

1 **Rapid Acquisition and Transmission of Drug Resistance Amongst Beijing Lineage**  
2 ***Mycobacterium tuberculosis* in Vietnam**

3  
4 Matthew Silcocks<sup>1\*</sup>, Xuling Chang<sup>1,2,3\*</sup>, Nguyen Thuy Thuong Thuong<sup>4,5\*</sup>, Youwen Qin<sup>6,7</sup>, Dang Thi  
5 Minh Ha<sup>8</sup>, Phan Vuong Khac Thai<sup>8</sup>, Srinivasan Vijay<sup>4,5,9</sup>, Do Dang Anh Thu<sup>4</sup>, Hoang Ngoc Nhung<sup>4</sup>,  
6 Nguyen Huu Lan<sup>8</sup>, Nguyen Thi Quynh Nhu<sup>4</sup>, David Edwards<sup>10\*\*</sup>, Artika Nath<sup>6</sup>, Kym Pham<sup>11</sup>, Nguyen  
7 Duc Bang<sup>8</sup>, Tran Thi Hong Chau<sup>4,12</sup>, Guy Thwaites<sup>4,5</sup>, A. Dorothee Heemskerk<sup>13</sup>, Chiea Chuen Khor<sup>14</sup>,  
8 Yik Ying Teo<sup>15</sup>, Michael Inouye<sup>6,16</sup>, Rick Twee-Hee Ong<sup>15</sup>, Maxine Caws<sup>17,18</sup>, Kathryn E. Holt<sup>10,19\*</sup>  
9 and Sarah J. Dunstan<sup>1\*#</sup>

10  
11 <sup>1</sup> Department of Infectious Diseases, University of Melbourne at the Peter Doherty Institute for  
12 Infection and Immunity, Parkville, Victoria, Australia.

13 <sup>2</sup> Department of Paediatrics, Yong Loo Lin School of Medicine, National University of Singapore,  
14 Singapore 119228, Singapore.

15 <sup>3</sup> Khoo Teck Puat – National University Children’s Medical Institute, National University Health  
16 System, Singapore 119074, Singapore.

17 <sup>4</sup> Oxford University Clinical Research Unit, Hospital for Tropical Diseases, District 5, HCMC,  
18 Vietnam,

19 <sup>5</sup> Nuffield Department of Medicine, Centre for Tropical Medicine and Global Health, University of  
20 Oxford, Oxford, UK

21 <sup>6</sup> Cambridge Baker Systems Genomics Initiative, Baker Heart and Diabetes Institute, Melbourne,  
22 Victoria, Australia.

23 <sup>7</sup> School of BioSciences, The University of Melbourne, Melbourne, Victoria, Australia.

24 <sup>8</sup> Pham Ngoc Thach Hospital for TB and Lung Disease, District 5, HCMC, Vietnam

25 <sup>9</sup> Theoretical Microbial Ecology, Friedrich Schiller University Jena, Germany

26 <sup>10</sup> Department of Infectious Diseases, Central Clinical School, Monash University, Melbourne,  
27 Victoria, Australia

28 <sup>11</sup> Department of Clinical Pathology, The University of Melbourne, Victoria, Australia

29 <sup>12</sup> Hospital for Tropical Diseases, District 5, HCMC, Vietnam

30 <sup>13</sup> Department of Medical Microbiology and Infection Prevention, Amsterdam University Medical  
31 Centre, Amsterdam, the Netherlands

32 <sup>14</sup> Genome Institute of Singapore, Singapore

33 <sup>15</sup> Saw Swee Hock School of Public Health, National University of Singapore

34 <sup>16</sup> Cambridge Baker Systems Genomics Initiative, Department of Public Health and Primary Care,  
35 University of Cambridge, Cambridge, United Kingdom

36 <sup>17</sup> Liverpool School of Tropical Medicine, Liverpool, UK

37 <sup>18</sup>Birat Nepal Medical Trust, Kathmandu, Nepal

38 <sup>19</sup>Department of Infection Biology, London School of Hygiene & Tropical Medicine, London, UK

39 \*equal contribution \*\*deceased #corresponding author

40 **Abstract**

41 Whole genome sequencing (WGS) and phenotypic drug susceptibility testing was performed  
42 on a collection of 2,542 *Mycobacterium tuberculosis* (*Mtb*) isolates from tuberculosis (TB)  
43 patients recruited in Ho Chi Minh City (HCMC), Vietnam, to investigate *Mtb* diversity, the  
44 prevalence and phylodynamics of drug resistance, and *in silico* resistance prediction with  
45 sequencing data. Amongst isolates tested phenotypically against first-line drugs, we observed  
46 high rates of streptomycin [STR, 37.7% (N=573/1,520)] and isoniazid resistance [INH,  
47 25.7% (N=459/1,786)], and lower rates of resistance to rifampicin [RIF, 4.9% (N=87/1,786)]  
48 and ethambutol [EMB, 4.2% (N=75/1,785)]. Resistance to STR and INH was predicted  
49 moderately well when applying the TB-Profiler algorithm to WGS data (sensitivities of 0.81  
50 and 0.87 respectively), while resistance to RIF and EMB was predicted relatively poorly  
51 (sensitivities of 0.70 and 0.44 respectively). Rates of multidrug-resistance [(MDR, 3.9%  
52 (N=69/1,786)], and resistance to a number of second-line drugs [Para-aminosalicylic acid  
53 (29.6% N=79/267), Amikacin (15.4% N=41/267) and Moxifloxacin (21.3%), N=57/267],  
54 were found to be high within a global context. Comparing rates of drug resistance among  
55 lineages, and exploring the dynamics of resistance acquisition through time, suggest the  
56 Beijing lineage (lineage 2.2) acquires *de novo* resistance mutations at higher rates and suffers  
57 no apparent fitness cost acting to impede the transmission of resistance. We infer resistance  
58 to INH and STR to have arisen earlier, on average, than resistance to RIF, and to be more  
59 widespread across the phylogeny. The high prevalence of ‘background’ INH resistance,  
60 combined with high rates of RIF mono-resistance (20.7%, N=18/87) suggests that rapid  
61 assays for INH resistance will be valuable in this setting. These tests will allow the detection  
62 of INH mono-resistance, and will allow MDR isolates to be distinguished from isolates with  
63 RIF mono-resistance.

64

## 65 **Introduction**

66 Tuberculosis (TB) remains a global epidemic with one quarter of the world's population  
67 estimated to be infected. The impact of the COVID-19 pandemic on essential TB services has  
68 reversed years of progress, with the number of newly diagnosed TB patients falling to 5.8  
69 million in 2020, much less than the estimated 10 million who developed TB [Global  
70 Tuberculosis Report, 2021 (WHO, 2021)]. An increase in TB deaths in 2020 was also  
71 estimated (1.3 million in HIV negative, and 214,000 in HIV-positive) due largely to a  
72 reduction in the number of people treated for drug resistant TB [Global Tuberculosis Report,  
73 2021 (WHO, 2021)]. Geographically, the burden of disease lies mainly in South-East Asia  
74 (44% of TB cases in 2018) [Global Tuberculosis Report, 2019 (WHO, 2019)], with 86% of  
75 new TB cases worldwide being reported from 30 high TB burden countries in 2020 [Global  
76 Tuberculosis Report, 2021 (WHO, 2021)].

77

78 Vietnam is a high TB burden country, designated by its high number of incident TB and  
79 multi-drug resistant TB (MDR-TB) cases [WHO global lists of high burden countries for TB,  
80 multidrug/rifampicin-resistant TB (MDR/RR-TB) and TB/HIV, 2021–2025, (WHO, 2021)].  
81 In 2018, 174,000 and 8,600 people in Vietnam fell ill with TB and drug-resistant TB,  
82 respectively. Of the patients with drug-resistant TB, only 36.3% were laboratory confirmed  
83 and 36.2% started on second-line treatment.

84

85 Although the COVID-19 pandemic has set back recent progress, steps towards the WHO  
86 END-TB targets have been made. To be able to achieve a 95% reduction of TB deaths by  
87 2035 (The END TB strategy, World Health Organisation, 2015), modern technologies must  
88 be embraced to find innovative ways to accelerate TB control and elimination. One such

89 technology is genomic sequencing, which offers a myriad of opportunities for innovation in  
90 diagnostics, treatment, prevention and control of TB.

91

92 Genotype data, for example, can be used to predict drug resistance in *Mtb* isolates, by  
93 querying databases of confirmed or suspected resistance conferring variants (CRyPTIC  
94 Consortium et al. 2018; Walker et al. 2015; Walker et al. 2022). This approach provides an  
95 efficient alternative to traditional phenotypic methods which are prone to human error,  
96 contamination, and require time-consuming bacterial culturing (Zignol et al. 2018; Gygli et  
97 al. 2019; Walker et al. 2022). This technology has the capacity to reduce the probability of  
98 misassignment of drugs to patients infected with resistant *Mtb* (compared to standardised  
99 drug regimens), and to reduce the time before patients receive effective treatment, potentially  
100 leading to more favourable treatment outcomes (Iketleng et al. 2018).

101

102 Despite the promise of genotype-based drug resistance prediction, its accuracy has been  
103 shown to vary according to human population, *Mtb* lineage, type of drug, and the prediction  
104 protocol used (Papaventsis et al. 2017). A more comprehensive understanding of the effect  
105 these factors have on prediction accuracy, and the levels of accuracy which are attainable  
106 across diverse cohorts, is necessary prior to implementation of these test in all clinical  
107 settings.

108

109 Genomics also provides insight into the emergence, transmission and overall dynamics of  
110 drug resistance in *Mtb*, via the use of a phylogenetic toolkit (Eldholm et al. 2015;  
111 Woolenberg et al. 2017; Brown et al. 2019; Cohen et al. 2019; Casali et al. 2014; Manson et  
112 al. 2017). Prior studies have produced time-calibrated phylogenies to date the acquisition of  
113 drug resistance in *Mtb* lineages (Eldholm et al. 2015; Cohen et al. 2015), and explored their

114 expansion through time (Brown et al. 2019). Others have used models of ancestral sequence  
115 reconstruction (Woolenberg et al. 2017; Cohen et al. 2015; Manson et al. 2017) and SNP  
116 clustering methods (Casali et al. 2014; Yang et al. 2017) to compare the rates of acquired (*de*  
117 *novo*), versus primary (i.e. transmitted) drug resistance within a population.

118

119 While these studies naturally vary in their scope, cohort size, geographic scale and setting,  
120 their results have highlighted similar trends in drug resistance evolution. They show that the  
121 global drug resistance burden has arisen through both the *de novo* acquisition of drug  
122 resistance, and through the transmission of drug resistant *Mtb* to new hosts (Woolenberg et al.  
123 2017; Casali et al. 2014; Ektefaie et al. 2021; Yang et al. 2017; Mai et al. 2018). They also  
124 show consistent trends in the order of drug resistance acquisition, with INH resistance  
125 generally arising prior to RIF resistance, and being more deeply rooted in the phylogeny  
126 (Cohen et al. 2015; Ektefaie et al. 2021; Manson et al. 2017). Finally, when characterising  
127 *Mtb* lineage diversity, these studies have implicated Beijing lineage isolates in many  
128 outbreaks (Woolenberg et al. 2017; Yang et al. 2017; Casali et al. 2014).

129

130 Here we apply these tools to investigate resistance evolution in a South East Asian context,  
131 using a cohort from HCMC, Vietnam, which has a high frequency of TB, drug resistance, and  
132 the Beijing lineage. This study builds on our previous investigation of *Mtb* diversity in  
133 HCMC (Holt et al. 2018), by using an expanded isolate collection and resistance phenotype  
134 data to explore the prevalence and transmission dynamics of drug resistance, and our ability  
135 to predict it *in silico*.

