bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

1 **Multi-clonal evolution of MDR/XDR *M. tuberculosis* in a high prevalence setting in Papua**
2 **New Guinea over three decades**

3 Arnold Bainomugisa[1,9]*, Evelyn Lavu[2]*, Stenard Hiashiri[3], Suman Majumdar[4], Alice
4 Honjepari[3], Rendi Moke[5], Paison Dakulala[5], Grant A. Hill-Cawthorne[6,7], Pandey Sushil[8], Ben J.
5 Marais[6,7], Christopher Coulter[8†], Lachlan J. M. Coin[9†¶]
6
7 [1]Faculty of Medicine, University of Queensland, Brisbane, Australia
8 [2]Central Public Health Laboratory, Port Moresby, Papua New Guinea
9 [3]Western Province Health office, Western Province, Papua New Guinea;
10 [4]Burnet Institute, Melbourne, Australia;
11 [5]National Department of Health, Port Moresby, Papua New Guinea
12 [6]School of Public Health, University of Sydney, Sydney, Australia
13 [7]Marie Bashir Institute for Infectious diseases and Biosecurity, University of Sydney, Sydney,
14 Australia;
15 [8]Queensland Mycobacterium Reference Laboratory, Pathology Queensland, Brisbane, Australia
16 [9]Institute for Molecular Biosciences, University of Queensland, Brisbane, Australia
17 * † Contributed equally
18 ¶Corresponding Author

19

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

20 **Abstract**

21 An outbreak of multi-drug resistant tuberculosis has been reported on Daru Island, Papua New
22 Guinea. The *Mycobacterium tuberculosis* strains driving this outbreak and the temporal accrual
23 of drug resistance mutations have not been described. We analyzed 100 isolates using whole
24 genome sequencing and found 95 belonged to a single modern Beijing strain cluster. Molecular
25 dating suggested acquisition of streptomycin and isoniazid resistance in the 1960s, with
26 virulence potentially enhanced by a *mycP1* mutation. The outbreak cluster demonstrated a high
27 degree of co-resistance between isoniazid and ethionamide (80/95; 84.2%) attributed to an *inhA*
28 promoter mutation combined with *inhA* and *ndh* coding mutations. Multidrug resistance (MDR),
29 observed in 78/95 samples, emerged with the acquisition of a typical *rpoB* mutation together
30 with a compensatory *rpoC* mutation in the 1980s. There was independent acquisition of
31 fluoroquinolone and aminoglycoside resistance; with evidence of local transmission of
32 extensively-drug resistant (XDR) strains from 2009. These findings underscore the importance of
33 whole-genome sequencing in informing an effective public health response to MDR/XDR *M.*
34 *tuberculosis.*

35

36

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

37 **Introduction**

38 Globally, an estimated 10.4 million cases of tuberculosis (TB) occurred in 2015 and
39 *Mycobacterium tuberculosis* was the leading single pathogen killer on the planet (1). The
40 emergence and spread of drug resistant *M. tuberculosis* strains pose a major challenge to global
41 TB control (2). Multi-drug resistant (MDR) TB, which is resistant to at least isoniazid and
42 rifampicin accounted for an estimated 480,000 new cases and 250, 000 deaths in 2015 (1).
43 Additional resistance to fluoroquinolones and second-line injectables defines extensively drug-
44 resistant (XDR) TB (3). In *M. tuberculosis*, drug resistance occurs mainly due to the
45 accumulation of chromosomal resistance conferring mutations without evidence of lateral gene
46 transfer (4). The emergence of drug resistance is dependent on the rate of acquisition of
47 resistance conferring mutations and the frequency with which these drug resistant strains are
48 transmitted within the community (5, 6). Compensatory mutations that limit the fitness cost
49 imposed by drug resistance may enhance the clonal spread of the most successful drug resistant
50 strains (6-8).

51 Genotyping of *M. tuberculosis* using 24-locus mycobacterial interspersed repetitive unit (MIRU-
52 24) facilitates TB outbreak investigation (9). However, MIRU-24 interrogate small genomic
53 regions that are susceptible to homoplasy and offers sub-optimal discriminatory power (10). The
54 discriminatory power of MIRU-24 is of particular concern with Beijing lineage strains (11).
55 Whole genome sequencing offers optimal resolution to explore local transmission dynamics
56 while also revealing the molecular mechanisms and evolution of drug resistance (6, 12, 13).

57 Papua New Guinea (PNG) has an estimated TB incidence of 432 per 100,000 population (1).
58 Daru Island, is one of the major hotspots for TB outbreaks and has an estimated incidence rate of
59 2600 per 100,000 population (14, 15). A recent national MDR TB survey showed that Daru

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

60 Island had the highest concentration of MDR TB cases in PNG (16) with estimates that 1·5% of
61 the Daru population are diagnosed with MDR TB every year (17) Recent Geo-spatial clustering
62 analysis of TB program data demonstrated evidence of primary transmission on Daru Island (18)
63 but this has not been confirmed by a detailed assessment of molecular epidemiology. The *M.*
64 *tuberculosis* strains driving the outbreak on Daru Island, as well as the associated drug resistance
65 mutations and temporal accrual of these mutations have not been described. This information is
66 urgently needed to guide patient management and optimal local public health responses.

67 **Results**

68 *Global phylogeny*

69 We characterized 100 strains from Daru, PNG (Figure 1) by MIRU-24 and whole genome
70 sequencing (Figure 2). MIRU-24 analysis revealed that 95 isolates formed a single dominant
71 outbreak cluster (Figure 3) with low allelic heterogeneity ($h$ <0·3) (Table S3). Analysis of SNPs
72 derived from whole genome sequencing revealed that the outbreak cluster formed a
73 monophyletic clade in the Beijing (East-Asian) lineage (modern Beijing sub-lineage), while all
74 the remaining strains belonged to the Euro-American lineage (Figure 4). According to Coll et al
75 classification (19), the outbreak cluster was revealed to be Beijing sub-lineage 2.2.1.1 (Table
76 S4). The Beijing outbreak cluster had a median of 23 differing SNPs between samples (range 0-
77 62), highlighting its limited genetic diversity (Figure S2). The Beijing outbreak cluster was
78 separated by at least 32 SNPs from the nearest neighboring modern Beijing genome included in
79 the representative phylogenetic tree (Table S5).

80

81

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

82 *Molecular dating and drug resistance*

83 Table 1 summarizes known and new putative drug resistance mutations detected within the
84 Beijing outbreak cluster, which were parsimoniously mapped to a molecular clock phylogenetic
85 tree that identified four clades; A (n=11), B (n=3), C (n=60) and D (n=21) (Figure 5). Strains in
86 clade A were the most distantly related to the rest of the outbreak cluster and accumulated the
87 least drug resistance mutations. Clade C demonstrated massive clonal expansion and included
88 the majority of the MDR strains (54 MDR and 6 XDR). Ten single nucleotide polymorphisms,
89 including a mutation in *mycP1* (p.Thr238Ala) differentiated clade B, C and D from clade A
90 (Table S6).

91 Strains in clades B, C and D demonstrated universal streptomycin resistance conferred by an
92 ancestral *rpsL* (p.Lys43Arg) mutation acquired in the 1960s. The strain ancestral to clade B, C
93 and D also acquired an *inhA* promoter mutation (*fabG1-inhA*, C-15T) associated with low-level
94 isoniazid resistance in the 1960s. The same *inhA* promoter mutation occurred independently in a
95 single clade A strain, while some (3/11) clade A strains acquired streptomycin resistance
96 conferred by an *rrs* (A514C) mutation. Clade C displayed universal high level isoniazid
97 resistance potentially conferred by an intragenic *inhA* mutation (p.Ile21Val) acquired in the
98 1980s. Occasional high level isoniazid resistance in clade B (1/3) and clade D (2/21) strains were
99 not associated with an *inhA* coding mutation while *katG* (p.Ser315Thr) mutation was detected in
100 only 2 clade A strains.

