.** LS, JC, AS, ZI, MH, KM, CU, DS, CR, KN, PF, MM and SN conceived the idea,
518 designed the study, and analyzed and interpreted the data. All authors contributed to obtaining and
519 assembling the data. LS, JC, AS, ZI, MM, PF and SN wrote the initial draft of the paper. All authors
520 contributed to data interpretation, final drafting of the paper and approved the final version of the
521 manuscript.

522 **Competing interest.** Authors declare no competing interests.

523 **Acknowledgments.** We thank Dr. Musco Giovanna and Dr. Ballabio Federico (San Raffale Hospital) for
524 their help in the molecular dynamics set up. We would like to acknowledge the technical support at
525 Research Center Borstel, especially Vanessa Mohr, Tanja Niemann, Carina Hahn, Silvia Mass, and Doreen
526 Beyer for their aid in the laboratory.

527 **Funding.** All *in vitro* experimental work was financially supported though EvoLUNG: Evolutionary Medicine
528 of the Lung, a Leibniz science campus for evolutionary medicine research (http://evolung.fz-
529 borstel.de/). JC is funded by the Rhodes Trust and the Stanford University Medical Scientist Training
530 program. PWF is funded by the NIHR Oxford Biomedical Research Centre, Oxford University Hospitals NHS
531 Foundation Trust, John Radcliffe Hospital, Oxford, UK. The views expressed are those of the author(s) and
532 not necessarily those of the NHS, the NIHR or the Department of Health. The work of the CRyPTIC
533 consortium was supported by the Bill & Melinda Gates Foundation [OPP1133541] and the Welcome Trust
534 [200205/Z/15/Z].

535 **References**

536 1. *WHO | Global tuberculosis report 2019*. http://www.who.int/tb/publications/global_report/en/
537 (2019).

538 2. Dheda, K. *et al.* The Lancet Respiratory Medicine Commission: 2019 update: epidemiology,
539 pathogenesis, transmission, diagnosis, and management of multidrug-resistant and incurable
540 tuberculosis. *The Lancet Respiratory Medicine* **7**, 820–826 (2019).

541 3. Dheda, K. *et al.* The epidemiology, pathogenesis, transmission, diagnosis, and management of
542 multidrug-resistant, extensively drug-resistant, and incurable tuberculosis. *The Lancet Respiratory*
543 *Medicine* **5**, 291–360 (2017).

544 4. Kendall, E. A., Fofana, M. O. & Dowdy, D. W. Burden of transmitted multidrug resistance in epidemics
545 of tuberculosis: a transmission modelling analysis. *The Lancet Respiratory Medicine* **3**, 963–972
546 (2015).

547 5. Merker, M. *et al.* Compensatory evolution drives multidrug-resistant tuberculosis in Central Asia. *eLife*
548 **7**, e38200 (2018).

549 6. Shah, N. S. *et al.* Transmission of Extensively Drug-Resistant Tuberculosis in South Africa. *New England*
550 *Journal of Medicine* **376**, 243–253 (2017).

551 7. WHO Consolidated Guidelines on Tuberculosis, Module 4: Treatment - Drug-Resistant Tuberculosis
552 Treatment. https://www.who.int/publications-detail-redirect/9789240007048.

553 8. WHO | Rapid Communication: Key changes to the treatment of drug-resistant tuberculosis. *WHO*
554 http://www.who.int/tb/publications/2019/rapid_communications_MDR/en/.

555 9. Huitric, E. *et al.* Rates and mechanisms of resistance development in Mycobacterium tuberculosis to
556 a novel diarylquinoline ATP synthase inhibitor. *Antimicrobial Agents and Chemotherapy* (2010)
557 doi:10.1128/AAC.01611-09.

558 10. Almeida, D. *et al.* Mutations in pepQ confer low-level resistance to bedaquiline and clofazimine in
559 Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy* (2016)
560 doi:10.1128/AAC.00753-16.

561 11. Zhang, S. *et al.* Identification of novel mutations associated with clofazimine resistance in
562 Mycobacterium tuberculosis. *Journal of Antimicrobial Chemotherapy* (2015) doi:10.1093/jac/dkv150.

563 12. de Vos, M., Ley, S. D. & Cox, H. Bedaquiline micro-heteroresistance after tuberculosis treatment
564 cessation. *N Engl J Med* **380**, 2178–2180 (2019).

565 13. Ismail, N. A. *et al.* Defining Bedaquiline Susceptibility, Resistance, Cross-Resistance and Associated
566 Genetic Determinants: A Retrospective Cohort Study. *EBioMedicine* (2018)
567 doi:10.1016/j.ebiom.2018.01.005.