136

## 137 **Results**

### 138 **Clinical characteristics**

139 To characterise the diversity and drug resistance of *Mtb* in HCMC, Vietnam, we analysed the  
140 genomic sequences of N=2,542 isolates (post quality-filtering) cultured from TB patients  
141 between 2001-2013. A subset of these genomes (N=1,627), derived from pulmonary TB  
142 (PTB) patients, were published previously (Holt et al. 2018). Here these are supplemented  
143 with an additional N=914 novel genomes, comprising N=747 from PTB cases and 167 from  
144 TBM cases (Heemskerk et al. 2016; Thwaites et al. 2004). A single lineage 5 genome was  
145 included as an outgroup for phylogenetic analyses.

146

147 PTB patients were sputum culture positive and  $\geq 18$  years old (Thai et al. 2018) whereas  
148 TBM patients were cerebral spinal fluid culture positive and  $\geq 15$  years old (Heemskerk et al.  
149 2016; Thwaites et al. 2004). All TB patients were human immunodeficiency virus (HIV)-  
150 negative. Overall the patients were predominantly male (71.5% male, 28.5% female) with a  
151 median (interquartile) age of 39 (26-49) years. We found the frequencies of males and  
152 females to differ across clinical phenotypes for PTB cases, with 73.1% of cases being male,  
153 and 26.9% being female. We calculated TBM cases to be more evenly distributed across  
154 genders (47.9% male and 52.1% female), however noted males to be more frequent in the  
155 larger TBM cohort from which these samples were derived (Heemskerk et al. 2016).

156

### 157 **Genetic diversity and *Mtb* lineage**

158 In our HCMC PTB/TBM genome dataset (N=2,542), we found East Asian lineage 2 to be the  
159 most prevalent (N=1,615, 63.6%), followed by the East African-Indian Ocean lineage 1  
160 (N=649, 25.5%), the European-American lineage 4 (N=275, 10.8%) and the Delhi Central  
161 Asian lineage 3 (N=2, 0.1%; Figure 1). The pattern of *Mtb* genomic diversity in the HCMC  
162 PTB/TBM genome dataset was thus consistent with our published subset of Vietnamese  
163 genomes (Holt et al. 2018).

164

165 To gain a more nuanced understanding of patterns of *Mtb* diversity in our genome dataset, we  
166 also considered lineage distributions at a more granular level of phylogenetic resolution.  
167 While we recognise that several recent studies have proposed new *Mtb* lineage nomenclature  
168 schemes (Napier et al. 2020; Freschi et al. 2021; Gisch et al. 2022; Palittapongarnpim et al.  
169 2018), here we decided to define lineages according to the most detailed level of resolution  
170 described by Coll et al. (2014), until a consensus is established, and a clearer understanding  
171 of sublineage distributions under these new schemes is obtained.

172

173 Consistent with our prior survey (Holt et al. 2018), lineage 1 isolates mainly belonged to  
174 sublineages 1.1.1.1 (N=525) and 1.1.1 (N=83), which predominate in Mainland South East  
175 Asia (Netikul et al. 2021; Freschi et al. 2021; Figure 1). We also observed isolates belonging  
176 to sublineage 1.2.1 (N=33), which attains highest frequencies in Island South East Asia  
177 (Netikul et al. 2021; Menardo et al. 2021).

178

179 Isolates from lineage 2 could be partitioned into both the ‘Proto-Beijing’ (L2.1; N=86) and  
180 Beijing (L2.2; N=1,529) sublineages (Luo et al. 2015; Figure 1). Beijing lineage isolates  
181 belonged to both L2.2.1 (N=1,446), which has a wide global distribution (Holt et al. 2018), as  
182 well as the less prevalent L2.2.2 (N=55) (Shiticov et al. 2017). We documented a wide array  
183 of L4 sublineages, consistent with a history of repeated introductions into Southeast Asia, as  
184 inferred by Holt *et al.* (2018). The most frequent L4 sublineages were 4.1, 4.2, 4.4 and 4.5.

185

186 Also consistent with Holt *et al.* (2018), we found *Mtb* lineage to be significantly associated  
187 with both the sex (Table 1), and age of the patient. While patients were predominantly male,  
188 lineage 2 and sublineage 2.2.1 were more prevalent amongst female patients (lineage 2;

189 67.4% of females vs 62.0% of males;  $p=0.010$ ), (sublineage 2.2.1; 62.3% of females vs  
190 55.8% of males;  $p=0.005$  (when excluding other L2 isolates from test)). Lineage 1 strains  
191 showed higher prevalence amongst males (27.3% of males vs 21.1% of females;  $p=0.001$ ;  
192 Table 1). Lineage 2 was significantly associated with younger people (median 37 years), and  
193 lineage 1 with older people (median 43 years,  $p=8.36 \times 10^{-8}$ ).

194

### 195 **Phenotypic drug resistance of PTB isolates**

196 Phenotypic drug susceptibility testing was performed for a subset of the sequenced isolates  
197 (N=1,786). We assessed phenotypic resistance to three first line drugs (INH, RIF and EMB)  
198 using MGIT (N=1,786), and also tested a subset of these isolates using UKMYC5 (N=267).  
199 Isolates were deemed resistant to an antimicrobial if they were classified as resistant by one  
200 or both methods. In total, N=459/1,786 (25.7%) isolates were resistant to INH, N=87/1,786  
201 (4.9%) to RIF, and N=75/1785 (4.2%) to EMB. In total, 3.9% (N=69/1,786) were MDR (INH  
202 and RIF resistant; Table 2), and 20.7% (N=18/87) of RIF resistant isolates were INH  
203 sensitive. STR resistance, which was tested by MGIT only, was identified in 37.7%  
204 (573/1,520) of the tested isolates. The frequencies of resistance to second-line antibiotics,  
205 determined for 267 isolates only, are given in Table 2.

206

207 We noted that rates of resistance to all first line drugs were higher amongst Beijing lineage  
208 isolates (L2.2) than isolates belonging to lineages 1 and 4 [Pearson's Chi-Squared test; INH  
209 ( $p=2.07 \times 10^{-4}$ ), RIF ( $p=0.003$ ), EMB ( $p=0.002$ ) and STR ( $p=8.29 \times 10^{-12}$ )]. This finding was in  
210 keeping with several prior studies (Casali et al. 2014; Borrell et al. 2009; Neimann et al.  
211 2010; Woolenberg et al. 2017).

212

### 213 **Prediction of drug resistance using genotype data**

214 We assessed our ability to correctly predict drug resistance using genotype data, and found it  
215 to vary according to the drug, and the phenotype (either ‘resistant’ or ‘susceptible’) being  
216 predicted. We analyzed raw sequencing data with TB-Profiler (Phelan et al. 2019), and  
217 calculated sensitivity and specificity metrics using standard approaches (Mahe et al. 2019).  
218 Sensitivity, which measure the proportion of phenotypically resistant strains correctly  
219 predicted to be resistant, for first line drugs was 0.87, 0.70, 0.44 and 0.81 for INH, RIF, EMB  
220 and STR respectively (Figure 2a). Specificity, which measure the proportion of  
221 phenotypically susceptible strains correctly predicted to be susceptible, was 0.98, 0.99, 0.97  
222 and 0.92 for these same four drugs respectively (Figure 2a).

223

224 Positive predictive values, which measure the proportion of isolates with a resistance  
225 prediction which were phenotypically resistant, were 0.93, 0.79, 0.40 and 0.87 for INH, RIF,  
226 EMB and STR (Figure 2a). Negative predictive values, which measure the proportion of  
227 isolates with a susceptible prediction which were phenotypically susceptible were 0.96, 0.98,  
228 0.98 and 0.89 (Figure 2a).

229

230 The frequency distribution of mutations explaining resistance to first-line drugs (Figure 2b)  
231 was skewed towards a number of commonly reported variants (WHO, 2021a). In particular,  
232 the *katG*-Ser315Thr substitution was most frequently observed in INH resistant cases  
233 (N=299/459, 65.1% of resistant isolates), followed by *fabG1*-15C>T (N=71/459, 15.5%). RIF  
234 resistance was most commonly explained by *rpoB*-Ser450Leu (N=41/87, 47.1%), and STR  
235 resistance by *rpsL*-Lys43Arg (N=243/573, 42.4%) and *rpsL*-Lys88Arg (N=100/573, 17.5%).  
236 EMB displayed a wider spectrum of resistance conferring mutations, with the most frequent  
237 being *embB*-Met306Val (N=13/75, 17.3%) and *embB*-Met306Ile (N=7/75, 9.3%).

238 Importantly, in the majority of cases (N=42/75, 56%), resistance to EMB was not explained  
239 by any known markers.

240

241 Resistance prediction for second-line drugs was overall poor [sensitivity values: BDQ=0  
242 (N=1 resistant isolate), ETH=0.41 (N=22), AMI=0 (N=41), LEV=0.07 (N=31), MXF=0.04  
243 (N=57), DLM=0.8 (N=5), LZD=0 (N=32), CFZ=0 (N=7), BDQ=0 (N=1), KAN=0 (N=15)  
244 and PAS=0.05 (N=79)].

245

246 We found the frequency of putative resistance-conferring variants to be similar between TBM  
247 and PTB isolates (30.5% vs 24.9% for INH, 1.2% vs 4.7% for RIF, 3.0% vs 4.9% for EMB,  
248 40.1% vs 36.4% for STR). We further verified that TBM isolates were dispersed throughout  
249 the phylogeny, and were not clustered in monophyletic clades (supp fig S1), justifying their  
250 inclusion in the subsequent analysis of antibacterial resistance dynamics.

251

## 252 **Dynamics of drug resistance acquisition and transmission**

253 We investigated the dynamics of drug resistance evolution within our cohort, to understand  
254 temporal and lineage-specific trends in drug resistance development, and to characterise the  
255 relative burden of acquired versus transmitted resistance. To do this, we used ancestral state  
256 reconstruction to map individual resistance mutation events to the phylogeny and identify  
257 instances of transmitted (Figure 3a; red points. see Methods) and acquired resistance (Figure  
258 3a; blue points).

259

260 Across the phylogeny, we inferred 826 unique mutation events leading to INH, STR or RIF  
261 resistance (EMB was not modelled due to the low positive predictive power of known  
262 variants, see Fig 2a). We observed a trend in the order of resistance acquisition (Figure 3),

263 with INH and STR mutations arising earlier (closer to the root of the tree) than RIF mutations  
264 (Figures 3a and b; median heights of 0.01285, 0.01291 and 0.01299 for INH, STR and RIF  
265 variants, respectively). This inference of early and widespread INH resistance was consistent  
266 with the conclusions of prior studies (Ektefaie et al. 2021; Manson et al. 2017).

267

268 We estimated the relative burden of transmitted (as opposed to acquired) resistance by  
269 observing the relative density of resistance mutations mapped to nodes, as opposed to  
270 terminal branches of the phylogeny (see Methods). Across the entire tree, we find that 63.6%  
271 (402/632) of isolates with INH resistance were categorised as cases of transmitted resistance  
272 (i.e. with their resistance mutation mapping to a node of the phylogeny). A similar percentage  
273 was calculated for STR variants (66.6%; 519/779).

274

275 These rates, while clearly influenced by the design and density of the sampling regime,  
276 indicate that DR-TB burden in HCMC is explained by both *de novo* acquisition, and  
277 transmission of resistance. As this, and our previous analysis of *Mtb* genetic variation within  
278 Vietnam (Holt et al. 2018) sample only a limited fraction of incident TB cases in HCMC,  
279 these estimates of transmission rates are likely to be conservative.

280

### 281 **Lineage specific trends in drug resistance evolution**

282 We observed clear lineage-specific trends in the dynamics of drug resistance evolution.  
283 Consistent with prior estimates of drug resistance rates across lineages (Casali et al. 2014;  
284 Borrell et al. 2009; Neimann et al. 2010; Woolenberg et al. 2017), we observed lineage 2  
285 isolates to acquire resistance conferring mutations at a faster rate than isolates from other  
286 lineages. 1.13% of all mutation events across the lineage 2 clade were resistance conferring; a  
287 higher figure than for lineages 1 and 4 (0.35% and 0.32% respectively).