101 Of the 87 isoniazid resistant strains tested, 67 had high-level resistance, of which 64 (95·5%)
102 were ethionamide co-resistant; while 16 of remaining 20 (80%) had low-level isoniazid were
103 ethionamide co-resistant. Of the 84 strains tested for ethionamide resistance, 4 strains were
104 susceptible (2 with *katG* mutation but no *inhA* coding or promoter mutations). All strains with

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

105 co-resistance to at least low-level isoniazid and ethionamide had a *fabG1-inhA* mutation. We
106 explored the co-occurring mutations observed in these strains (Table 3). All strains with co-
107 occurring *fabG1-inhA*, *inhA* (p.Ile21Val) and *ndh* (del.G304) mutations were ethionamide
108 resistant, while 2/3 with both *fabG1-inhA* and *ndh* mutations were ethionamide resistant. No
109 other mutations associated with ethionamide resistance, such as *ethA* or *ethR* were observed.
110 Multi-drug resistance was first acquired by clade C in the 1980s with acquisition of a frequently
111 encountered *rpoB* mutation (p.Ser450Leu), together with a compensatory *rpoC* mutation
112 (p.Val483Gly). The ancestral strain in clade B acquired a different *rpoB* mutation (p.His445Arg)
113 in the 1990s without a compensatory mutation. One rifampicin resistant clade A strain acquired a
114 different *rpoC* compensatory mutation (p.Ile491Thr). The majority of clade D strains (17/21)
115 also acquired the p.Ser450Leu *rpoB* mutation, but at a later time point and only 1 strain had a
116 compensatory *rpoA* (p.Val183Gly) mutation. One rifampicin resistant clade D strain had two
117 *rpoB* mutations (p.Ser450Leu and p.Ile480Val); confirmed by sequence reads spanning both
118 mutations (Figure S3). In a single clade D strain, located in the middle of the *rpoB*
119 (p.Ser450Leu) clone, we could not identify the mutation despite adequate sequence coverage
120 (Figure S4).
121 There was good correlation between genotypic mutations and phenotypic drug susceptibility
122 determination (Table 1). However, only 50% (27/54) of the strains with putative ethambutol
123 resistance conferring mutations were phenotypically resistant at the critical concentration (5.0
124 mg/L); MIC testing to encompass lower concentrations of ethambutol was not performed. The
125 *embB* (p.Met306Val) mutation first occurred in clade C during the 1980s with independent
126 acquisition at later time points in clade A, C and D. Another *embB* mutation (p.Gln497Arg)
127 occurred independently on two occasions in clade C.

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

128 The majority of clade C (53/60; 88·3%) strains had pyrazinamide resistance mutations. The
129 earliest mutation *pncA* (p.Trp68Arg) was acquired in the 1990s. The only two pyrazinamide
130 resistant clade A strains shared a 3,990bp genomic deletion (position 2287064-2291054, Figure
131 S5) spanning four genes including *pncA*. MDR/XDR strains were more likely to have a *pncA*
132 mutation than non-MDR/XDR strains (56/84 versus 28/84, $P < 0.0001$).

133 *XDR TB genotypes*

134 Fluoroquinolone resistance due to mutations in the *gyrA* gene was detected in 11/60 (18%) clade
135 C strains; 6 were XDR TB (Table 1). A single four-member XDR clone emerged around 2009,
136 characterized by an additional capreomycin resistance mutation (*tlyA*, insertion C.397). One
137 XDR strain with phenotypic resistance to amikacin, kanamycin and capreomycin had both *rrs*
138 (G1484T) and *tlyA* (insertion397C) mutations, while another with an *rrs* (G1484A) mutation was
139 phenotypically only resistant to kanamycin and capreomycin. (Figure S6 and Table S7). We
140 assessed the dynamics of drug resistance transmission through enumerating the number of strains
141 that shared known resistance conferring SNPs to infer transmitted versus acquired drug
142 resistance. Primary drug resistance to isoniazid, rifampicin, ethambutol and streptomycin was
143 more dominant in clade B, C and D ($P < 0.0001$, Table 2). Clade C had additional primary
144 resistance to pyrazinamide and also displayed primary XDR-TB transmission.

145 **Discussion**

146 Our findings indicate that the large number of MDR TB cases detected on Daru Island in the
147 Western Province of Papua New Guinea is driven by the transmission of a highly drug resistant
148 cluster of modern Beijing lineage strain. Studies from other parts of Papua New Guinea
149 demonstrated a dominance of Euro-American lineage strains, (20, 21) although the contribution
150 of Beijing lineage strains may have been underestimated in these limited surveys. We cannot

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

151 comment on the geographic spread of this outbreak cluster, but phylogenetic analysis combined
152 with detailed molecular dating suggests that it has been in local circulation since the 1940s and
153 first acquired drug resistance mutations in the 1960s. The identification of four different clades
154 with distinct evolutionary trajectories suggests a 'permissive' environment for the *M.*
155 *tuberculosis* strains to acquire and spread drug resistance within the study setting.
156 Clade C was the most successful clade, acquiring resistance to all four first line drugs and
157 demonstrating clonal spread of both MDR and XDR strains. Phylogenetic analysis indicated that
158 isoniazid and streptomycin resistance was acquired by strains ancestral to clades B, C and D.
159 This is consistent with the use of non-rifampicin containing regimens in the 1960s that were
160 highly reliant on isoniazid and streptomycin (22). Similar analysis done in Russia and South
161 Africa where large MDR outbreaks have been recorded also indicated that streptomycin and
162 isoniazid resistance were acquired before the introduction of rifampicin containing regimens.
163 These settings mostly reported high-level isoniazid resistance due to *katG* mutations (6, 12). In a
164 recent multinational study, *katG* (p.Ser315Thr) was identified as the harbinger mutation that is
165 most frequently associated with future MDR occurrence and risks subsequent clonal spread (23).
166
167 Contrary to the experience in most other settings, (23, 24) this Beijing outbreak strain displayed
168 a signature *fabG1-inhA* (C-15T) mutation that confers low-level isoniazid resistance with
169 ethionamide co-resistance (25). However, on phenotypic drug susceptibility testing clade C
170 strains demonstrated universal high-level isoniazid resistance. This may be explained by the
171 accumulation of an additional *inhA* (p.Ile21Val) mutation. Such a double mutation may have
172 conferred high-level isoniazid resistance without the fitness cost associated with *katG* mutations,
173 supporting successful clonal expansion and the acquisition of resistance conferring mutations to

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

174 other anti-tuberculosis drugs (26). An additional consideration is the possibility that the *ndh*
175 frame shift mutation (deletion Glu102fs) observed only in clades B and C may also have
176 contributed, potentially as a compensatory mutation, since SNPs in the *ndh* gene have been
177 associated with increased intracellular NADH/NAD+ ratios hence competitive inhibition of
178 activated isoniazid (27).
179
180 The selective advantage of clade C would have been enhanced by the acquisition of a classic
181 *rpoB* (p.Ser450Leu) rifampicin resistance determining mutation together with a compensatory
182 *rpoC* (p.Val483Gly) mutation that would have abrogated its negative fitness effects (7). Similar
183 to previous observations, (28) one strain had two *rpoB* mutations of which p.Ile480Val, which
184 lies outside the 81bp rifampicin resistance determining region, may represent a compensatory
185 mutation. The single strain within the clade D *rpoB* (p.Ser450Leu) cluster, in which we were
186 unable to detect an *rpoB* mutation, may have been from a heteroresistant population since the
187 same *rpoB* mutation was detected in all the neighboring strains. Cultivation of the clinical
188 specimens probably favored selection of the observed strain (29).
189 An effective host immune response is critical to protect individuals against tuberculosis and to
190 limit ongoing transmission within communities. We identified a non-synonymous *mycp1*
191 (p.Thr238Ala) mutation that is ancestral to clades B, C and D. Mouse models used to identify *M.*
192 *tuberculosis* ESX-1 substrates and their effects on host cells found *mycP*1 proteins to be essential
193 for early macrophage replication and contribution to virulence by allowing escape of
194 mycobacteria from the phagosome into cytosol of infected macrophages (30, 31). Site directed
195 mutagenesis that inactivated *mycP*1 increased expression of ESAT-6. Altered host immune
196 responses effected by the putative *mycp*1 (p.Thr238Ala) mutation may have increased the