568 14. Villellas, C. *et al.* Unexpected high prevalence of resistance-associated Rv0678 variants in MDR-TB
569 patients without documented prior use of clofazimine or bedaquiline. *The Journal of antimicrobial*
570 *chemotherapy* (2017) doi:10.1093/jac/dkw502.

571 15. Kranzer, K. *et al.* New World Health Organization treatment recommendations for multidrug-resistant
572 tuberculosis: Are we well enough prepared? *American Journal of Respiratory and Critical Care*
573 *Medicine* (2019) doi:10.1164/rccm.201902-0260LE.

574 16. Polsfuss, S. *et al.* Emergence of Low-level Delamanid and Bedaquiline Resistance During Extremely
575 Drug-resistant Tuberculosis Treatment. *Clinical Infectious Diseases* **69**, 1229–1231 (2019).

576 17. Andres, S. *et al.* Bedaquiline-Resistant Tuberculosis: Dark Clouds on the Horizon. *Am J Respir Crit Care*
577 *Med* **201**, 1564–1568 (2020).

578 18. Islam, M. M. *et al.* Drug resistance mechanisms and novel drug targets for tuberculosis therapy.
579 *Journal of Genetics and Genomics* **44**, 21–37 (2017).

580 19. Milano, A. *et al.* Azole resistance in Mycobacterium tuberculosis is mediated by the MmpS5–MmpL5
581 efflux system. *Tuberculosis* **89**, 84–90 (2009).

582 20. Radhakrishnan, A. *et al.* Crystal Structure of the Transcriptional Regulator Rv0678 of Mycobacterium
583 tuberculosis*. *Journal of Biological Chemistry* **289**, 16526–16540 (2014).

584 21. Gao, Y.-R. *et al.* Structural analysis of the regulatory mechanism of MarR protein Rv2887 in M.
585 tuberculosis. *Sci Rep* **7**, 6471 (2017).

586 22. Hartkoorn, R. C., Uplekar, S. & Cole, S. T. Cross-Resistance between Clofazimine and Bedaquiline
587 through Upregulation of MmpL5 in Mycobacterium tuberculosis. *Antimicrobial Agents and*
588 *Chemotherapy* **58**, 2979–2981 (2014).

589 23. Andries, K. *et al.* Acquired Resistance of Mycobacterium tuberculosis to Bedaquiline. *PLoS ONE* **9**,
590 e102135 (2014).

591 24. Bloemberg, G. V. *et al.* Acquired Resistance to Bedaquiline and Delamanid in Therapy for Tuberculosis.
592 *N Engl J Med* **373**, 1986–1988 (2015).

593 25. Hoffmann, H. *et al.* Delamanid and Bedaquiline Resistance in Mycobacterium tuberculosis Ancestral
594 Beijing Genotype Causing Extensively Drug-Resistant Tuberculosis in a Tibetan Refugee. *Am J Respir*
595 *Crit Care Med* **193**, 337–340 (2016).

596 26. Liu, Y. *et al.* Reduced Susceptibility of Mycobacterium tuberculosis to Bedaquiline During
597 Antituberculosis Treatment and Its Correlation With Clinical Outcomes in China. *Clinical Infectious*
598 *Diseases* (2020) doi:10.1093/cid/ciaa1002.

599 27. Nimmo, C. *et al.* Dynamics of within-host Mycobacterium tuberculosis diversity and heteroresistance
600 during treatment. *EBioMedicine* **55**, (2020).

601 28. Xu, J. *et al.* Primary Clofazimine and Bedaquiline Resistance among Isolates from Patients with
602 Multidrug-Resistant Tuberculosis. *Antimicrob Agents Chemother* **61**, e00239-17, e00239-17 (2017).

603 29. Nimmo, C. *et al.* Population-level emergence of bedaquiline and clofazimine resistance-associated
604 variants among patients with drug-resistant tuberculosis in southern Africa: a phenotypic and
605 phylogenetic analysis. *The Lancet Microbe* **1**, e165–e174 (2020).

606 30. Beckert, P. *et al.* MDR M. tuberculosis outbreak clone in Eswatini missed by Xpert has elevated
607 bedaquiline resistance dated to the pre-treatment era. *Genome Medicine* **12**, 104 (2020).

608 31. Liu, A. *et al.* Selective Advantage of Resistant Strains at Trace Levels of Antibiotics: a Simple and
609 Ultrasensitive Color Test for Detection of Antibiotics and Genotoxic Agents. *Antimicrobial Agents and*
610 *Chemotherapy* **55**, 1204–1210 (2011).