288

289 Lineage 2 isolates were also found to contribute disproportionately to the number of  
290 transmitted resistance cases within our cohort. Of the total 402 cases of transmitted INH  
291 resistance across the phylogeny, 347 (86.3%) of these occurred within the lineage 2 subtree,  
292 and 20 (5.0%) and 35 (8.7%) occurred within the lineage 1 and 4 subtrees respectively. These  
293 proportions were higher than expected when considering the overall lineage distribution  
294 across all isolates in the phylogeny (25.5% L1, 63.5% L2, 10.8% L4). Similar findings were  
295 reported when considering STR resistance variants (487 of 519 (93.8%) cases of transmitted  
296 resistance occurred within the lineage 2 subtree). Thus, lineage 2 isolates possessed both a  
297 higher absolute rate of resistance acquisition, and a higher propensity to transmit resistance.

298

299 We also found a greater proportion of resistant lineage 2 isolates could be attributed to  
300 transmission when compared to other lineages. 74.6% (N=347/465) of INH resistant lineage  
301 2 isolates were classified as cases of transmitted resistance, a higher figure than for lineages 1  
302 (20%; N=20/100) and 4 (52.2%; N=35/67). Similar results were observed when considering  
303 STM resistance (L2 – 71.9%; N=487/677, L1 – 15.9%; N=11/69, L4 – 63.6%; N=21/33).

304

305 Finally, we investigated whether *Mtb* isolates with genotypic resistance were less likely to be  
306 transmitted than susceptible isolates. We are interested in this question, as resistance  
307 conferring variants are typically associated with fitness costs in *Mtb*, and other bacterial  
308 species (Alame Emame et al. 2021; Nguyen et al. 2018; Gagneux et al. 2009; Andersson &  
309 Hughes, 2010). Determining the potential of resistant isolates to transmit to new human hosts  
310 will therefore be relevant in determining future drug resistance trajectories worldwide.

311

312 Firstly, similar to prior investigations (Casali et al. 2014; Yang et al. 2017; Mai et al. 2018),  
313 we measured the proportion of resistant isolates which were part of a transmission cluster,  
314 and compared this to the proportion of susceptible isolates in a transmission cluster (Figure  
315 4a). For this and the below analysis we restricted our focus to sublineage 2.2.1, which has  
316 previously been shown to display, on average, shorter terminal branches and higher rates of  
317 resistance than other lineages in this setting (Holt et al. 2018). We also restricted our analysis  
318 to focus on the two drugs with substantial numbers of resistance mutations across the  
319 phylogeny, INH and STR. We found isolates with INH and STR resistance mutations to be  
320 significantly less likely to be found in a transmission cluster when using a 5-SNP threshold  
321 (Figure 4a; Pearson's Chi-Squared Test  $p=6\times 10^{-4}$  for INH and  $p=0.003$  for STR), but found  
322 no significant differences when using 10- and 15- SNP thresholds ( $p=0.51$  and  $p=0.18$  for  
323 INH;  $p=0.24$  and  $p=0.28$  for STR).

324

325 Secondly, as per approaches used in other studies of *Mtb* phylogenetics [(Holt et al. 2018;  
326 Freschi et al. 2021) (Figure 4b)], we measured the lengths of terminal branches in the tree  
327 leading to resistant and susceptible tips (Figure 4b). We did not find significant differences in  
328 the lengths of terminal branches leading to isolates with INH and STR resistance mutations  
329 (Figure 4b; Mann Whitney U tests;  $p=0.64$  for INH;  $p=0.36$  for STR). This evidence suggests  
330 that the presence of known drug resistance mutations does not substantially impede the  
331 transmission of *Mtb* lineage 2.2.1 isolates from HCMC.

332

333

334

335 **Discussion**

336

337 **High prevalence of phenotypic drug resistance in *Mtb* isolates from HCMC**

338 Our investigation has highlighted the high rates of resistance to several first and second line  
339 drugs, in addition to MDR, within a Vietnamese cohort. The results reported here are in  
340 keeping with a recent national survey of Vietnam (Nhung et al. 2015), which also reports  
341 high rates of STR (27.4%) and INH (18.9%) resistance, and lower rates of RIF (4.1%) and  
342 EMB (3.4%) resistance amongst new TB cases. This study adds an additional dimension to  
343 this survey, by recording rates of phenotypic resistance to second line drugs, and finding rates  
344 for PAS, MXF, AMI, LZD and LEV to all be greater than 10%.

345

346 When considered globally, individual first line drug resistance values were higher than has  
347 been reported in parts of Africa (Ismail et al. 2018; Gehre et al. 2016; Chonde et al. 2010),  
348 East Asia (He et al. 2008; Bai et al. 2007), and South Asia (Tahseen et al. 2016), but lower  
349 than a South American cohort (Quispe et al. 2020). The second line drug resistance rates we  
350 observe are higher than reported in the small number of studies which survey second line  
351 resistance in new TB cases (Quispe et al. 2020; Bai et al. 2007; Zhao et al. 2012).

352

353 **Concordance between drug resistance phenotype and genotype-based predictions**

354 Our analysis has also provided an assessment of the accuracy of genotype-based drug  
355 resistance prediction within a Vietnamese population. Contextualising the results we obtained  
356 is difficult, as prior studies have typically relied on much smaller sample sizes (Kohl et al.  
357 2018; Schleusener et al. 2017; Macedo et al. 2018), have not reported data for the same  
358 subset of drugs analysed here (Zignol et al. 2018), or have systematically removed isolates  
359 with uncharacterised mutations in target genes (Cryptic Consortium, 2018). We contrast our

360 results here with those of Mahe et al. (2019), who analyse a large sample (>6,000 isolates)  
361 using the same prediction catalogue and methodology.

362

363 Our sensitivity values for INH and STR prediction are comparable with those of Mahe et al.  
364 (2019) (0.89 and 0.77 respectively), and specificity values for all four first-line drugs are  
365 equal to or higher than those of Mahe et al. (2019) (INH=0.97; RIF=0.98; EMB=0.93;  
366 STR=0.91). Our sensitivity values for the two drugs with the lowest levels of phenotypic  
367 resistance, RIF and EMB are, however, lower than those of Mahe et al. (2019) (0.91 and 0.93  
368 respectively). Also, our ability to predict resistance to second-line drugs was considerably  
369 lower than Mahe et al. (2019), who report sensitivity values of 0.92, 0.92 and 0.89 for KAN,  
370 AMI and aggregated Fluoroquinolones respectively.

371

372 We conclude that prediction accuracies for the two drugs with the highest rates of resistance,  
373 INH and STR, are comparable with global benchmarking standards. We also note that the  
374 frequency distributions of markers explaining resistance in correctly predicted isolates are  
375 consistent with a recent survey of the Vietnamese population (Mai et al. 2018), and the  
376 recently published WHO catalogue (WHO, 2021a). In all instances, the most commonly  
377 reported resistance conferring variants in our cohort were also amongst the most frequent in  
378 the catalogue collated by the WHO (2021a).

379

380 A future avenue of research may involve investigating the genetic basis of resistance in the  
381 high number of EMB resistant isolates which were predicted to be susceptible using the TB-  
382 Profiler catalogue (N=42/75, 56%). It is clear that large studies correlating genotypic  
383 predictions of resistance with clinical outcomes are essential to improve the accuracy of  
384 genotypic resistance prediction and to resolve discrepancies between phenotypic and

385 genotypic prediction, especially for the antimycobacterial drugs with poorly characterised  
386 resistance mechanisms.

387

388 **High rates of acquisition and transmission of drug resistance in Beijing lineage isolates.**

389 This investigation has explored the dynamics of *Mtb* drug resistance evolution in a high  
390 burden South-East Asian setting. Analysis of cohorts from other geographical regions have  
391 alternately emphasized the roles of transmission (Woolenberg et al. 2017; Casali et al. 2014;  
392 Yang et al. 2017) or *de novo* acquisition (Cohen et al. 2015; Ektefaie et al. 2021; Manson et  
393 al. 2017) in explaining rates of drug resistance. The drug resistance burden in our Vietnamese  
394 cohort was explained by both the *de novo* acquisition of resistance, and ongoing transmission  
395 of resistant isolates. As can reasonably be expected, the rate of transmitted INH resistance  
396 reported here (63.6%) is lower than reported in Woolenberg et al. (2017) (90%), who explore  
397 an outbreak within a small Russian population, yet higher than values reported by Ektefaie et  
398 al. 2021 (~40%) who analyse a diverse global cohort.

399

400 Importantly, we find that drug resistance burden in Vietnam is predominantly driven by  
401 Beijing lineage isolates, which develop resistance conferring mutations at a faster rate than  
402 other isolates, and display a tendency to transmit resistance between hosts. These findings  
403 contribute to our understanding of the biology and life history of Beijing lineage *Mtb* isolates,  
404 which have previously been shown to display enhanced virulence and transmissibility than  
405 isolates from other lineages (Holt et al. 2018). This data therefore highlights an additional  
406 danger associated with the Beijing lineage, which is gradually supplanting endemic lineage 1  
407 isolates within Vietnam (Holt et al. 2018).

408

409 **Limitations of this analysis.**

410 A caveat which must be applied to our findings is that our phylogenetic methods are unable  
411 to model the evolution of resistance in isolates for which their phenotype cannot be explained  
412 by markers in the TB-Profiler catalogue. It is possible that inferences surrounding the  
413 evolution of RIF resistance, for instance, may change with the identification of more  
414 resistance conferring markers. Resistance explained by indels is similarly unable to be  
415 modelled using our methods, however, this class of variant represents a small fraction of our  
416 dataset.

417

#### 418 **Implications for TB control in Vietnam.**

419 In addition to highlighting dynamics of drug resistance evolution within our cohort, we have  
420 also established that resistance to INH and STR arose earlier, on average, than resistance to  
421 RIF, and is now more widespread within the Vietnamese population. The inference of early  
422 INH resistance evolution is consistent with prior studies, and mimics results obtained from a  
423 South African population (Cohen et al. 2015), and two surveys of global *Mtb* isolates  
424 (Manson et al. 2017; Ektefaie et al. 2017).

425

426 We conclude, as do Manson et al. (2017), that a rapid assay for INH resistance will allow the  
427 detection of ‘pre-MDR’ TB, and offer the high number of patients with INH resistant TB,  
428 treatment options which include other drugs. Unlike the cohort analysed by Manson et al.  
429 (2017), however, we find that a high percentage of RIF resistance isolates are susceptible to  
430 INH (20.7%). This finding further supports the utility of an INH resistance assay, which may  
431 be applied after a patient tests positive for RIF resistance using the MTB/RIF Xpert test.  
432 Several WHO approved tests for INH resistance could be implemented (WHO, 2016).  
433 Application of such INH resistance tests will allow the 20% of RIF resistant cases which  
434 remain susceptible to INH to be treated with an INH-containing regimen.

435

436 Finally, the high rates of resistance to a number of second line drugs highlights the utility of

437 WGS-based individualised therapy for drug resistant TB in Vietnam. The poor sensitivity

438 values calculated, however, stresses the need to develop a greater understanding of the

439 genetic variants implicated in resistance to these drugs.