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

197 virulence and transmissibility of the outbreak cluster since the observed mutation is within the
198 active site (32), but this remains speculative since clade B demonstrated limited clonal
199 expansion.
200 Previous studies have found a strong correlation between MDR TB and resistance to
201 pyrazinamide or ethambutol, which highlights the problem of drug resistance amplification with
202 continued use of first-line treatment regimens in patients with undiagnosed MDR TB (33, 34).
203 The widespread use of the so-called retreatment regimen or category II posed particular
204 problems, since it added a single drug (streptomycin) to the standard first-line regimen in
205 patients who failed treatment or experienced a second tuberculosis episode. As per World Health
206 Organization (WHO) guidelines, the retreatment regimen was routinely used in the study setting,
207 which would have encouraged amplification of drug resistance given that streptomycin
208 resistance had been present since the 1960s. After a long delay the WHO has finally discouraged
209 the use of the retreatment regimen in its most recent treatment guidance (35).
210
211 There was good correlation between genotypic and phenotypic drug resistance profiles, except
212 for ethambutol. This reflects the complex genetic basis of ethambutol resistance, but also
213 emphasizes the highly variable results achieved with phenotypic drug susceptibility testing (13).
214 In our study, phenotypic susceptibility testing for ethambutol demonstrated huge variability,
215 especially when using the Bactec MGIT 960 automated system (36). Safi et al. demonstrated the
216 epistatic nature of ethambutol resistance by identifying that mutations in *embB* are often
217 accompanied by polymorphisms in other genes such as *nuoD* and *Rv3806c* that lead to a
218 progressive increase in the critical concentration (37). We could not identify any of the

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

219 accompanying mutations previously identified, but hypothesize that a reduction in the critical
220 concentration used may improve the correlation between resistance phenotype and genotype.
221 There is evidence of ongoing transmission of XDR strains characterized by a *tlyA* (ins.C397)
222 mutation that confers capreomycin resistance. One XDR strain demonstrated resistance to all
223 second line injectables, highlighting the ongoing need for timely access to new and repurposed
224 agents like bedaquiline, delamanid and linezolid for effective management of XDR cases. No
225 putative resistance mutations were detected which have been implicated in resistance to these
226 newer agents. Mutations at *rrs* position 1484 may affect aminoglycoside susceptibility
227 differently, since one XDR sample with an *rrs* (G1484T) mutation had phenotypic amikacin
228 resistance, while another with *rrs* (G1484A) mutation was susceptible to amikacin. In a
229 systematic review of studies that assessed the use of second line injectables, the *rrs* (G1484T)
230 mutation was described as a specific predictor of injectable drug resistance but not the *rrs*
231 (G1484A) mutation (38). The Hain MTBDRsl line probe assay recently endorsed by WHO for
232 rapid detection of second-line drug resistance has a probe for *rrs* G1484T but may detect
233 G1484A as a failure of wild type binding (39) and hence incorrectly infer resistance to all
234 second-line injectables; which could mislead treatment.
235
236 Our study is limited by its retrospective nature and the fact that clinical isolates tested failed to
237 capture the whole population with MDR TB on Daru Island which is thought to be the nexus of
238 TB transmission. Although this is one of the largest MDR outbreak clusters described to date, the
239 limited number of strains analyzed and the short specimen collection time frame may reduce the
240 accuracy of our time point estimates. However, our estimates have biological plausibility and
241 temporal consistency with the introduction of the respective drugs. Unfortunately a detailed

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

242 description of when specific drugs were first used in the study setting was unavailable. Finally,
243 we lacked clinical and epidemiological information to enhance our genomic analyses. This
244 information is critical to identify individual and population characteristics that facilitate ongoing
245 transmission of these drug resistant strains.
246 Whole genome sequencing confirmed a major outbreak of transmitted drug resistant tuberculosis
247 on Daru Island, Papua New Guinea, driven by a single Beijing strain cluster. A major concern is
248 further spread of the outbreak strain to adjacent geographical areas including the capital city,
249 Port Moresby may amplify transmission within Papua New Guinea. There is an urgent need to
250 improve early detection of drug resistant tuberculosis cases with linkage to effective care
251 programs in order to limit drug resistance amplification and terminate ongoing transmission.

252 **Methods**

253 We performed a retrospective assessment of all clinical isolates referred from Daru Island to the
254 Supra-National Referral Laboratory (SRL) in Brisbane, Australia from October 1, 2012 to March
255 15, 2015. MIRU-24 and whole genome sequencing were performed to understand the genetic
256 diversity of circulating strains, reveal the mutations associated with drug resistance and to
257 explore the evolution of these mutations. The study was approved by institutional review boards
258 of the University of Queensland, Australia (study number 2015000572) and PNG Medical
259 Research Advisory Committee (study number 16·42).

260 *Study setting and specimen selection*

261 Daru town located on Daru island (14·7$km^2$ in area) is the provincial capital of PNG's Western
262 Province and has an estimated population of around 15,000 people (40). Daru General Hospital
263 provides health care services to residents of Daru Island and surrounding mainland communities

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

264 in South Fly District. Clinical specimens are routinely referred to SRL for culture and sensitivity
265 testing. Criteria for referral is on detection of rifampicin resistance by Xpert MTB/RIF assay
266 (Cephid, Sunnyvale, California) on a concurrent sample or at the individual discretion of a
267 clinician.

268 *Study procedures*

269 DNA was extracted from cultures with confluent growth on Löwenstien-Jensen (LJ) slopes using
270 Spin column High pure PCR template prep kit (Roche Diagnostics, Pennsbury, Germany).
271 MIRU-24 genotyping was performed as per standard protocol (41) and GeneMapper® version 4.0
272 (Applied Biosystems, California, USA) was used for fragment analysis.
273 Phenotypic drug susceptibility to first-line drugs rifampicin (1·0 μg/ml), isoniazid (0·1 μg/ml;
274 low-level and 0·4 μg/ml; high-level), streptomycin (1·0 μg/ml), ethambutol (5·0 μg/ml) and
275 pyrazinamide (100 μg/ml) was performed based on the proportion method using the automated
276 Bactec Mycobacterial Growth Indicator Tube (MGIT) 960 system (Becton Dickinson, New
277 Jersey, USA). For MDR TB isolates, susceptibility to the second-line drugs amikacin (1·0
278 μg/ml), capreomycin (2·5 μg/ml), kanamycin (2·5 μg/ml), ethionamide (5·0 μg/ml), ofloxacin
279 (2·0 μg/ml), *p*-aminosalicylic acid (4·0 μg/ml) and cycloserine (50 μg/ml) was determined using
280 MGIT system (20, 42)

281 DNA for whole genome sequencing was isolated using a lysozyme-phenol chloroform based
282 method and DNA purified using QIAamp DNA mini prep kit (QIAGEN, Hilden, Germany).
283 Paired end libraries were prepared using Illumina Nextera® XT DNA library preparation kit and
284 sequenced using the Illumina MiSeq sequencing platform (San Diego, CA, USA) at Westmead
285 Institute for Medical Research and Australia Genome Research Facility (Sydney, Australia).
286 Raw reads were used to derive octal codes using *insilico* SpolPred (43) and spoligotypes inferred

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

287 using the international database (SpolDB4) (44). Paired end reads were quality checked using
288 FastQC v0·11·2 (45), trimmomatic v0·27 (46) was used to remove low quality base pairs (Phred
289 score<30) especially at 3' ends. Trimmed reads were mapped to the *H37Rv* reference genome
290 (GeneBank: NC_000962·3) using BWA-mem with default settings (47). The mean reference
291 coverage was 98·4% (range 96·4-99·8%) and mean high quality base coverage was 70·7X
292 (range 25-182X).

293 GATK UnifiedGenotyper was used to call SNPs and small indels (48). SNPs and small indels
294 with at least 10X read depth, 80% allele frequency and with at least 10bp difference between
295 neighboring SNPs/indels were retained using customized java and perl scripts. Indels of greater
296 than 4 mutations were excluded from the analysis unless occurring within a known drug
297 resistance associated gene. High quality SNPs and indels were annotated using SnpEff v4·1 (49)
298 and those in repetitive regions such as PE/PPE were excluded from analysis. The SNP allele
299 frequency spectrum was constructed to illustrate the allelic diversity (Figure S1). We
300 characterized mutations in known genes (including regulatory mutations) that confer resistance
301 to rifampicin, isoniazid, ethambutol, streptomycin, pyrazinamide, fluoroquinolones, amikacin,
302 capreomycin, kanamycin, ethionamide, *para*-aminosalicyclic acid, cycloserine, bedaquiline,
303 linezolid and delamanid according to a literature review (Table S1). Raw reads in the form of
304 FASTQ files were submitted to NCBI sequence read archive under the project file number
305 PRJNA385247.