611 32. Gullberg, E. *et al.* Selection of Resistant Bacteria at Very Low Antibiotic Concentrations. *PLoS*
612 *Pathogens* **7**, e1002158 (2011).

613 33. Dartois, V. The path of anti-tuberculosis drugs: from blood to lesions to mycobacterial cells. *Nat Rev*
614 *Micro* **12**, 159–167 (2014).

615 34. Sarathy, J. P. *et al.* Prediction of Drug Penetration in Tuberculosis Lesions. *ACS Infect Dis* **2**, 552–563
616 (2016).

617 35. Prideaux, B. *et al.* The association between sterilizing activity and drug distribution into tuberculosis
618 lesions. *Nature Medicine* (2015) doi:10.1038/nm.3937.

619 36. Dheda, K. *et al.* Drug-penetration gradients associated with acquired drug resistance in patients with
620 tuberculosis. *American Journal of Respiratory and Critical Care Medicine* (2018)
621 doi:10.1164/rccm.201711-2333OC.

622 37. Cicchese, J. M., Dartois, V., Kirschner, D. E. & Linderman, J. J. Both Pharmacokinetic Variability and
623 Granuloma Heterogeneity Impact the Ability of the First-Line Antibiotics to Sterilize Tuberculosis
624 Granulomas. *Front. Pharmacol.* **11**, (2020).

625 38. Strydom, N. *et al.* Tuberculosis drugs' distribution and emergence of resistance in patient's lung
626 lesions: A mechanistic model and tool for regimen and dose optimization. *PLOS Medicine* **16**,
627 e1002773 (2019).

628 39. Greenwood, D. J. *et al.* Subcellular antibiotic visualization reveals a dynamic drug reservoir in infected
629 macrophages. *Science* **364**, 1279–1282 (2019).

630 40. Fearns, A., Greenwood, D. J., Rodgers, A., Jiang, H. & Gutierrez, M. G. Correlative light electron ion
631 microscopy reveals in vivo localisation of bedaquiline in Mycobacterium tuberculosis–infected lungs.
632 *PLOS Biology* **18**, e3000879 (2020).

633 41. Kadura, S. *et al.* Systematic review of mutations associated with resistance to the new and repurposed
634 Mycobacterium tuberculosis drugs bedaquiline, clofazimine, linezolid, delamanid and pretomanid.
635 *Journal of Antimicrobial Chemotherapy* **75**, 2031–2043 (2020).

636 42. Song, H., Sandie, R., Wang, Y., Andrade-Navarro, M. A. & Niederweis, M. Identification of outer
637 membrane proteins of Mycobacterium tuberculosis. *Tuberculosis* **88**, 526–544 (2008).

638 43. Gu, S. *et al.* Comprehensive Proteomic Profiling of the Membrane Constituents of a Mycobacterium
639 tuberculosis Strain*S. *Molecular & Cellular Proteomics* **2**, 1284–1296 (2003).

640 44. Hunt, M. *et al.* ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads.
641 *Microbial Genomics,* **3**, e000131 (2017).

642 45. Consortium, T. Cr. & Fowler, P. W. Epidemiological cutoff values for a 96-well broth microdilution
643 plate for high-throughput research antibiotic susceptibility testing of M. tuberculosis. *medRxiv*
644 2021.02.24.21252386 (2021) doi:10.1101/2021.02.24.21252386.

645 46. Vargas, R. *et al.* The role of epistasis in amikacin, kanamycin, bedaquiline, and clofazimine resistance
646 in Mycobacterium tuberculosis complex. *bioRxiv* 2021.05.07.443178 (2021)
647 doi:10.1101/2021.05.07.443178.

648 47. Merker, M. *et al.* Phylogenetically informative mutations in genes implicated in antibiotic resistance
649 in Mycobacterium tuberculosis complex. *Genome Med* **12**, 27 (2020).

650 48. Zimenkov, D. V. *et al.* Examination of bedaquiline- and linezolid-resistant Mycobacterium tuberculosis
651 isolates from the Moscow region. *Journal of Antimicrobial Chemotherapy* **72**, 1901–1906 (2017).

652 49. Hecht, M., Bromberg, Y. & Rost, B. Better prediction of functional effects for sequence variants. *BMC*
653 *Genomics* **16**, S1 (2015).

654 50. Kumarevel, T. The MarR Family of Transcriptional Regulators - A Structural Perspective. *Antibiotic*
655 *Resistant Bacteria - A Continuous Challenge in the New Millennium* (2012) doi:10.5772/28565.

656 51. Hong, M., Fuangthong, M., Helmann, J. D. & Brennan, R. G. Structure of an OhrR-ohrA operator
657 complex reveals the DNA binding mechanism of the MarR family. *Molecular Cell* (2005)
658 doi:10.1016/j.molcel.2005.09.013.