440

## 441 **Materials and Methods**

### 442 **Study population**

443 Bacterial isolates (N=2,619) from patients with pulmonary TB (PTB) (N=2,446) and  
444 tuberculous meningitis (TBM) (N=173) were collected as part of larger clinical studies  
445 (Heemskerk et al. 2016; Thai et al. 2018; Thwaites et al. 2004). Patients with PTB were  
446 defined as HIV-negative adults (>18 years) sputum culture positive for *M. tuberculosis*.  
447 Isolates (N=1,654) were collected in 8 district TB units (DTUs) in Ho Chi Minh City  
448 (HCMC), Vietnam between December 2008 and July 2011 (Thai et al. 2018). A further 792  
449 isolates from PTB patients were similarly collected at these DTUs in HCMC as an extension  
450 to this clinical study, from 2011-2013. Patients with TBM were defined as HIV-negative  
451 patients, >15 years old, with cerebral spinal fluid (CSF) culture positive *Mtb* and were  
452 recruited into two clinical trials conducted at the Hospital for Tropical Diseases and Pham  
453 Ngoc Thach Hospital for TB and Lung Diseases in HCMC (Heemskerk et al. 2016; Thwaites  
454 et al. 2004). 62 were collected between 2001-2003 (Thwaites et al. 2004) and 111 were  
455 collected between 2011-2015 (Heemskerk et al. 2016), with a total of 173 isolates from TBM  
456 patients included in this genomic study.

457

### 458 **Ethical approvals**

459 All study protocols for PTB and TBM studies were approved by the Institutional Research  
460 Board of Pham Ngoc Thach Hospital for TB and Lung Disease, HCMC Health Services and  
461 the Oxford Tropical Research Ethics Committee, UK. Approval for the genomics study was  
462 granted from the Health Sciences Human Ethics Sub-Committee at the University of  
463 Melbourne, Australia (ID:1340458). Written informed consent was obtained from all  
464 patients.

465

#### 466 **Phenotypic drug susceptibility testing (DST)**

467 Phenotypic DST was performed using two techniques. In the first method, isolates were  
468 subcultured in Mycobacterial Growth Indicator Tubes (MGIT) for phenotypic DST on the  
469 first line drugs, INH 0.1µg/ml; STR 1.0µg/ml; RIF 1.0µg/ml; EMB; 5.0µg/ml, using the  
470 BACTEC MGIT 960 SIRE Kit (Becton Dickinson) according to the manufacturer's  
471 instructions. The second method used the UKMYC5 plate designed by the CRYPTIC  
472 consortium which enables minimum inhibitory concentration (MIC) measurement for 14  
473 different anti-tuberculosis compounds. The UKMTC5 MIC plate method was used as  
474 described in Rancoita et al. (2018). The critical concentrations used to determine drug  
475 resistance were BDQ 1.0µg/ml; KAN 2.5µg/ml; ETH 5.0µg/ml; AMI 1.0µg/ml; EMB  
476 5.0µg/ml; INH 0.1µg/ml; LEV 1.0µg/ml; MXF 0.5µg/ml; DLM 0.12µg/ml; LZD 1.0µg/ml;  
477 CFZ 1.0µg/ml; RIF 1.0µg/ml; RFB 1.0µg/ml; PAS 4.0µg/ml.

478

#### 479 **DNA extraction and sequencing**

480 Lowenstein Jensen media was used to subculture isolates at the Oxford University Clinical  
481 Research Unit, Vietnam, prior to DNA extraction using the cetyl trimethylammonium  
482 bromide extraction protocol as described previously (Caws et al. 2008). DNA was shipped to  
483 the University of Melbourne (UoM) (N=1,827) and the National University of Singapore  
484 (NUS) (N=792) for whole genome sequencing. At GIS, genomic DNA was first quantified by  
485 Picogreen assay, followed by shearing using the Covaris. Library preparation was done using  
486 a commercially available kit, NEBNext® Ultra™ DNA Library Prep Kit for Illumina®  
487 following the manufacturer's protocol. The quality of the libraries were QC via LabChip GX  
488 or Agilent D1000 ScreenTape before pooling. After pooling, the pooled library was QC  
489 using Agilent high sensitivity DNA kit and KAPA quantification before sequencing on the  
490 Illumina HiSeq 4000 (Illumina, San Diego).

491

## 492 ***Mtb* genome data and SNP calling**

493 *Mtb* genome data from a subset of PTB patients have been previously described and is  
494 denoted here as the ‘published subset’ (Holt et al. 2018). These genomes are available at the  
495 European Nucleotide Archive (accession ID PRJNA355614;  
496 <https://www.ebi.ac.uk/ena/browser/view/PRJNA355614>). The remaining *Mtb* genome data  
497 will be available via the NCBI database upon publication of this manuscript. The complete  
498 genome collection is denoted as the HCMC PTB/TBM genome dataset. Variant calling for  
499 *Mtb* isolates was carried out using the RedDog pipeline V1beta.11  
500 (<https://github.com/katholt/RedDog>) with default settings, which uses BowTie (Langmead et  
501 al. 2009) for read mapping and Samtools (Li, 2011) for variant calling. After variant calling,  
502 samples for which less than 90% of their reads mapped to the *Mtb* H37Rv reference genome  
503 (NC\_000962.3), and with high proportions of heterozygous sites were removed. Variants  
504 called in repetitive regions, as defined by Holt et al. (2018), were also removed. *Mtb* lineages  
505 and sublineages were assigned using fast-lineage-caller (Freschi et al. 2021) on per-sample  
506 vcf files, using the scheme of Coll et al. (2014). Sequence reads from an isolate belonging to  
507 lineage 5 were also incorporated into the above variant calling pipeline as an outgroup for all  
508 subsequent phylogenetic analysis.

509

## 510 **Drug resistance prediction using whole genome sequencing data**

511 Whole genome sequencing-based resistant prediction was performed with TB-Profiler v4.2.0  
512 (Phelan et al. 2019) using the most up to date database available (16/02/2022). Read data  
513 were screened for mutations associated with resistance to 4 first-line drugs, (INH, RIF, EMB  
514 and STR), and second-line drugs for which phenotypes were available, and which are  
515 covered by the TB-Profiler catalogue.

516

517 **Phylogenomic analysis**

518 We inferred a phylogeny from 2,542 *Mtb* genomes using RAxML v8.2 (Stamatakis, 2014),  
519 using a GTR model of nucleotide substitution. Ancestral state reconstruction was then  
520 performed using SNPPar v1.0 (Edwards et al. 2021) with default settings. Biallelic SNPs  
521 conferring drug resistance were defined according to TB-Profiler, and extracted from the  
522 SNPPar output. All downstream analyses of tree diversity metrics were carried out using  
523 custom Unix and R scripts (R Core Team, 2021), and the R packages ape v5.6.2 (Paradis &  
524 Schliep, 2018), treeio v1.18.1 (Wang et al. 2019) and phangorn v2.8.1 (Schliep, 2010).

525

526 To make inferences regarding the dynamics of drug resistance evolution and transmission, we  
527 followed the theory applied by recent studies (Manson et al. 2017; Ektefaie et al. 2021;  
528 Woolenberg et al. 2017). *Mtb* isolates for which resistance mutations were inferred to have  
529 arisen on a terminal branch of the phylogeny were taken to indicate putative cases of acquired  
530 resistance. These mutations were plotted as blue points in Figure 3a on the tip of the  
531 phylogeny downstream of the branch on which they arose.

532

533 We interpreted resistance mutations which could be localised to nodes of the phylogeny, and  
534 which are therefore inherited by multiple descendant tips, to indicate cases of transmitted  
535 resistance. These mutations were plotted as red points in Figure 3a on the node downstream  
536 of the branch on which they arose. The small number of isolates (N=5) across the phylogeny  
537 which developed resistance-conferring variants on their terminal branch, but which had  
538 already descended from a resistant node, were classified as cases of transmitted resistance.

539

540 We excluded EMB from this analysis as our ability to accurately model transmission  
541 dynamics would be limited due to the low predictive power of EMB resistance variants  
542 (59.8% of isolates with putative EMB resistance variants were phenotypically susceptible;  
543 Figure 2a), in addition to the low sensitivity demonstrated previously (Figure 2a).

544

### 545 **Statistical analysis**

546 Analyses were done using R version 4.1.1 and two-sided  $P < 0.05$  was considered statistically  
547 significant. The overall lineage distribution of *Mtb* isolates across gender, clinical phenotype  
548 and resistance to the first and second anti-TB drugs was presented as number and percentage  
549 (%). Pearson's  $\chi^2$  test was performed to investigate the relationship between categorical  
550 variables, including lineages and other clinical characteristics, naming gender, clinical  
551 phenotype and drug resistance. It was also utilized to assess the relationship between rates of  
552 drug resistance and lineages, as well as resistance and transmission ability. Mann-Whitney U  
553 test was employed to test the difference for non-normally distributed variables among  
554 lineages, including count of total mutations relative to the number of resistance conferring  
555 mutations and lengths of terminal branches leading to resistance. The association between  
556 age and lineage (lineage 1 and 2) was assessed using linear regression. To investigate the  
557 prediction accuracy for drug resistance using WGS data we calculated sensitivity, defined as  
558 the proportion of resistant isolates correctly predicted; specificity, defined as the proportion  
559 of susceptible isolates correctly predicted; positive predictive value (PPV) and negative  
560 predictive value (NPV).

561

### 562 **Acknowledgements**

563 We would like to thank the clinical staff who recruited patients into our study from the  
564 following District TB Units (DTUs) in HCMC, Vietnam: Districts 1, 4, 5, 6 and 8, Tan Binh,

565 Binh Thanh and Phu Nhuan; and also our colleagues from Pham Ngoc Thach Hospital for  
566 Tuberculosis and Lung Disease, HCMC Vietnam. We would like to acknowledge the  
567 Cryptic consortium, <http://www.crypticproject.org/>, for providing the UKMYC5 plates for  
568 DST. This work was supported by the National Health and Medical Research Council,  
569 Australia (Investigator grant APP1172853 to SJD); NHMRC (APP1056689) /A\*STAR  
570 (12/1/21/24/6689) joint call to SJD/YYT; the Wellcome Trust UK (research training  
571 fellowship 081814/Z/06/Z to MC), (intermediate fellowship 206724/Z/17/Z to NTTTT) and as  
572 part of their Major Overseas Program in Vietnam (089276/Z/09/Z to JF and 106680/B/14/Z  
573 to GT). The National University of Singapore Yong Loo Lin School of Medicine Aspiration  
574 Fund (NUHSRO/2014/069/AF-New Idea/04) contributed funding to the whole genome  
575 sequencing conducted in this study. MI was supported by the Munz Chair of Cardiovascular  
576 Prediction and Prevention at the Baker Heart and Diabetes Institute and the NIHR Cambridge  
577 Biomedical Research Centre (BRC-1215-20014) [\*]. \*The views expressed are those of the  
578 author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

579

#### 580 **Conflicts of Interest**

581 No conflicts of interest

## 582 **References**

583

584 Andersson, D. I., & Hughes, D. (2010). Antibiotic resistance and its cost: is it possible to  
585 reverse resistance? *Nature Reviews Microbiology*, 8(4), 260–271.  
586 <https://doi.org/10.1038/nrmicro2319>

587

588 Bai, G. H., Park, Y. K., Choi, Y. W., Bai, J. I., Kim, H. J., Chang, C. L., Lee, J. K., Kim, S. J.  
589 (2007). Trend of anti-tuberculosis drug resistance in Korea, 1994-2004. *Int J Tuberc Lung*  
590 *Dis*, 11, 571-576.

591

592 Borrell, S., & Gagneux, S. (2009). Infectiousness, reproductive fitness and evolution of drug-  
593 resistant *Mycobacterium tuberculosis*. *Int J Tuberc Lung Dis*, 13, 1456-1466.