306 *Allelic diversity, phylogeny and molecular dating*

307 The clonal structure from MIRU-24 was inferred using a minimum spanning tree algorithm
308 implemented in Bionumerics v6·7 (Applied Maths NV, Keistraat 120, Belgium). MIRU-24
309 profiles with a minimum allele variation of <3 were defined as a cluster. The Hunter-Gaston

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

310 discriminatory index (HGDI) was used to evaluate the allelic diversity (*h*) of the MIRU-24 locus
311 (50) We obtained 34 global representative genomes from collaborators and from previous studies
312 for phylogenetic analysis (19, 24) Concatenated SNP alignment was used to construct a
313 maximum likelihood phylogenetic tree using RAxML v7·4·2; GTRCAT model at 1000
314 bootstraps (51) and visualized using FigTree v1·4·2.

315 Molecular dating of the Beijing outbreak cluster was first performed by Beauti using an
316 alignment containing both invariable and variable sites, the resultant file was used as an input file
317 for BEAST v1·8·2 (52). Base substitution was modelled using Hasegawa-Kishino-Yano (HKY)
318 or General Time Reversible (GTR) model with an estimated base frequency and a gamma
319 distribution among site rate variation with four rate categories. The lognormal relaxed clock
320 (uncorrelated) model which assumes independent mutation rates on different branches, was used
321 (12, 13, 53, 54) The tree was calibrated using time of sample collection as tip dates for each
322 genome, specified in years before the present.

323 We used uniform prior distribution for all the trees using a mean mutation rate of 0·35
324 SNP/genome/year (12, 55-57) and compared the model performance under the different
325 demographic models by calculating the Bayes factors from marginal likelihood estimates
326 obtained from path sampling/stepping stone sampling (58). For each analysis, two independent
327 runs of 50 million steps of Markov chain Monte Carlo (MCMC) were performed, discarding
328 10% burn-in and drawing samples at every 5,000 steps. Three independent runs of the best

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

329 model were performed for consistency. Tracer was used to examine Markov chain convergence,
330 adequate mixing, chain length and effective sample size (ESS > 200). TreeAnnotator was used to
331 obtain the best supported topology under the maximum clade credibility method.

332 A constant demographic model that used GTR substitution and gamma distribution at 4
333 categories was found to be the best fit since it had a better marginal likelihood estimate
334 compared to the other models (Table S2). The overall mutation rate was estimated to be 0·36
335 SNP/genome/per year (95% High Posterior Density, HPD 0·24-0·48) which is consistent with
336 other published reports (55, 59, 60)

337 Resistance conferring mutations were parsimoniously mapped on the phylogenetic tree
338 (considering no reversion) and divergence time used to infer the likely timing of drug resistance
339 acquisition. The number of strains that shared known resistance conferring SNPs from the tree
340 nodes (Figure 5) were enumerated to infer primary drug resistance versus non-shared drug
341 resistance conferring SNPs (tree branches) to infer acquired drug resistance (6, 13, 23)· A
342 Fisher's exact test in R statistical package was used to assess the association of observed drug
343 resistance mutations and MDR/XDR phenotypes.

## 344 Author contribution statement

345 BA, CC, BM, LC, RM, SP, HS, DP, SM, AH and EL designed the study. BA, GC and SP did
346 laboratory work and whole genome sequencing. BA, EL, BM, CC, SP, GC and LC did analysis.
347 All coauthors reviewed and approved the final manuscript

## 348 Competing financial interests

349 We declare no competing interests

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

350 **Acknowledgement**

351 This study was funded by the National Health and Medical Research Council, Australia (grant
352 APP1044986) and AusAid (grant 01 HHISP-2013-0015). We thank the Papua New Guinea
353 National Tuberculosis Program, Provincial Health Staff (Western Province) and staff of
354 Queensland Mycobacterium Reference Laboratory for their assistance. Devika Ganesamoorthy,
355 Derek Benson, Nour Ben Zakour and Alex Outhred for their assistance with data analysis is
356 gratefully acknowledged. Associate Professor Vitali Sintchenko is thanked for his leadership in
357 pursuing WGS strategies in Australia as outlined in NHMRC grant APP1044986. We thank
358 Professor Matthew Cooper for discussions and interpretation of results.

359 **References**

360 1. World Health Organization., Global Tuberculosis Programme. Global tuberculosis control
361 : WHO report. Geneva: Global Tuberculosis Programme. p. 15 volumes.
362 2. Abubakar I, Zignol M, Falzon D, Raviglione M, Ditiu L, Masham S, et al. Drug-resistant
363 tuberculosis: time for visionary political leadership. The Lancet Infectious diseases.
364 2013;13(6):529-39.
365 3. Trauner A, Borrell S, Reither K, Gagneux S. Evolution of drug resistance in tuberculosis:
366 recent progress and implications for diagnosis and therapy. Drugs. 2014;74(10):1063-72.
367 4. Zhang Y, Yew WW. Mechanisms of drug resistance in Mycobacterium tuberculosis:
368 update 2015. The international journal of tuberculosis and lung disease : the official journal of the
369 International Union against Tuberculosis and Lung Disease. 2015;19(11):1276-89.
370 5. Ragonnet R, Trauer JM, Denholm JT, Marais BJ, McBryde ES. High rates of multidrug-
371 resistant and rifampicin-resistant tuberculosis among re-treatment cases: where do they come
372 from? BMC infectious diseases. 2017;17(1):36.

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

373 6. Casali N, Nikolayevskyy V, Balabanova Y, Harris SR, Ignatyeva O, Kontsevaya I, et al.
374 Evolution and transmission of drug-resistant tuberculosis in a Russian population. Nature genetics.
375 2014;46(3):279-86.

376 7. Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, et al. Whole-genome
377 sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory
378 mutations in RNA polymerase genes. Nature genetics. 2012;44(1):106-10.

379 8. Muller B, Chihota VN, Pillay M, Klopper M, Streicher EM, Coetzee G, et al.
380 Programmatically selected multidrug-resistant strains drive the emergence of extensively drug-
381 resistant tuberculosis in South Africa. PloS one. 2013;8(8):e70919.

382 9. Hanekom M, van der Spuy GD, Gey van Pittius NC, McEvoy CR, Hoek KG, Ndabambi
383 SL, et al. Discordance between mycobacterial interspersed repetitive-unit-variable-number
384 tandem-repeat typing and IS6110 restriction fragment length polymorphism genotyping for
385 analysis of Mycobacterium tuberculosis Beijing strains in a setting of high incidence of
386 tuberculosis. Journal of clinical microbiology. 2008;46(10):3338-45.

387 10. Comas I, Homolka S, Niemann S, Gagneux S. Genotyping of genetically monomorphic
388 bacteria: DNA sequencing in Mycobacterium tuberculosis highlights the limitations of current
389 methodologies. PloS one. 2009;4(11):e7815.

390 11. Gurjav U, Outhred AC, Jelfs P, McCallum N, Wang Q, Hill-Cawthorne GA, et al. Whole
391 Genome Sequencing Demonstrates Limited Transmission within Identified Mycobacterium
392 tuberculosis Clusters in New South Wales, Australia. PloS one. 2016;11(10):e0163612.

393 12. Cohen KA, Abeel T, Manson McGuire A, Desjardins CA, Munsamy V, Shea TP, et al.
394 Evolution of Extensively Drug-Resistant Tuberculosis over Four Decades: Whole Genome

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

395 Sequencing and Dating Analysis of Mycobacterium tuberculosis Isolates from KwaZulu-Natal.
396 PLoS medicine. 2015;12(9):e1001880.
397 13. Eldholm V, Monteserin J, Rieux A, Lopez B, Sobkowiak B, Ritacco V, et al. Four decades
398 of transmission of a multidrug-resistant Mycobacterium tuberculosis outbreak strain. Nature
399 communications. 2015;6:7119.
400 14. McBryde E. Evaluation of Risks of Tuberculosis in Western Province of PNG. Burnet
401 Institute (accessed 02, July 2014)2012. Available from: http://www.adi.org.au/wp-
402 content/uploads/2013/07/2012_BurnettInstitute_TB_RiskEvalWP.pdf.
403 15. Program NT. TB Report: Western Province, Papua New Guinea. National Tuberculosis
404 Program, 2016.
405 16. Aia P, Kal M, Lavu E, John LN, Johnson K, Coulter C, et al. The Burden of Drug-Resistant
406 Tuberculosis in Papua New Guinea: Results of a Large Population-Based Survey. PloS one.
407 2016;11(3):e0149806.
408 17. Furin J, Cox H. Outbreak of multidrug-resistant tuberculosis on Daru Island. The Lancet
409 Respiratory medicine. 2016.
410 18. Hiasihri S AP, John NL, et al, editor Primary transmission of drug resistant tuberculosis in
411 South Fly District, Western Province, Papua New Guinea. . 45th Union World Conference on Lung
412 Health; 2015 Dec 2–6; Cape Town, South Africa.
413 19. Coll F, McNerney R, Guerra-Assuncao JA, Glynn JR, Perdigao J, Viveiros M, et al. A
414 robust SNP barcode for typing Mycobacterium tuberculosis complex strains. Nature
415 communications. 2014;5:4812.