659 52. Nimmo, C. *et al.* Bedaquiline resistance in drug-resistant tuberculosis HIV co-infected patients.
660 *European Respiratory Journal* (2020) doi:10.1183/13993003.02383-2019.

661 53. Makhado, N. A. *et al.* Outbreak of multidrug-resistant tuberculosis in South Africa undetected by
662 WHO-endorsed commercial tests: an observational study. *The Lancet Infectious Diseases* **18**, 1350–
663 1359 (2018).

664 54. Gröschel, M. I. *et al.* Pathogen-based precision medicine for drug-resistant tuberculosis. *PLoS Pathog*
665 **14**, e1007297 (2018).

666 55. Feuerriegel, S. *et al.* Rapid genomic first- and second-line drug resistance prediction from clinical
667 Mycobacterium tuberculosis specimens using Deeplex®-MycTB. *European Respiratory Journal* (2020)
668 doi:10.1183/13993003.01796-2020.

669 56. Jouet, A. *et al.* Deep amplicon sequencing for culture-free prediction of susceptibility or resistance to
670 13 anti-tuberculous drugs. *European Respiratory Journal* (2020) doi:10.1183/13993003.02338-2020.

671 57. Cabibbe, A. M. *et al.* Application of Targeted Next-Generation Sequencing Assay on a Portable
672 Sequencing Platform for Culture-Free Detection of Drug-Resistant Tuberculosis from Clinical Samples.
673 *Journal of Clinical Microbiology* **58**, (2020).

674 58. Kumar, S. & Jena, L. Understanding Rifampicin Resistance in Tuberculosis through a Computational
675 Approach. *Genomics Inform* **12**, 276–282 (2014).

676 59. Zhang, Q. *et al.* Uncovering the Resistance Mechanism of Mycobacterium tuberculosis to Rifampicin
677 Due to RNA Polymerase H451D/Y/R Mutations From Computational Perspective. *Front. Chem.* **7**,
678 (2019).

679 60. Marney, M. W., Metzger, R. P., Hecht, D. & Valafar, F. Modeling the structural origins of drug
680 resistance to isoniazid via key mutations in Mycobacterium tuberculosis catalase-peroxidase, KatG.
681 *Tuberculosis* **108**, 155–162 (2018).

682 61. Mehmood, A. *et al.* Structural Dynamics Behind Clinical Mutants of PncA-Asp12Ala, Pro54Leu, and
683 His57Pro of Mycobacterium tuberculosis Associated With Pyrazinamide Resistance. *Front. Bioeng.*
684 *Biotechnol.* **7**, (2019).

685 62. Carter, J. J. *et al.* Prediction of pyrazinamide resistance in Mycobacterium tuberculosis using
686 structure-based machine learning approaches. *bioRxiv* 518142 (2019) doi:10.1101/518142.

687 63. Tunstall, T. *et al.* Combining structure and genomics to understand antimicrobial resistance.
688 *Computational and Structural Biotechnology Journal* **18**, 3377–3394 (2020).

689 64. Karmakar, M. *et al.* Empirical ways to identify novel Bedaquiline resistance mutations in AtpE. *PLOS*
690 *ONE* **14**, e0217169 (2019).

691 65. Dartois, V. The path of anti-tuberculosis drugs: From blood to lesions to mycobacterial cells. *Nature*
692 *Reviews Microbiology* (2014) doi:10.1038/nrmicro3200.

693 66. Strydom, N. *et al.* Tuberculosis drugs' distribution and emergence of resistance in patient's lung
694 lesions: A mechanistic model and tool for regimen and dose optimization. *PLoS Medicine* (2019)
695 doi:10.1371/journal.pmed.1002773.

696 67. Müller, B., Borrell, S., Rose, G. & Gagneux, S. The heterogeneous evolution of multidrug-resistant
697 Mycobacterium tuberculosis. *Trends in Genetics* **29**, 160–169 (2013).

698 68. Veziris, N. *et al.* Rapid emergence of Mycobacterium tuberculosis bedaquiline resistance: lessons to
699 avoid repeating past errors. *Eur Respir J* **49**, 1601719 (2017).

700 69. Ghodousi, A. *et al.* Acquisition of Cross-Resistance to Bedaquiline and Clofazimine following
701 Treatment for Tuberculosis in Pakistan. *Antimicrob Agents Chemother* **63**, e00915-19,
702 /aac/63/9/AAC.00915-19.atom (2019).