594

595 Brown, T. S., Challagundla, L., Baugh, E. H., Omar, S. V., Mustaev, A., Auld, S. C., Shah,  
596 N. S., Kreiswirth, B. N., Brust, J. C. M., Nelson, K. N., Narechania, A., Kurepina, N.,  
597 Mlisana, K., Bonneau, R., Eldholm, V., Ismail, N., Kolokotronis, S.-O., Robinson, D. A.,  
598 Gandhi, N. R., & Mathema, B. (2019). Pre-detection history of extensively drug-resistant  
599 tuberculosis in KwaZulu-Natal, South Africa. *Proceedings of the National Academy of*  
600 *Sciences*, 116(46), 23284–23291. <https://doi.org/10.1073/pnas.1906636116>

601

602 Casali, N., Nikolayevskyy, V., Balabanova, Y., Harris, S. R., Ignatyeva, O., Kontsevaya, I.,  
603 Corander, J., Bryant, J., Parkhill, J., Nejentsev, S., Horstmann, R. D., Brown, T., &  
604 Drobniewski, F. (2014). Evolution and transmission of drug-resistant tuberculosis in a  
605 Russian population. *Nature Genetics*, 46(3), 279–286. <https://doi.org/10.1038/ng.2878>

606

607 Chonde, T. M., Basra, D., Mfinanga, S. G. M., Range, N., Lwilla, F., Shirima, R. P., van  
608 Deun, A., Zignol, M., Cobelems, F. G., Egwaga, S. M., van Leth, F. (2010). National anti-  
609 tuberculosis drug resistance study in Tanzania. *Int J Tuberc Lung Dis*, 14, 967-72.

610

611 Cohen, K. A., Manson, A. L., Abeel, T., Desjardins, C. A., Chapman, S. B., Hoffner, S.,  
612 Birren, B. W., & Earl, A. M. (2019). Extensive global movement of multidrug-resistant *M.*  
613 *tuberculosis* strains revealed by whole-genome analysis. *Thorax*, 74(9), 882–889.  
614 <https://doi.org/10.1136/thoraxjnl-2018-211616>

615

616 Coll, F., McNerney, R., Guerra-Assunção, J. A., Glynn, J. R., Perdigão, J., Viveiros, M.,  
617 Portugal, I., Pain, A., Martin, N., & Clark, T. G. (2014). A robust SNP barcode for typing  
618 *Mycobacterium tuberculosis* complex strains. *Nature Communications*, 5(1), 4812.  
619 <https://doi.org/10.1038/ncomms5812>

620

621 Edwards, D. J., Duchene, S., Pope, B., & Holt, K. E. (2021). SNPPar: identifying convergent  
622 evolution and other homoplasies from microbial whole-genome alignments. *Microbial*  
623 *Genomics*, 7(12). <https://doi.org/10.1099/mgen.0.000694>

624

625 Ektefaie, Y., Dixit, A., Freschi, L., & Farhat, M. R. (2021). Globally diverse *Mycobacterium*  
626 *tuberculosis* resistance acquisition: a retrospective geographical and temporal analysis of  
627 whole genome sequences. *The Lancet Microbe*, 2(3), e96–e104.  
628 [https://doi.org/10.1016/S2666-5247\(20\)30195-6](https://doi.org/10.1016/S2666-5247(20)30195-6)

629

- 630 Eldholm, V., Monteserin, J., Rieux, A., Lopez, B., Sobkowiak, B., Ritacco, V., & Balloux, F.  
631 (2015). Four decades of transmission of a multidrug-resistant Mycobacterium tuberculosis  
632 outbreak strain. *Nature Communications*, 6(1), 7119. <https://doi.org/10.1038/ncomms8119>  
633
- 634 Emane, A. K. A., Guo, X., Takiff, H. E., & Liu, S. (2021). Drug resistance, fitness and  
635 compensatory mutations in Mycobacterium tuberculosis. *Tuberculosis*, 129, 102091.  
636 <https://doi.org/10.1016/j.tube.2021.102091>  
637
- 638 Freschi, L., Vargas, R., Husain, A., Kamal, S. M. M., Skrahina, A., Tahseen, S., Ismail, N.,  
639 Barbova, A., Niemann, S., Cirillo, D. M., Dean, A. S., Zignol, M., & Farhat, M. R. (2021).  
640 Population structure, biogeography and transmissibility of Mycobacterium tuberculosis.  
641 *Nature Communications*, 12(1), 6099. <https://doi.org/10.1038/s41467-021-26248-1>  
642
- 643 Gagneux, S. (2009). Fitness cost of drug resistance in Mycobacterium tuberculosis. *Clinical  
644 Microbiology and Infection*, 15, 66–68. <https://doi.org/10.1111/j.1469-0691.2008.02685.x>  
645
- 646 Gehre, F., Otu, J., Kendall, L., Forson, A., Kwara, A., Kudzawu, S., Kehinde, A. O., Adebisi,  
647 O., Salako, K., Baldeh, I., Jallow, A., Jallow, M., Dagnra, A., Dissé, K., Kadanga, E. A.,  
648 Idigbe, E. O., Onubogu, C., Onyejebu, N., Gaye-Diallo, A., ... Antonio, M. (2016). The  
649 emerging threat of pre-extensively drug-resistant tuberculosis in West Africa: preparing for  
650 large-scale tuberculosis research and drug resistance surveillance. *BMC Medicine*, 14(1), 160.  
651 <https://doi.org/10.1186/s12916-016-0704-5>  
652
- 653 Gisch, N., Utpatel, C., Gronbach, L. M., Kohl, T. A., Schombel, U., Malm, S., Dobos, K. M.,  
654 Hesser, D. C., Diel, R., Götsch, U., Gerdes, S., Shuaib, Y. A., Ntinginya, N. E., Khosa, C.,  
655 Viegas, S., Kerubo, G., Ali, S., Al-Hajoj, S. A., Ndung'u, P. W., ... Homolka, S. (2022).  
656 Sub-Lineage Specific Phenolic Glycolipid Patterns in the Mycobacterium tuberculosis  
657 Complex Lineage 1. *Frontiers in Microbiology*, 13.  
658 <https://doi.org/10.3389/fmicb.2022.832054>  
659
- 660 Gygli, S. M., Keller, P. M., Ballif, M., Blöchliger, N., Hömke, R., Reinhard, M., Loiseau, C.,  
661 Ritter, C., Sander, P., Borrell, S., Loo, J. C., Avihingsanon, A., Gnokoro, J., Yotebieng, M.,  
662 Egger, M., Gagneux, S., & Böttger, E. C. (2019). Whole-Genome Sequencing for Drug  
663 Resistance Profile Prediction in Mycobacterium tuberculosis. *Antimicrobial Agents and  
664 Chemotherapy*, 63(4). <https://doi.org/10.1128/AAC.02175-18>  
665
- 666 He, G. X., Zhao, Y. L., Jiang, G. L., Liu, Y. H., Xia, H., Wang, S. F., Wang, L. X.,  
667 Borgdorff, M. W., van der Werf, M. J., & van den Hof, S. (2008). Prevalence of tuberculosis  
668 drug resistance in 10 provinces of China. *BMC Infectious Diseases*, 8(1), 166.  
669 <https://doi.org/10.1186/1471-2334-8-166>  
670
- 671 Heemskerck, A. D., Bang, N. D., Mai, N. T. H., Chau, T. T. H., Phu, N. H., Loc, P. P., Chau,  
672 N. V. V., Hien, T. T., Dung, N. H., Lan, N. T. N., Lan, N. H., Lan, N. N., Phong, L. T., Vien,  
673 N. N., Hien, N. Q., Yen, N. T. B., Ha, D. T. M., Day, J. N., Caws, M., & Merson, L. (2016).  
674 Intensified Antituberculosis Therapy in Adults with Tuberculous Meningitis. *New England  
675 Journal of Medicine*, 374(2), 124–134.  
676
- 677 Holt, K. E., McAdam, P., Thai, P. V. K., Thuong, N. T. T., Ha, D. T. M., Lan, N. N., Lan, N.  
678 H., Nhu, N. T. Q., Hai, H. T., Ha, V. T. N., Thwaites, G., Edwards, D. J., Nath, A. P., Pham,  
679 K., Ascher, D. B., Farrar, J., Khor, C. C., Teo, Y. Y., Inouye, M., ... Dunstan, S. J. (2018).

- 680 Frequent transmission of the Mycobacterium tuberculosis Beijing lineage and positive  
681 selection for the EsxW Beijing variant in Vietnam. *Nature Genetics*, 50(6), 849–856.  
682 <https://doi.org/10.1038/s41588-018-0117-9>  
683
- 684 Iketleng, T., Lessells, R., Dlamini, M. T., Mogashoa, T., Mupfumi, L., Moyo, S., Gaseitsiwe,  
685 S., de Oliveira, T. (2018). Mycobacterium tuberculosis Next-Generation Whole Genome  
686 Sequencing: Opportunities and Challenges. *Tuberculosis Research and Treatment*, 1,  
687 <https://doi.org/10.1155/2018/1298542>.  
688
- 689 Ismail, N. A., Mvusi, L., Nanoo, A., Dreyer, A., Omar, S. V., Babatunde, S., Molebatsi, T.,  
690 van der Walt, M., Adelekan, A., Deyde, V., Ihekweazu, C., & Madhi, S. A. (2018).  
691 Prevalence of drug-resistant tuberculosis and imputed burden in South Africa: a national and  
692 sub-national cross-sectional survey. *The Lancet Infectious Diseases*, 18(7), 779–787.  
693 [https://doi.org/10.1016/S1473-3099\(18\)30222-6](https://doi.org/10.1016/S1473-3099(18)30222-6)  
694
- 695 Kohl, T. A., Utpatel, C., Schleusener, V., Filippo, M. R. D., Beckert, P., Cirillo, D. M., &  
696 Niemann, S. (2018). MTBseq: a comprehensive pipeline for whole genome sequence analysis  
697 of Mycobacterium tuberculosis complex isolates. *PeerJ*, 6, e5895.  
698 <https://doi.org/10.7717/peerj.5895>  
699
- 700 Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-  
701 efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10(3),  
702 R25. <https://doi.org/10.1186/gb-2009-10-3-r25>  
703
- 704 Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association  
705 mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, 27(21), 2987–2993. <https://doi.org/10.1093/bioinformatics/btr509>  
706  
707
- 708 Luo, T., Comas, I., Luo, D., Lu, B., Wu, J., Wei, L., Yang, C., Liu, Q., Gan, M., Sun, G.,  
709 Shen, X., Liu, F., Gagneux, S., Mei, J., Lan, R., Wan, K., & Gao, Q. (2015). Southern East  
710 Asian origin and coexpansion of Mycobacterium tuberculosis Beijing family with Han  
711 Chinese. *Proceedings of the National Academy of Sciences*, 112(26), 8136–8141.  
712 <https://doi.org/10.1073/pnas.1424063112>  
713
- 714 Macedo, R., Nunes, A., Portugal, I., Duarte, S., Vieira, L., & Gomes, J. P. (2018). Dissecting  
715 whole-genome sequencing-based online tools for predicting resistance in Mycobacterium  
716 tuberculosis: can we use them for clinical decision guidance? *Tuberculosis*, 110, 44–51.  
717 <https://doi.org/10.1016/j.tube.2018.03.009>  
718
- 719 Mahé, P., Azami, M. E., Barlas, P., & Tournoud, M. (2019). A large scale evaluation of  
720 TBProfiler and Mykrobe for antibiotic resistance prediction in Mycobacterium tuberculosis.  
721 *PeerJ*, 7, e6857. <https://doi.org/10.7717/peerj.6857>  
722
- 723 Mai, T. Q., Martinez, E., Menon, R., Anh, N. T. V., Hien, N. T., Marais, B. J., & Sintchenko,  
724 V. (2018). Mycobacterium tuberculosis Drug Resistance and Transmission among Human  
725 Immunodeficiency Virus–Infected Patients in Ho Chi Minh City, Vietnam. *The American*  
726 *Journal of Tropical Medicine and Hygiene*, 99(6), 1397–1406.  
727 <https://doi.org/10.4269/ajtmh.18-0185>  
728