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

416 20. Ballif M, Harino P, Ley S, Coscolla M, Niemann S, Carter R, et al. Drug resistance-
417 conferring mutations in Mycobacterium tuberculosis from Madang, Papua New Guinea. BMC
418 microbiology. 2012;12:191.
419 21. Ley SD, Harino P, Vanuga K, Kamus R, Carter R, Coulter C, et al. Diversity of
420 Mycobacterium tuberculosis and drug resistance in different provinces of Papua New Guinea.
421 BMC microbiology. 2014;14:307.
422 22. Ley SD, Riley I, Beck HP. Tuberculosis in Papua New Guinea: from yesterday until today.
423 Microbes and infection / Institut Pasteur. 2014;16(8):607-14.
424 23. Manson AL, Cohen KA, Abeel T, Desjardins CA, Armstrong DT, Barry CE, 3rd, et al.
425 Genomic analysis of globally diverse Mycobacterium tuberculosis strains provides insights into
426 the emergence and spread of multidrug resistance. Nature genetics. 2017;49(3):395-402.
427 24. Guerra-Assuncao JA, Crampin AC, Houben RM, Mzembe T, Mallard K, Coll F, et al.
428 Large-scale whole genome sequencing of M. tuberculosis provides insights into transmission in a
429 high prevalence area. eLife. 2015;4.
430 25. Dominguez J, Boettger EC, Cirillo D, Cobelens F, Eisenach KD, Gagneux S, et al. Clinical
431 implications of molecular drug resistance testing for Mycobacterium tuberculosis: a
432 TBNET/RESIST-TB consensus statement. The international journal of tuberculosis and lung
433 disease : the official journal of the International Union against Tuberculosis and Lung Disease.
434 2016;20(1):24-42.
435 26. Ogbunugafor CB, Wylie CS, Diakite I, Weinreich DM, Hartl DL. Adaptive Landscape by
436 Environment Interactions Dictate Evolutionary Dynamics in Models of Drug Resistance. PLoS
437 computational biology. 2016;12(1):e1004710.

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

438 27. Lee AS, Teo AS, Wong SY. Novel mutations in ndh in isoniazid-resistant Mycobacterium
439 tuberculosis isolates. Antimicrobial agents and chemotherapy. 2001;45(7):2157-9.
440 28. Ramaswamy SV, Dou SJ, Rendon A, Yang Z, Cave MD, Graviss EA. Genotypic analysis
441 of multidrug-resistant Mycobacterium tuberculosis isolates from Monterrey, Mexico. Journal of
442 medical microbiology. 2004;53(Pt 2):107-13.
443 29. Rinder H, Mieskes KT, Loscher T. Heteroresistance in Mycobacterium tuberculosis. The
444 international journal of tuberculosis and lung disease : the official journal of the International
445 Union against Tuberculosis and Lung Disease. 2001;5(4):339-45.
446 30. Ohol YM, Goetz DH, Chan K, Shiloh MU, Craik CS, Cox JS. Mycobacterium tuberculosis
447 MycP1 protease plays a dual role in regulation of ESX-1 secretion and virulence. Cell host &
448 microbe. 2010;7(3):210-20.
449 31. Simeone R, Bobard A, Lippmann J, Bitter W, Majlessi L, Brosch R, et al. Phagosomal
450 rupture by Mycobacterium tuberculosis results in toxicity and host cell death. PLoS pathogens.
451 2012;8(2):e1002507.
452 32. Solomonson M, Huesgen PF, Wasney GA, Watanabe N, Gruninger RJ, Prehna G, et al.
453 Structure of the mycosin-1 protease from the mycobacterial ESX-1 protein type VII secretion
454 system. The Journal of biological chemistry. 2013;288(24):17782-90.
455 33. Hoek KG, Schaaf HS, Gey van Pittius NC, van Helden PD, Warren RM. Resistance to
456 pyrazinamide and ethambutol compromises MDR/XDR-TB treatment. South African medical
457 journal = Suid-Afrikaanse tydskrif vir geneeskunde. 2009;99(11):785-7.
458 34. Zhao LL, Sun Q, Liu HC, Wu XC, Xiao TY, Zhao XQ, et al. Analysis of embCAB
459 mutations associated with ethambutol resistance in multidrug-resistant mycobacterium
460 tuberculosis isolates from China. Antimicrobial agents and chemotherapy. 2015;59(4):2045-50.

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

461 35. World Health Organization. Guidelines for treatment of drug-susceptible tuberculosis and
462 patient care (2017 update). Geneva: World Health Organization.
463 http://www.who.int/tb/publications/2017/dstb_guidance_2017/en/ (accessed April 15, 2017),
464 2017.
465 36. Van Deun A, Wright A, Zignol M, Weyer K, Rieder HL. Drug susceptibility testing
466 proficiency in the network of supranational tuberculosis reference laboratories. The international
467 journal of tuberculosis and lung disease : the official journal of the International Union against
468 Tuberculosis and Lung Disease. 2011;15(1):116-24.
469 37. Safi H, Lingaraju S, Amin A, Kim S, Jones M, Holmes M, et al. Evolution of high-level
470 ethambutol-resistant tuberculosis through interacting mutations in decaprenylphosphoryl-beta-D-
471 arabinose biosynthetic and utilization pathway genes. Nature genetics. 2013;45(10):1190-7.
472 38. Georghiou SB, Magana M, Garfein RS, Catanzaro DG, Catanzaro A, Rodwell TC.
473 Evaluation of genetic mutations associated with Mycobacterium tuberculosis resistance to
474 amikacin, kanamycin and capreomycin: a systematic review. PloS one. 2012;7(3):e33275.
475 39. Organization WH. The use of molecular line probe assays for the detection of resistance to
476 second-line anti-tuberculosis drugs. Geneva, Switzerland: 2016.
477 40. Final Figures Papua New Guinea National Population and Housing Census 2011 [press
478 release]. 2011.
479 41. Supply P, Allix C, Lesjean S, Cardoso-Oelemann M, Rusch-Gerdes S, Willery E, et al.
480 Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-
481 number tandem repeat typing of Mycobacterium tuberculosis. Journal of clinical microbiology.
482 2006;44(12):4498-510.

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

483 42. Kruuner A, Yates MD, Drobniewski FA. Evaluation of MGIT 960-based antimicrobial
484 testing and determination of critical concentrations of first- and second-line antimicrobial drugs
485 with drug-resistant clinical strains of Mycobacterium tuberculosis. Journal of clinical
486 microbiology. 2006;44(3):811-8.
487 43. Coll F, Mallard K, Preston MD, Bentley S, Parkhill J, McNerney R, et al. SpolPred: rapid
488 and accurate prediction of Mycobacterium tuberculosis spoligotypes from short genomic
489 sequences. Bioinformatics. 2012;28(22):2991-3.
490 44. Demay C, Liens B, Burguiere T, Hill V, Couvin D, Millet J, et al. SITVITWEB--a publicly
491 available international multimarker database for studying Mycobacterium tuberculosis genetic
492 diversity and molecular epidemiology. Infection, genetics and evolution : journal of molecular
493 epidemiology and evolutionary genetics in infectious diseases. 2012;12(4):755-66.
494 45. FastQC: a quality control tool for high throughput sequence data. Available online at:
495 http://www.bioinformatics.babraham.ac.uk/projects/fastqc [Internet]. 2010.
496 46. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence
497 data. Bioinformatics. 2014;30(15):2114-20.
498 47. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
499 arXiv. 2013;:1303.3997v1 [q-bio.GN].
500 48. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The
501 Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA
502 sequencing data. Genome research. 2010;20(9):1297-303.
503 49. Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, et al. A program for
504 annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the
505 genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. 2012;6(2):80-92.