703 70. van Soolingen, D., Hermans, P. W., de Haas, P. E., Soll, D. R. & van Embden, J. D. Occurrence and
704 stability of insertion sequences in Mycobacterium tuberculosis complex strains: evaluation of an
705 insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis.
706 *Journal of Clinical Microbiology* **29**, 2578–2586 (1991).

707 71. Kohl, T. A. *et al.* MTBseq: a comprehensive pipeline for whole genome sequence analysis of
708 Mycobacterium tuberculosis complex isolates. *PeerJ* **6**, e5895 (2018).

709 72. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).

710 73. Iqbal, Z., Caccamo, M., Turner, I., Flicek, P. & McVean, G. De novo assembly and genotyping of variants
711 using colored de Bruijn graphs. *Nature Genetics* **44**, 226–232 (2012).

712 74. Rancoita, P. M. V. *et al.* Validating a 14-Drug Microtiter Plate Containing Bedaquiline and Delamanid
713 for Large-Scale Research Susceptibility Testing of Mycobacterium tuberculosis. *Antimicrob Agents*
714 *Chemother* **62**, e00344-18, /aac/62/9/e00344-18.atom (2018).

715 75. Fowler, P. W. *et al.* Automated detection of bacterial growth on 96-well plates for high-throughput
716 drug susceptibility testing of mycobacterium tuberculosis. *Microbiology (United Kingdom)* (2018)
717 doi:10.1099/mic.0.000733.

718 76. Pettersen, E. F. *et al.* UCSF Chimera—A visualization system for exploratory research and analysis.
719 *Journal of Computational Chemistry* **25**, 1605–1612 (2004).

720 77. Pires, D. E. V., Blundell, T. L. & Ascher, D. B. MCSM-lig: Quantifying the effects of mutations on protein-
721 small molecule affinity in genetic disease and emergence of drug resistance. *Scientific Reports* **6**, 1–8
722 (2016).

723 78. Pires, D. E. V., Ascher, D. B. & Blundell, T. L. MCSM: Predicting the effects of mutations in proteins
724 using graph-based signatures. *Bioinformatics* **30**, 335–342 (2014).

725 79. Decherchi, S., Bottegoni, G., Spitaleri, A., Rocchia, W. & Cavalli, A. BiKi Life Sciences: A New Suite for
726 Molecular Dynamics and Related Methods in Drug Discovery. *J. Chem. Inf. Model.* **58**, 219–224 (2018).

727 80. WHO | Technical report on critical concentrations for TB drug susceptibility testing of medicines used
728 in the treatment of drug-resistant TB. *WHO*

729 http://www.who.int/tb/publications/2018/WHO_technical_report_concentrations_TB_drug_susce
730 ptibility/en/.

731 81. Kaniga, K. *et al.* A Multilaboratory, Multicountry Study To Determine Bedaquiline MIC Quality Control
732 Ranges for Phenotypic Drug Susceptibility Testing. *J Clin Microbiol* **54**, 2956–2962 (2016).

733 82. Schön, T. *et al.* Wild-type distributions of seven oral second-line drugs against Mycobacterium
734 tuberculosis. *Int J Tuberc Lung Dis* **15**, 502–509 (2011).

735 83. Battaglia, S. *et al.* Characterization of Genomic Variants Associated with Resistance to Bedaquiline
736 and Delamanid in Naive Mycobacterium tuberculosis Clinical Strains. *J Clin Microbiol* **58**, (2020).

737 84. Yang, J. S., Kim, K. J., Choi, H. & Lee, S. H. Delamanid, Bedaquiline, and Linezolid Minimum Inhibitory
738 Concentration Distributions and Resistance-related Gene Mutations in Multidrug-resistant and
739 Extensively Drug-resistant Tuberculosis in Korea. *Ann Lab Med* **38**, 563–568 (2018).

740 85. Wang, G. *et al.* Prevalence and molecular characterizations of seven additional drug resistance among
741 multidrug-resistant tuberculosis in China: A subsequent study of a national survey. *J Infect* **82**, 371–
742 377 (2021).

743 86. Xu, J. *et al.* Verapamil Increases the Bioavailability and Efficacy of Bedaquiline but Not Clofazimine in
744 a Murine Model of Tuberculosis. *Antimicrobial Agents and Chemotherapy* **62**, e01692-17.

745 87. Pang, Y. *et al.* In Vitro Drug Susceptibility of Bedaquiline, Delamanid, Linezolid, Clofazimine,
746 Moxifloxacin, and Gatifloxacin against Extensively Drug-Resistant Tuberculosis in Beijing, China.
747 *Antimicrob Agents Chemother* **61**, e00900-17, e00900-17 (2017).

748