- 729 Manson, A. L., Cohen, K. A., Abeel, T., Desjardins, C. A., Armstrong, D. T., Barry, C. E.,  
730 Brand, J., Chapman, S. B., Cho, S.-N., Gabrielian, A., Gomez, J., Jodals, A. M., Joloba, M.,  
731 Jureen, P., Lee, J. S., Malinga, L., Maiga, M., Nordenberg, D., Noroc, E., ... Earl, A. M.  
732 (2017). Genomic analysis of globally diverse *Mycobacterium tuberculosis* strains provides  
733 insights into the emergence and spread of multidrug resistance. *Nature Genetics*, 49(3), 395–  
734 402. <https://doi.org/10.1038/ng.3767>  
735
- 736 Menardo, F., Rutaihwa, L. K., Zwyrer, M., Borrell, S., Comas, I., Conceição, E. C., Coscolla,  
737 M., Cox, H., Joloba, M., Dou, H.-Y., Feldmann, J., Fenner, L., Fyfe, J., Gao, Q., de Viedma,  
738 D. G., Garcia-Basteiro, A. L., Gygli, S. M., Hella, J., Hiza, H., ... Gagneux, S. (2021). Local  
739 adaptation in populations of *Mycobacterium tuberculosis* endemic to the Indian Ocean Rim.  
740 *F1000Research*, 10, 60. <https://doi.org/10.12688/f1000research.28318.2>  
741
- 742 Napier, G., Campino, S., Merid, Y., Abebe, M., Woldeamanuel, Y., Aseffa, A., Hibberd, M.  
743 L., Phelan, J., & Clark, T. G. (2020). Robust barcoding and identification of *Mycobacterium*  
744 *tuberculosis* lineages for epidemiological and clinical studies. *Genome Medicine*, 12(1), 114.  
745 <https://doi.org/10.1186/s13073-020-00817-3>  
746
- 747 Netikul, T., Palittapongarnpim, P., Thawornwattana, Y., & Plitphonganphim, S. (2021).  
748 Estimation of the global burden of *Mycobacterium tuberculosis* lineage 1. *Infection, Genetics*  
749 *and Evolution*, 91, 104802. <https://doi.org/10.1016/j.meegid.2021.104802>  
750
- 751 Nguyen, Q. H., Contamin, L., Nguyen, T. V. A., & Bañuls, A.-L. (2018). Insights into the  
752 processes that drive the evolution of drug resistance in *Mycobacterium tuberculosis*.  
753 *Evolutionary Applications*, 11(9), 1498–1511. <https://doi.org/10.1111/eva.12654>  
754
- 755 Nhung, N. V., Hoa, N. B., Sy, D. N., Hennig, C. M., & Dean, A. S. (2015). The Fourth  
756 National Anti-Tuberculosis Drug Resistance Survey in Viet Nam. *The International Journal*  
757 *of Tuberculosis and Lung Disease*, 19(6), 670–675. <https://doi.org/10.5588/ijtld.14.0785>  
758
- 759 Niemann, S., Diel, R., Khechinashvili, G., Gegia, M., Mdivani, N., & Tang, Y.-W. (2010).  
760 *Mycobacterium tuberculosis* Beijing Lineage Favors the Spread of Multidrug-Resistant  
761 Tuberculosis in the Republic of Georgia. *Journal of Clinical Microbiology*, 48(10), 3544–  
762 3550. <https://doi.org/10.1128/JCM.00715-10>  
763
- 764 Palittapongarnpim, P., Ajawatanawong, P., Viratyosin, W., Smittipat, N., Disratthakit, A.,  
765 Mahasirimongkol, S., Yanai, H., Yamada, N., Nedsuwan, S., Imasanguan, W., Kantipong, P.,  
766 Chaiyasirinroje, B., Wongyai, J., Toyo-oka, L., Phelan, J., Parkhill, J., Clark, T. G., Hibberd,  
767 M. L., Ruengchai, W., ... Tokunaga, K. (2018). Evidence for Host-Bacterial Co-evolution via  
768 Genome Sequence Analysis of 480 Thai *Mycobacterium tuberculosis* Lineage 1 Isolates.  
769 *Scientific Reports*, 8(1), 11597. <https://doi.org/10.1038/s41598-018-29986-3>  
770
- 771 Paradis, E. & Schliep, K. (2018) “APE 5.0: An environment for modern phylogenetics and  
772 evolutionary analyses in R,” *Bioinformatics*, 35(3), 526–528.  
773 <https://doi.org/10.1093/bioinformatics/bty633>.  
774
- 775 Papaventsis, D., Casali, N., Kontsevaya, I., Drobniowski, F., Cirillo, D. M., &  
776 Nikolayevskyy, V. (2017). Whole genome sequencing of *Mycobacterium tuberculosis* for  
777 detection of drug resistance: a systematic review. *Clinical Microbiology and Infection*, 23(2),  
778 61–68. <https://doi.org/10.1016/j.cmi.2016.09.008>

- 779  
780 Phelan, J. E., O’Sullivan, D. M., Machado, D., Ramos, J., Oppong, Y. E. A., Campino, S.,  
781 O’Grady, J., McNerney, R., Hibberd, M. L., Viveiros, M., Huggett, J. F., & Clark, T. G.  
782 (2019). Integrating informatics tools and portable sequencing technology for rapid detection  
783 of resistance to anti-tuberculous drugs. *Genome Medicine*, 11(1), 41.  
784 <https://doi.org/10.1186/s13073-019-0650-x>  
785  
786 Quispe, N., Asencios, L., Obregon, C., Velásquez, G. E., Mitnick, C. D., Lindeborg, M.,  
787 Jave, H., & Solari, L. (2020). The fourth national anti-tuberculosis drug resistance survey in  
788 Peru. *The International Journal of Tuberculosis and Lung Disease*, 24(2), 207–213.  
789 <https://doi.org/10.5588/ijtld.19.0186>  
790  
791 R Core Team (2021). R: A language and environment for statistical computing. R Foundation  
792 for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.  
793  
794 Rancoita, P. M. V., Cugnata, F., Gibertoni Cruz, A. L., Borroni, E., Hoosdally, S. J., Walker,  
795 T. M., Grazian, C., Davies, T. J., Peto, T. E. A., Crook, D. W., Fowler, P. W., & Cirillo, D.  
796 M. (2018). Validating a 14-Drug Microtiter Plate Containing Bedaquiline and Delamanid for  
797 Large-Scale Research Susceptibility Testing of Mycobacterium tuberculosis. *Antimicrobial*  
798 *Agents and Chemotherapy*, 62(9), e00344-18.  
799  
800 Schleusener, V., Köser, C. U., Beckert, P., Niemann, S., & Feuerriegel, S. (2017).  
801 Mycobacterium tuberculosis resistance prediction and lineage classification from genome  
802 sequencing: comparison of automated analysis tools. *Scientific Reports*, 7(1), 46327.  
803 <https://doi.org/10.1038/srep46327>  
804  
805 Schliep, K. P. (2010). Phangorn: Phylogenetic analysis in R. *Bioinformatics*, 27(4), 592–593.  
806 <https://doi.org/10.1093/bioinformatics/btq706>  
807  
808 Shitikov, E., Kolchenko, S., Mokrousov, I., Bespyatykh, J., Ischenko, D., Ilina, E., &  
809 Govorun, V. (2017). Evolutionary pathway analysis and unified classification of East Asian  
810 lineage of Mycobacterium tuberculosis. *Scientific Reports*, 7(1), 9227.  
811 <https://doi.org/10.1038/s41598-017-10018-5>  
812  
813 Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis  
814 of large phylogenies. *Bioinformatics*, 30(9), 1312–1313.  
815 <https://doi.org/10.1093/bioinformatics/btu033>  
816  
817 Tahseen, S., Qadeer, E., Khanzada, F. M., Rizvi, A. H., Dean, A., Deun, A. V., & Zignol, M.  
818 (2016). Use of Xpert ® MTB/RIF assay in the first national anti-tuberculosis drug resistance  
819 survey in Pakistan. *The International Journal of Tuberculosis and Lung Disease*, 20(4), 448–  
820 455. <https://doi.org/10.5588/ijtld.15.0645>  
821  
822 The CRyPTIC Consortium and the 100,000 Genomes Project (2018). Prediction of  
823 Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. *New England Journal*  
824 *of Medicine*, 379, 1403-1415.  
825  
826 Thai, P. V. K., Ha, D. T. M., Hanh, N. T., Day, J., Dunstan, S., Nhu, N. T. Q., Kiet, V. S.,  
827 Lan, N. H., Dung, N. H., Lan, N. T. N., Thuong, N. T., Lan, N. N., Liễu, P. T. T., Hồng, N.  
828 T., Diệp, Đ. C., Thanh, N. T. K., Hội, N. V., Nghĩa, N. V., Đại, T. N., & Minh, H. Q. (2018).

- 829 Bacterial risk factors for treatment failure and relapse among patients with isoniazid resistant  
830 tuberculosis. *BMC Infectious Diseases*, 18(1), 112.  
831
- 832 Thwaites, G. E., Bang, N. D., Dung, N. H., Quy, H. T., Oanh, D. T. T., Thoa, N. T. C., Hien,  
833 N. Q., Thuc, N. T., Hai, N. N., Lan, N. T. N., Lan, N. N., Duc, N. H., Tuan, V. N., Hiep, C.  
834 H., Chau, T. T. H., Mai, P. P., Dung, N. T., Stepniewska, K., White, N. J., & Hien, T. T.  
835 (2004). Dexamethasone for the Treatment of Tuberculous Meningitis in Adolescents and  
836 Adults. *New England Journal of Medicine*, 351(17), 1741–1751.  
837
- 838 Walker, T. M., Kohl, T. A., Omar, S. V., Hedge, J., Elias, C. D. O., Bradley, P., Iqbal, Z.,  
839 Feuerriegel, S., Niehaus, K. E., Wilson, D. J., Clifton, D. A., Kapatai, G., Ip, C. L. C.,  
840 Bowden, R., Drobniewski, F. A., Allix-Béguec, C., Gaudin, C., Parkhill, J., Diel, R., ... Peto,  
841 T. E. A. (2015). Whole-genome sequencing for prediction of Mycobacterium tuberculosis  
842 drug susceptibility and resistance: a retrospective cohort study. *The Lancet Infectious  
843 Diseases*, 15(10), 1193–1202. [https://doi.org/10.1016/S1473-3099\(15\)00062-6](https://doi.org/10.1016/S1473-3099(15)00062-6)  
844
- 845 Walker, T. M., Miotto, P., Köser, C. U., Fowler, P. W., Knaggs, J., Iqbal, Z., Hunt, M.,  
846 Chindelevitch, L., Farhat, M. R., Cirillo, D. M., Comas, I., Posey, J., Omar, S. V., Peto, T. E.,  
847 Suresh, A., Uplekar, S., Laurent, S., Colman, R. E., Nathanson, C.-M., ... Puyen, Z. M.  
848 (2022). The 2021 WHO catalogue of Mycobacterium tuberculosis complex mutations  
849 associated with drug resistance: a genotypic analysis. *The Lancet Microbe*, 3(4), e265–e273.  
850 [https://doi.org/10.1016/S2666-5247\(21\)00301-3](https://doi.org/10.1016/S2666-5247(21)00301-3)  
851
- 852 Wang, L.-G., Lam, T. T.-Y., Xu, S., Dai, Z., Zhou, L., Feng, T., Guo, P., Dunn, C. W., Jones,  
853 B. R., Bradley, T., Zhu, H., Guan, Y., Jiang, Y., & Yu, G. (2019). Treeio: An R package for  
854 phylogenetic tree input and output with richly annotated and associated data. *Molecular  
855 Biology and Evolution*, 37(2), 599–603. <https://doi.org/10.1093/molbev/msz240>  
856
- 857 Wollenberg, K. R., Desjardins, C. A., Zalutskaya, A., Slodovnikova, V., Oler, A. J.,  
858 Quiñones, M., Abeel, T., Chapman, S. B., Tartakovsky, M., Gabrielian, A., Hoffner, S.,  
859 Skrahin, A., Birren, B. W., Rosenthal, A., Skrahina, A., & Earl, A. M. (2017). Whole-  
860 Genome Sequencing of Mycobacterium tuberculosis Provides Insight into the Evolution and  
861 Genetic Composition of Drug-Resistant Tuberculosis in Belarus. *Journal of Clinical  
862 Microbiology*, 55(2), 457–469. <https://doi.org/10.1128/JCM.02116-16>  
863
- 864 World Health Organisation. (2018) Global Tuberculosis Report, 2018. Retrieved from:  
865 <https://apps.who.int/iris/handle/10665/274453>  
866
- 867 World Health Organisation. (2019) Global Tuberculosis Report, 2019. Retrieved from:  
868 <https://www.who.int/publications/i/item/9789241565714>  
869
- 870 World Health Organisation. (2021) Global Tuberculosis Report, 2021. Retrieved from:  
871 <https://www.who.int/publications/i/item/9789240037021>  
872
- 873 World Health Organisation. (2015) The end TB strategy. Retrieved from:  
874 <https://www.who.int/publications/i/item/WHO-HTM-TB-2015.19>  
875
- 876 World Health Organisation. (2021a) Catalogue of mutations in Mycobacterium tuberculosis  
877 complex and their association with drug resistance. Retrieved from:  
878 <https://www.who.int/publications/i/item/9789240028173>