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

506 50. Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems:
507 an application of Simpson's index of diversity. Journal of clinical microbiology.
508 1988;26(11):2465-6.
509 51. Stamatakis A, Hoover P, Rougemont J. A rapid bootstrap algorithm for the RAxML Web
510 servers. Systematic biology. 2008;57(5):758-71.
511 52. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian phylogenetics with BEAUti
512 and the BEAST 1.7. Molecular biology and evolution. 2012;29(8):1969-73.
513 53. Merker M, Blin C, Mona S, Duforet-Frebourg N, Lecher S, Willery E, et al. Evolutionary
514 history and global spread of the Mycobacterium tuberculosis Beijing lineage. Nature genetics.
515 2015;47(3):242-9.
516 54. Luo T, Comas I, Luo D, Lu B, Wu J, Wei L, et al. Southern East Asian origin and
517 coexpansion of Mycobacterium tuberculosis Beijing family with Han Chinese. Proceedings of the
518 National Academy of Sciences of the United States of America. 2015;112(26):8136-41.
519 55. Ford CB, Shah RR, Maeda MK, Gagneux S, Murray MB, Cohen T, et al. Mycobacterium
520 tuberculosis mutation rate estimates from different lineages predict substantial differences in the
521 emergence of drug-resistant tuberculosis. Nature genetics. 2013;45(7):784-90.
522 56. Walker TM, Ip CL, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. Whole-genome
523 sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational
524 study. The Lancet Infectious diseases. 2013;13(2):137-46.
525 57. Bryant JM, Schurch AC, van Deutekom H, Harris SR, de Beer JL, de Jager V, et al.
526 Inferring patient to patient transmission of Mycobacterium tuberculosis from whole genome
527 sequencing data. BMC infectious diseases. 2013;13:110.

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

528 58. Baele G, Li WL, Drummond AJ, Suchard MA, Lemey P. Accurate model selection of
529 relaxed molecular clocks in bayesian phylogenetics. Molecular biology and evolution.
530 2013;30(2):239-43.
531 59. Ford CB, Lin PL, Chase MR, Shah RR, Iartchouk O, Galagan J, et al. Use of whole genome
532 sequencing to estimate the mutation rate of Mycobacterium tuberculosis during latent infection.
533 Nature genetics. 2011;43(5):482-6.
534 60. Merker M, Kohl TA, Roetzer A, Truebe L, Richter E, Rusch-Gerdes S, et al. Whole
535 genome sequencing reveals complex evolution patterns of multidrug-resistant Mycobacterium
536 tuberculosis Beijing strains in patients. PloS one. 2013;8(12):e82551.

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

537

538 Figure 1: A Map of Papua New Guinea illustrating the study location of Daru Island (Inset)

539

540 Daru town is the capital of South Fly district, western province of Papua New Guinea. The map
541 was obtained from http://d-maps.com

542

543

544

545

546

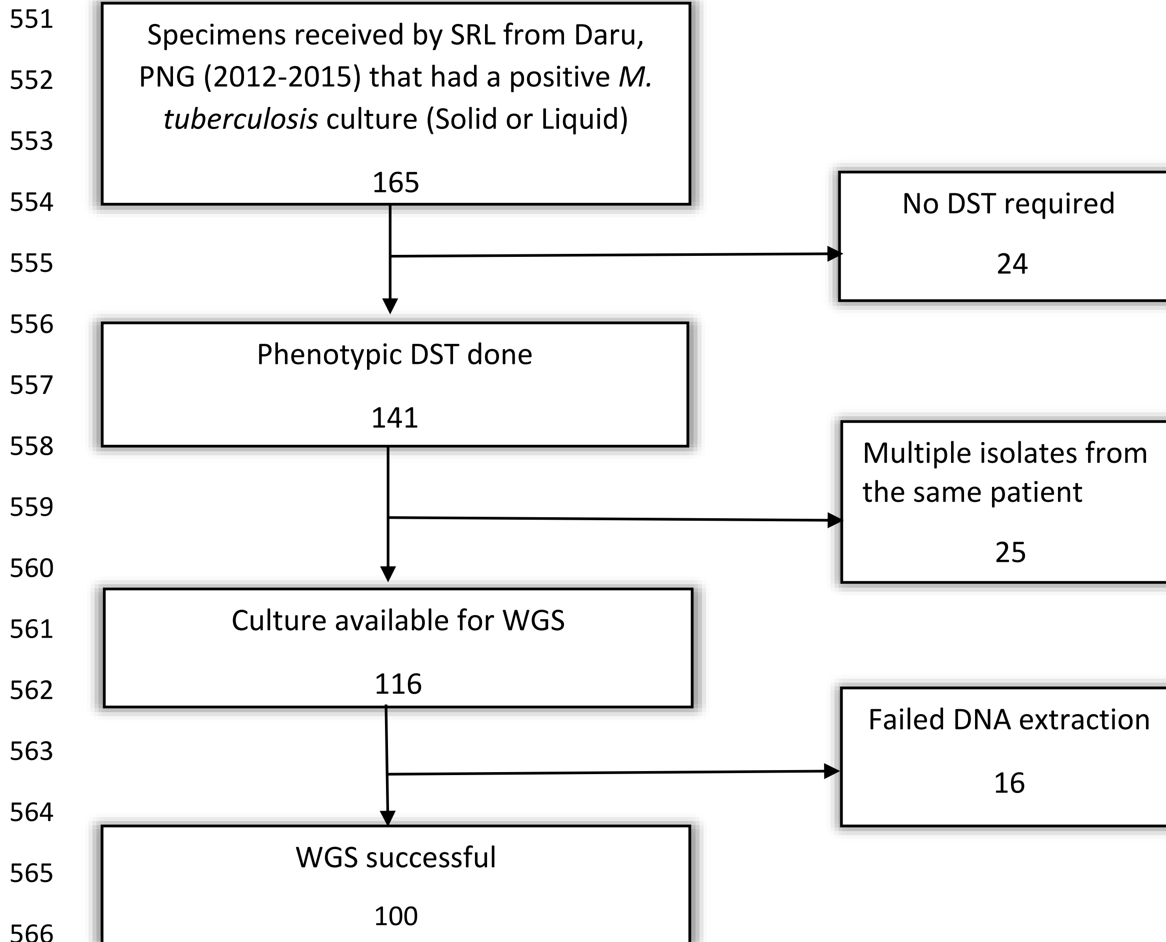
bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

547 Figure 2: A scheme of workflow to derive clinical isolates included in the study collected between 1/10/2012 to 15/3/2015

548

549 PNG – Papua New Guinea; SRL– Supra-National Reference Laboratory; DST – drug susceptibility test; WGS – whole genome sequencing