879

880 World Health Organisation. (2021b) WHO global lists of high burden countries for  
881 tuberculosis (TB), TB/HIV and multidrug/rifampicin-resistant TB (MDR/RR-TB), 2021–  
882 2025. Retrieved from: [https://www.who.int/news/item/17-06-2021-who-releases-new-global-](https://www.who.int/news/item/17-06-2021-who-releases-new-global-lists-of-high-burden-countries-for-tb-hiv-associated-tb-and-drug-resistant-tb)  
883 [lists-of-high-burden-countries-for-tb-hiv-associated-tb-and-drug-resistant-tb](https://www.who.int/news/item/17-06-2021-who-releases-new-global-lists-of-high-burden-countries-for-tb-hiv-associated-tb-and-drug-resistant-tb)

884

885 World Health Organisation. (2016) The use of molecular line probe assay for the detection of  
886 resistance to isoniazid and rifampicin: policy update. Retrieved from:  
887 <https://apps.who.int/iris/handle/10665/250586>

888

889 Yang, C., Luo, T., Shen, X., Wu, J., Gan, M., Xu, P., Wu, Z., Lin, S., Tian, J., Liu, Q., Yuan,  
890 Z., Mei, J., DeRiemer, K., & Gao, Q. (2017). Transmission of multidrug-resistant  
891 Mycobacterium tuberculosis in Shanghai, China: a retrospective observational study using  
892 whole-genome sequencing and epidemiological investigation. *The Lancet Infectious*  
893 *Diseases*, 17(3), 275–284. [https://doi.org/10.1016/S1473-3099\(16\)30418-2](https://doi.org/10.1016/S1473-3099(16)30418-2)

894

895 Zhao, Y., Xu, S., Wang, L., Chin, D. P., Wang, S., Jiang, G., Xia, H., Zhou, Y., Li, Q., Ou,  
896 X., Pang, Y., Song, Y., Zhao, B., Zhang, H., He, G., Guo, J., & Wang, Y. (2012). National  
897 Survey of Drug-Resistant Tuberculosis in China. *New England Journal of Medicine*, 366(23),  
898 2161–2170. <https://doi.org/10.1056/NEJMoa1108789>

899

900 Zignol, M., Cabibbe, A. M., Dean, A. S., Glaziou, P., Alikhanova, N., Ama, C., Andres, S.,  
901 Barbova, A., Borbe-Reyes, A., Chin, D. P., Cirillo, D. M., Colvin, C., Dadu, A., Dreyer, A.,  
902 Driesen, M., Gilpin, C., Hasan, R., Hasan, Z., Hoffner, S., ... Raviglione, M. C. (2018).  
903 Genetic sequencing for surveillance of drug resistance in tuberculosis in highly endemic  
904 countries: a multi-country population-based surveillance study. *The Lancet Infectious*  
905 *Diseases*, 18(6), 675–683. [https://doi.org/10.1016/S1473-3099\(18\)30073-2](https://doi.org/10.1016/S1473-3099(18)30073-2)

906

907 Zignol, M., Dean, A. S., Falzon, D., van Gemert, W., Wright, A., van Deun, A., Portaels, F.,  
908 Laszlo, A., Espinal, M. A., Pablos-Méndez, A., Bloom, A., Aziz, M. A., Weyer, K.,  
909 Jaramillo, E., Nunn, P., Floyd, K., & Raviglione, M. C. (2016). Twenty Years of Global  
910 Surveillance of Antituberculosis-Drug Resistance. *New England Journal of Medicine*,  
911 375(11), 1081–1089. <https://doi.org/10.1056/NEJMSr1512438>