550

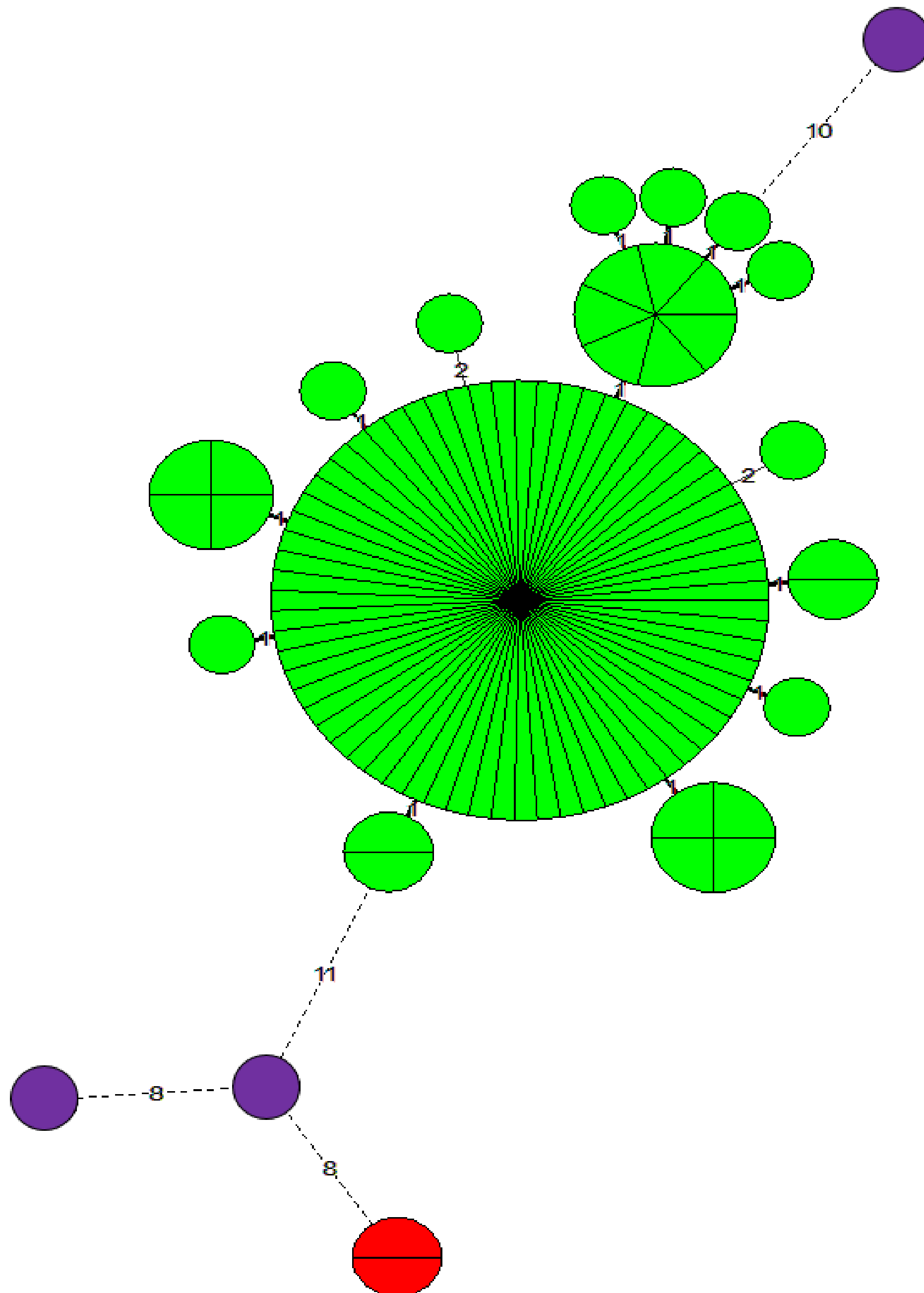



551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

574 Figure 3: MIRU-24 minimum spanning tree of *M. tuberculosis* isolates collected from Daru
575 Island, Papua New Guinea. Sectioned circles represent two or more isolates that share identical
576 allele profiles. Values on the branch*es* represent allelic difference between isolates. Green-major
577 outbreak cluster, Red-minor cluster, Purple-unique isolates

578




579

580

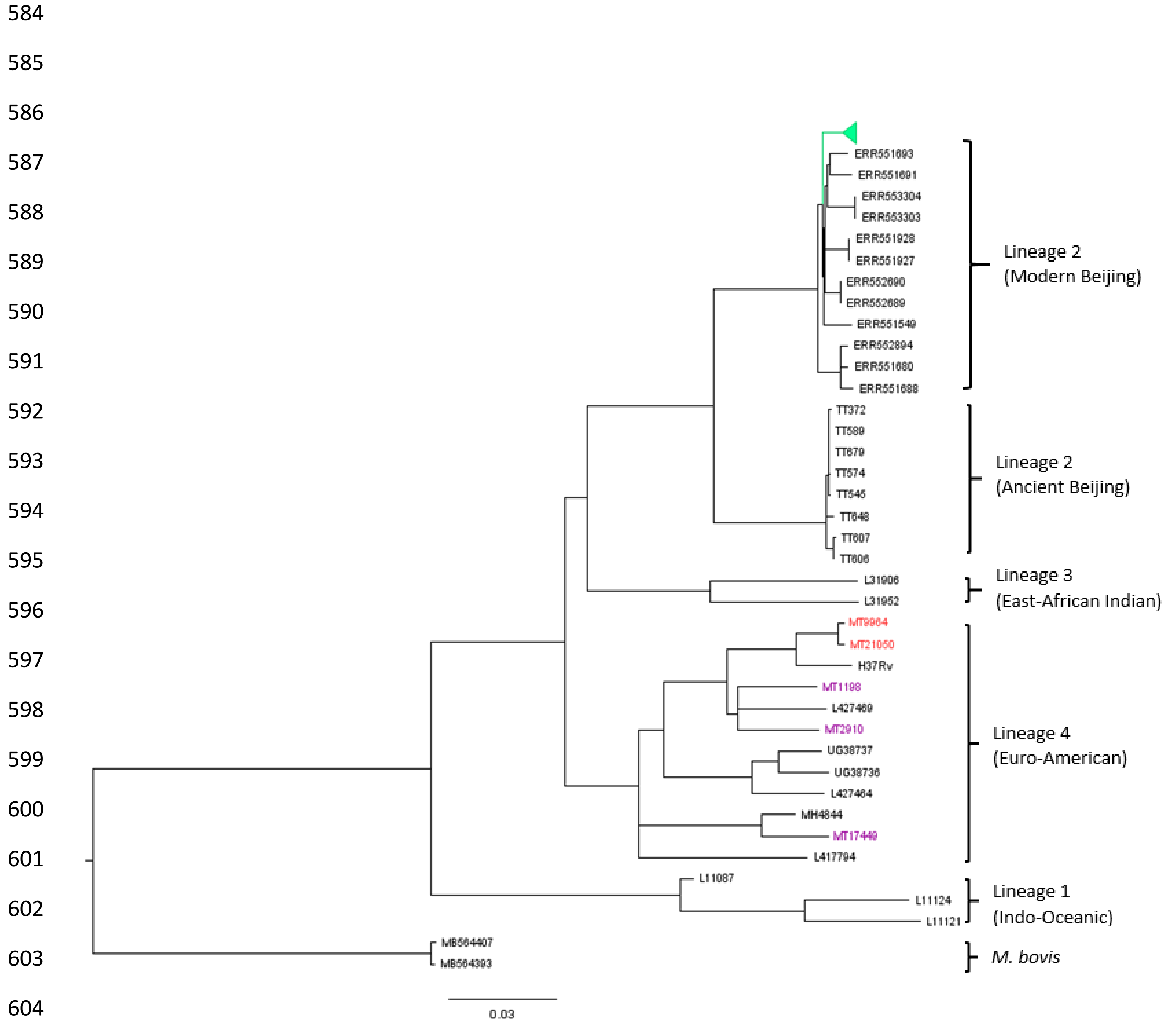
bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

581 Figure 4: Phylogenetic tree of *M. tuberculosis isolates* collected from Daru Island, Papua New Guinea,
582 together with global representative M. tuberculosis genomes. Tree constructed using 7282 single
583 nucleotide polymorphisms (using RAxML v.7.4.2) and rooted on *M. bovis*. The 95 members of the
"outbreak cluster" form a monophyletic clade (green) among the modern Beijing lineage while the colors
of the other isolates match those used in Figure 3

584

585

586



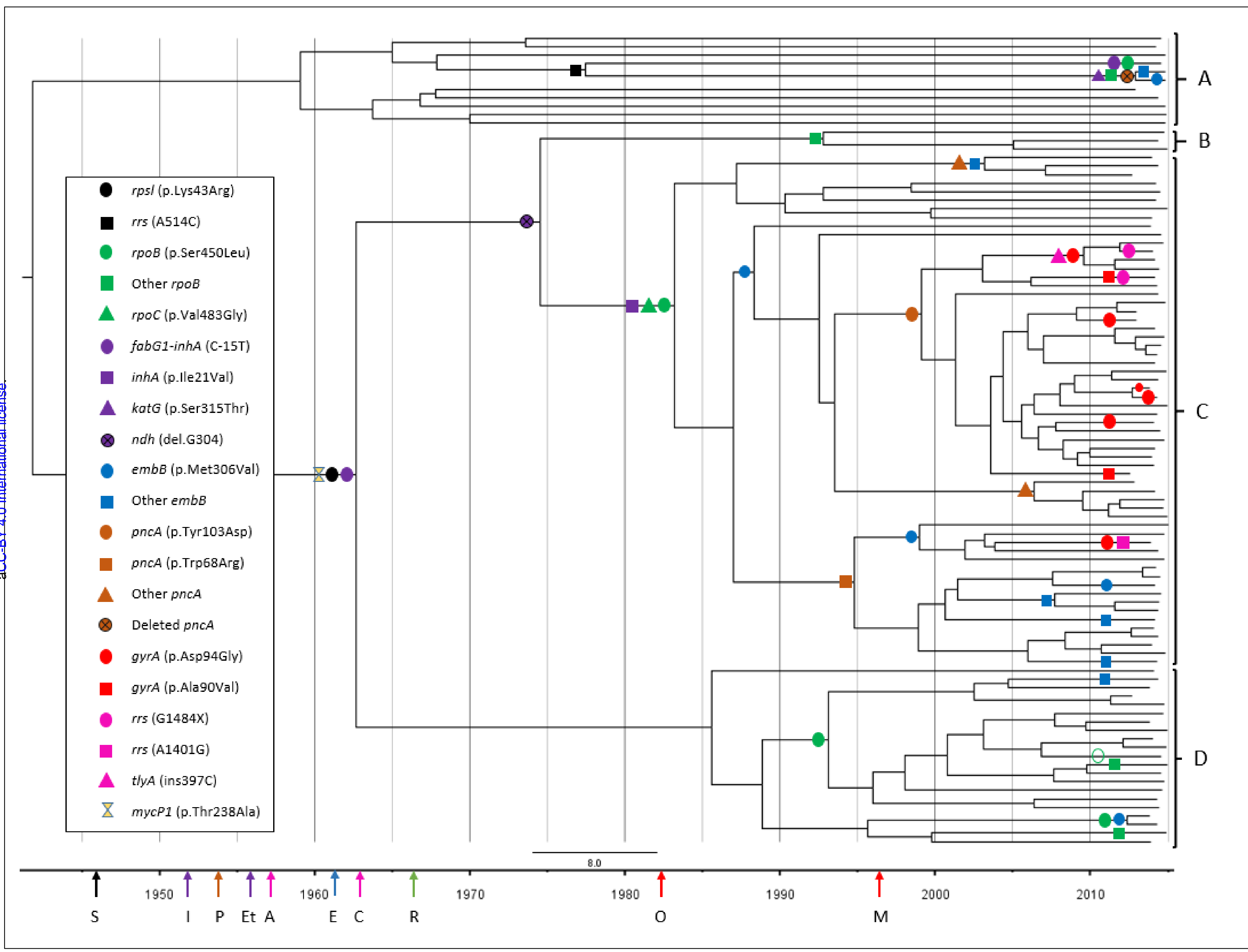

587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

609 Figure 5: Phylogeny of Beijing outbreak cluster strain with dated acquisition of drug resistance
610 mutations. Four clades (A, B, C and D) were recognizable. Evolutionary gain of resistance
611 conferring mutations and putative mutations for isoniazid (*ndh*) and virulent marker (*mycP1*);
612 parsimonious assigned onto the tree using colored shape with each representing a TB drug or class
613 of TB drug. The colored arrows represent the year of drug discovery; black-S (streptomycin, 1946),
614 green-R (rifampicin, 1966), purple-I (isoniazid, 1952) and Et (ethionamide, 1956), blue-E
615 (ethambutol, 1961), brown-P (pyrazinamide, 1954), red-O (ofloxacin, 1982) and M (moxifloxacin,
616 1996); pink-A (amikacin, 1957) and C (capreomycin, 1963). Small-red circle-*gyrA* (p.Asp94Ala);
617 hollow-green circle-no *rpoB* mutation

618

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

**Table 1:** Drug resistance mutations and phenotypic drug resistance observed among Beijing "outbreak cluster" strains

| TB drug | Gene | Mutation | Phenotypically resistant (N) | Phenotypically susceptible (N) | Critical concentration (µg/ml) |
|---|---|---|---|---|---|
| Rifampicin | *rpoB* | Ser450Leu* | 78 | 0 | 1 |
| | | Asp435Val | 2 | 0 | |
| | | Asp435Tyr | 2 | 0 | |
| | | His445Arg | 1 | 0 | |
| | | Ile480Val* | 1 | 0 | |
| | | His445Asp | 1 | 0 | |
| | *rpoC* | Val483Gly | 60 | 0 | |
| | | Ile491Thr | 1 | 0 | |
| | *rpoA* | Val183Gly | 1 | 0 | |
| Isoniazid | *inhA* | Ile21Val | 60 | 0 | 0.4 |
| | *fabG1-inhA* | C15T | 65 | 0 | 0.4 |
| | | | 20 | 0 | 0.1 |
| | *katG* | Ser315Thr | 2 | 0 | 0.4 |
| | *ndh* | del, G304** | 61 | 0 | 0.4 |
| | | | 2 | 0 | 0.1 |
| Ethambutol | *embB* | Met306Val* | 23 | 21 | 5 |
| | | Met306Ile | 1 | 1 | |
| | | Gly406Ala | 0 | 1 | |
| | | Gln497Arg* | 2 | 5 | |
| Pyrazinamide | *pncA* | Tyr103Asp | 27 | 1 | 100 |
| | | Trp68Arg | 17 | 0 | |
| | | Thr135Pro | 1 | 0 | |
| | | Gln10Pro | 3 | 0 | |
| | | Asp12Glu | 5 | 0 | |
| | | Del, gene** | 2 | 0 | |
| Streptomycin | *rpsL* | Lys43Arg | 84 | 0 | 8 |
| | *rrs* | A514C | 3 | 0 | |
| Ethionamide | *fabG1-inhA* | C15T | 80 | 2 | 10 |
| Quinolones | *gyrA* | Ala90Val | 2 | 0 | 2.5 |
| | | Asp94Ala | 1 | 0 | |
| | | Asp94Gly | 8 | 0 | |
| Aminoglycosides | *rrs* | A1401G | 1 (A,K,C) | 0 | 4 |
| | | G1484A | 1 (K,C) | 1 (A) | |
| | | G1484T | 1 (A,K,C) | 0 | |
| | *tlyA* | ins397C | 3 (C) | 1 (C) | |

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Aminoglycosides; A-amikacin, K-kanamycin, C-capreomycin. Isoniazid-all samples with phenotypic resistance had one of the mutations listed; Ethambutol─27/54 samples with one of the listed mutations were phenotypically susceptible. Known mutations to Bedaquiline, Delamanid, Linezolid and Clofazimine were not identified.

*Two dual mutations (Ile480Val and Ser450Val, and Gln497Arg and Met306Val)

**Putative mutations

**Table 2:** Frequency of clustered (transmitted) drug resistance mutations identified in different clades of the Beijing "outbreak cluster" strains*

| | | **Clade A** | | **Clade B** | **Clade C** | | **Clade D** | |
|---|---|---|---|---|---|---|---|---|
| | | Independent | Cluster | Cluster | Independent | Cluster | Independent | Cluster |
| **Locus** | *rpoB* | 1 | 2 | 3 | 0 | 60 | 3 | 16 |
| | *fabG1-inhA* | 1 | 0 | 3 | 0 | 60 | 0 | 21 |
| | *inhA* | 0 | 0 | 0 | 0 | 60 | 0 | 0 |
| | *katG* | 0 | 2 | 0 | 0 | 0 | 0 | 0 |
| | *rpsl* | 0 | 3 | 3 | 0 | 60 | 0 | 21 |
| | *embB* | 2 | 0 | 0 | 3 | 46 | 1 | 2 |
| | *pncA* | 0 | 2 | 0 | 0 | 53 | 0 | 0 |
| | *gyrA* | 0 | 0 | 0 | 7 | 4 | 0 | 0 |
| | *rrs* | 0 | 0 | 0 | 3 | 0 | 0 | 0 |
| | *tylA* | 0 | 0 | 0 | 0 | 4 | 0 | 0 |

*Inferred from figure 5

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

**Table 3**: Association of isoniazid and ethionamide phenotypes and common genes involved in resistance among Beijing "outbreak cluster" strains

| Gene combinations* | Total | High level INH resistant | Low level INH resistant | Ethionamide resistant | Ethionamide susceptible |
|---|---|---|---|---|---|
| *fabG1-inhA + inhA + ndh* | 60 | 60 | 0 | 60 | 0 |
| *fabG1-inhA +ndh* only | 3 | 1 | 2 | 2 | 1[#] |
| *fabG1-inhA* only | 22 | 4 | 18 | 18 | 1 |
| *katG* only | 2 | 2 | 0 | 0 | 2 |
| *inhA* only | 0 | 0 | 0 | 0 | 0 |
| *ethA* only | 0 | 0 | 0 | 0 | 0 |
| *ethr* only | 0 | 0 | 0 | 0 | 0 |

INH-isoniazid, ETH-ethionamide. *Observed mutations are in Table 1. [#]high level resistance to INH. A total of 64/67 strains with high-level INH were Ethionamide resistant

NB: Three strains (*fabG–inhA* only) with low-level isoniazid resistance were not tested against ethionamide as they were not MDR.

bioRxiv preprint doi: https://doi.org/10.1101/172601; this version posted August 4, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.