912

913

914

915

916

917

918

919

920

921

922

923

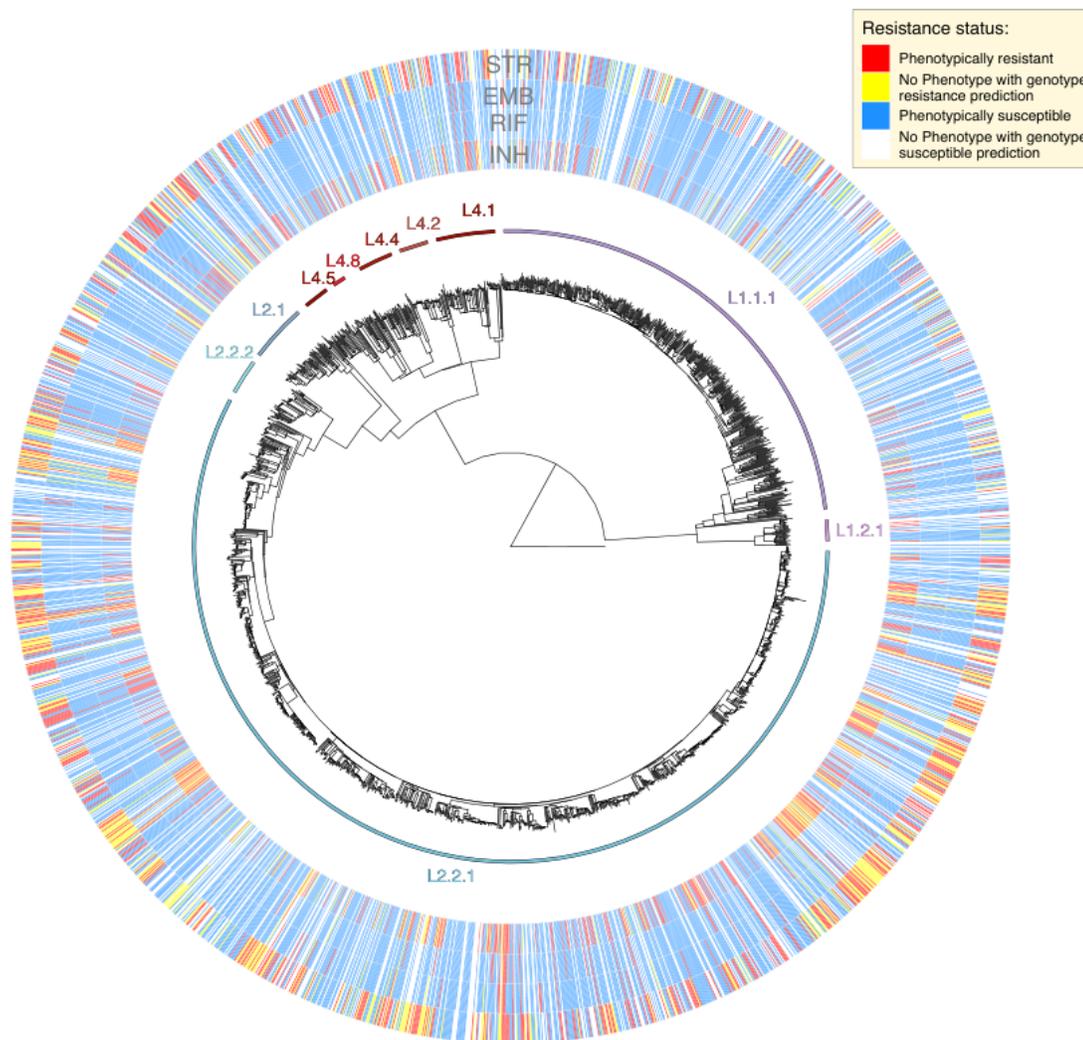
924

925

926

927

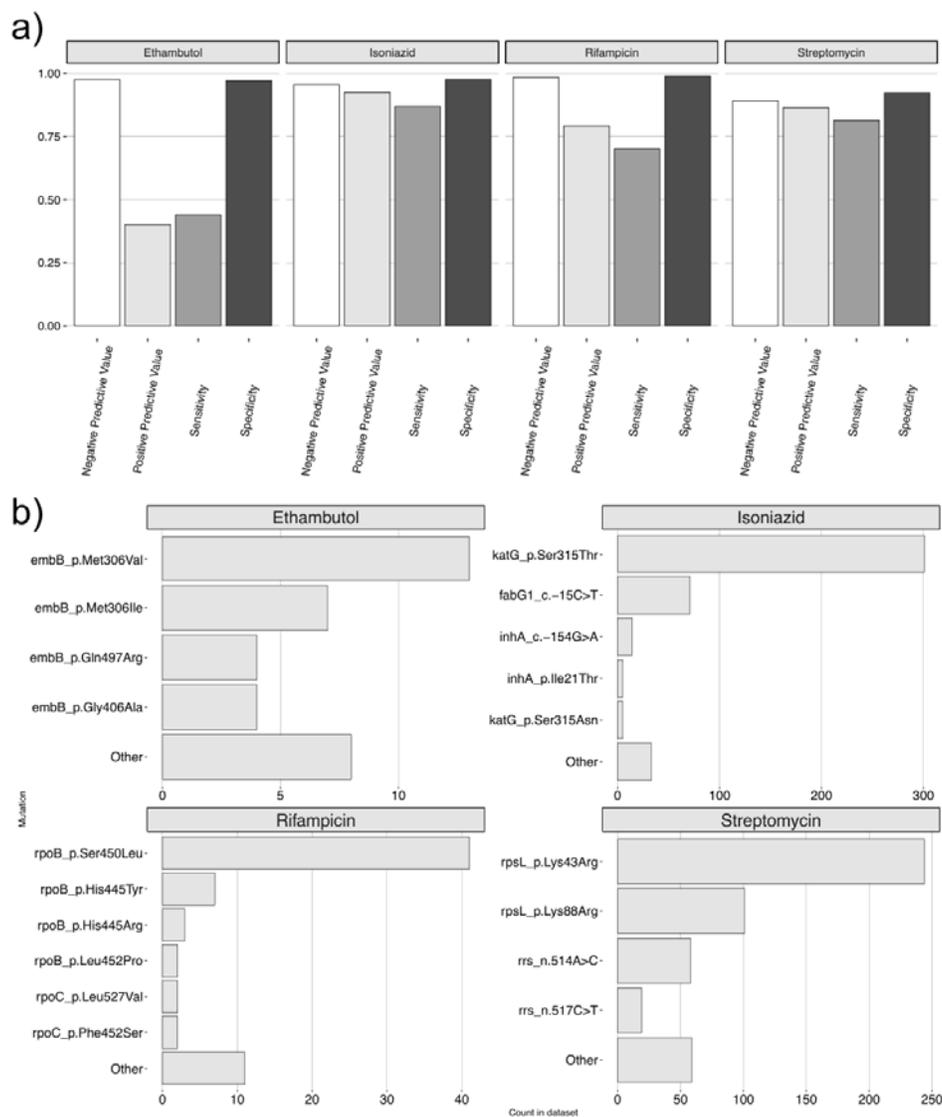
928



929

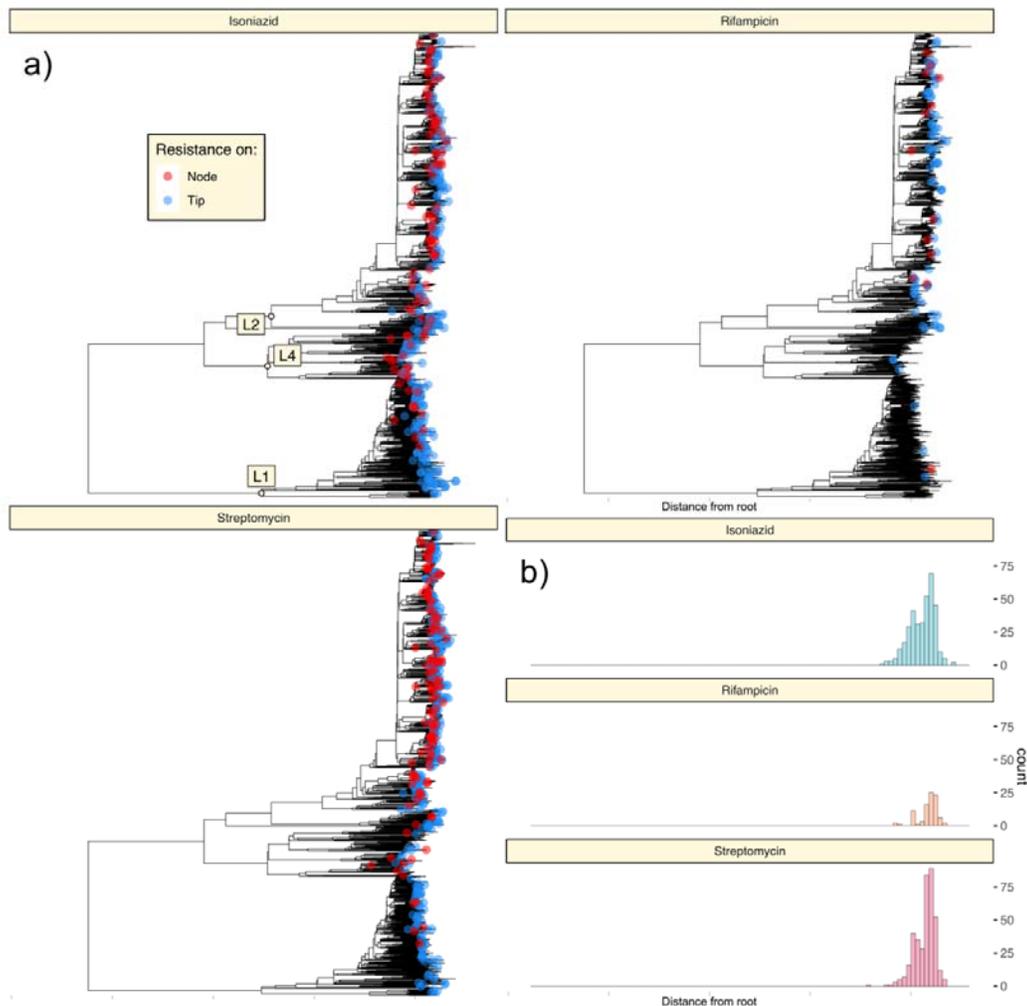
930 **Figure 1.** Phylogeny of 2,542 *Mtb* genomes, with associated phenotypic and genotypic drug  
931 resistance data for four first line drugs (moving outwards: INH, RIF, EMB and STR,  
932 respectively). For all samples with gold-standard phenotypic drug susceptibility testing  
933 (DST) data (N= 1,781), red denotes resistant isolates and blue denotes susceptible isolates.  
934 For all isolates without phenotypic DST data, yellow denotes genotypically predicted  
935 resistant isolates and white denotes genotypically predicted susceptible isolates. As noted  
936 elsewhere, phenotypic resistance and genotypic-based resistance prediction statuses were not  
937 perfectly correlated. Of all isolates predicted to be resistant using genotype data, 92.6, 79.2,  
938 40.2 and 86.5% were phenotypically resistant for INH, RIF, EMB, STR respectively (positive

939 predictive value; Figure 2). Of all isolates predicted to be susceptible using genotype data,  
 940 95.6, 98.5, 97.6 and 89.2% were phenotypically susceptible for these same four drugs  
 941 (negative predictive value; Figure 2).  
 942  
 943



944  
 945 **Figure 2. a)** Sensitivity, specificity, positive and negative predictive values for four first line  
 946 drugs inferred using TB-Profiler. **b)** The count of resistance conferring mutations/amino acid  
 947 substitutions in phenotypically resistant isolates.

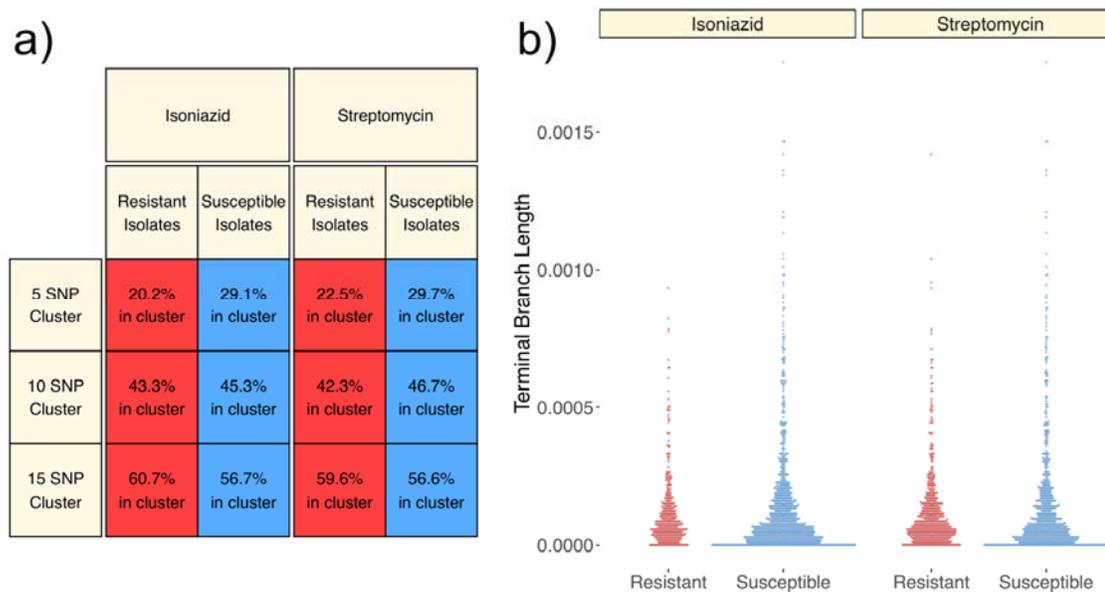
948



949

950 **Figure 3a).** Acquisition and transmission of mutations conferring resistance to three first line  
951 drugs: INH, RIF and STR. Biallelic SNPs present within the TB-Profiler catalogue were  
952 mapped onto nodes of the phylogeny using SNPPar and labelled according to whether they  
953 arose on a tip/terminal branch (blue circles), or interior branch (red circles). b) Histograms  
954 showing the distribution of resistance mutation emergence events relative to the root of the  
955 tree. X-axis scale of the histogram matches that of the above phylogeny.

956



957

958 **Figure 4. a)** Percentages of resistant and susceptible L2.2.1 isolates that belong to an  
 959 epidemiological cluster at 5, 10 and 15 SNP thresholds. Percentages were calculated for INH  
 960 and STR, the two drugs with the highest rates of resistance. **b)** Beeswarm plots showing the  
 961 distribution of terminal branch lengths of L2.2.1 isolates leading to tips with INH and STR  
 962 resistance mutations.

**Table 1. Lineage distribution by sex and clinical phenotype**

		Lineage						
		1	2.1	2.2.1	2.2 (Other than 2.2.1)	3	4	total
<b>Sex</b>	Male	496 (27.3%)	68 (3.7%)	1,014 (55.8%)	44 (2.4%)	1 (0.1%)	193 (10.6%)	1,816 (71.5%)
	Female	153 (21.1%)	18 (2.5%)	452 (62.3%)	19 (2.6%)	1 (0.2%)	82 (11.3%)	725 (28.5%)
<b>Clinical Phenotype</b>								
	PTB	581 (24.5%)	82 (3.50%)	1,389 (58.5%)	57 (2.4%)	2 (0.1%)	263 (11.1%)	2,374 (93.4%)
	TBM	68 (40.7%)	4 (2.4%)	77 (46.1%)	6 (3.6%)	-	12 (7.2%)	167 (6.6%)
	all TB	649 (25.5%)	86 (3.4%)	1,466 (57.7%)	63 (2.5%)	2 (0.1%)	275 (10.8%)	2,541

**Table 2 Antimicrobial resistance by lineage**

Drug resistance in PTB	1	2.1	2.2.1	2.2 (Other than 2.2.1)	4	total resistant	total tested
<b>MDR</b>							
Isoniazid/Rifampicin	7 (10.1%)	4 (5.8%)	53 (76.8%)	2 (2.9%)	3 (4.3%)	69 (3.9%)	1,786
<b>Antimicrobial (DST by either)</b>							
Isoniazid	74 (16.1%)	18 (3.9%)	311 (67.8%)	6 (1.3%)	50 (10.9%)	459 (25.7%)	1,786
Rifampicin	11 (12.6%)	5 (5.7%)	64 (73.6%)	2 (2.3%)	5 (5.7%)	87 (4.9%)	1,786
Ethambutol	7 (9.3%)	2 (2.7%)	58 (77.3%)	1 (1.3%)	7 (9.3%)	75 (4.2%)	1,785
<b>Antimicrobial (DST by MGIT)</b>							
Streptomycin	78 (13.6%)	29 (5.1%)	391 (68.2%)	11 (1.92%)	64 (11.2%)	573 (37.7%)	1,520
<b>Antimicrobial (DST by UKMYC5)</b>							
Bedaquiline (BDQ >=1)	-	-	-	1 (100%)	-	1 (0.4%)	267
Kanamycin (KAN >=2.5)	1 (6.7%)	2 (13.3%)	5 (33.3%)	1 (6.7%)	6 (40.0%)	15 (5.6%)	267
Ethionamide (ETH >=5)	12 (54.5%)	-	10 (45.5%)	-	-	22 (8.2%)	267
Amikacin (AMI >=1)	10 (24.4%)	1 (2.4%)	22 (53.7%)	4 (9.8%)	4 (9.8%)	41 (15.4%)	267
Levofloxacin (LEV >=1)	8 (25.8%)	1 (3.2%)	18 (58.1%)	2 (6.5%)	2 (6.5%)	31 (11.6%)	267
Moxifloxacin (MXF >=0.5)	10 (17.5%)	1 (1.8%)	33 (57.9%)	5 (8.8%)	8 (14.0%)	57 (21.3%)	267
Delamanid (DLM >= 0.12)	-	-	4 (80.0%)	1 (20.0%)	-	5 (1.9%)	267
Linezolid (LZD >=1)	14 (43.8%)	-	13 (40.6%)	-	5 (15.6%)	32 (12.0%)	267
Ceftazadime (CFZ >=1)	2 (28.6%)	-	3 (42.9%)	-	2 (28.6%)	7 (2.6%)	267
Rifabutin (RFB >=1)	-	1 (20.0%)	4 (80.0%)	-	-	5 (1.9%)	267
Para-aminosalicylic acid (PAS >=4)	31 (39.2%)	2 (2.5%)	36 (45.6%)	5 (6.3%)	5 (6.3%)	79 (29.6%)	267