

1 Genotypic and phenotypic diversity of the multidrug-resistant
2 *Mycobacterium tuberculosis* strains from eastern India

3

4 **Authors:** Arup Ghosh^{1,2,#}, Himadri Bal^{3,#}, Viplov Kumar Biswas^{1,2,#}, Dasarathi Das³,
5 Sanghmitra Pati³, Sunil Raghav^{1,2*}

6

7 **Affiliations:**

- 8 1. Immunogenomics & Systems Biology group, Institute of Life Sciences (ILS),
9 Bhubaneswar, Odisha, India
- 10 2. School of Biotechnology, Kalinga Institute of Industrial Technology (KIIT), Bhubaneswar,
11 Odisha, India
- 12 3. Regional Medical Research Centre (RMRC), Indian Council of Medical research (ICMR),
13 Bhubaneswar, Odisha, India

14

15

16 [#]Equal contributing authors

17 *Corresponding author

18

19 Contact information: sunilraghav@ils.res.in, raghuvanshi2010@yahoo.co.in

20

21

22 **Abstract**

23

24 *Mycobacterium tuberculosis* (Mtb) poses a great challenge to human health and wellbeing
25 and hinders economic growth of a region. India along with other south east Asian countries
26 are known as high Tuberculosis burden countries. Adoption of whole genome sequencing in
27 studying genetic diversity, evolution, transmission pattern and drug resistance development
28 provided a great opportunity for developing and improving diagnostic and therapeutic
29 approaches. In our study we have sequenced 118 Mtb whole genome from North East(NE)
30 and Odisha as a representative of the diversity in eastern region of India for the first time.
31 We observed high prevalence of multi-drug resistant(MDR) lineage-2(n=52) strains in NE
32 whereas presence of mostly lineage-1(n=30) & 3 (n=11) strains in Odisha. The MDR strains
33 from Sikkim demonstrated similar resistance profile of fluroquinolones and pair-wise SNP
34 distances showed presence of local transmission clusters. We also detected significant
35 enrichment of short INDELs in MDR samples in contrast to drug susceptible samples. This
36 study provides molecular level insight into Mtb strains of eastern region in comparison with
37 Indian and global perspective.

38

39 **Introduction**

40 The COVID-19 pandemic has disrupted progress made for tuberculosis(TB) diagnosis in
41 recent years and reduced access to TB diagnosis and treatment. The most visible impact
42 observed in the global drop in the number of newly diagnosed and reported TB cases. A
43 sharp decline of 18% observed from the data collected up to June 2021 (1). WHO Global
44 Tuberculosis Report 2020 states that 26% of new TB cases are from India, which is the
45 highest number among high TB burden countries followed by China with 9% of new cases
46 reported. Although the number of new cases globally was lower than the 2017 report, there
47 was only a marginal change observed in new cases from India(2). Also, India has the
48 second-highest incidence of multidrug resistance (MDR) cases with the highest mortality rate
49 (2). Although the drug-susceptible cases had a higher treatment success rate, the MDR and
50 XDR TB cases had a treatment success rate of only 54% and 30%, respectively (3). The
51 major challenge with the current TB diagnostic strategies are the time required for culture-
52 based tests (3-6 weeks) and commercially available molecular diagnostics fail to account for
53 novel compensatory mutations leading to drug resistance development (4, 5).

54 From the previous whole-genome sequencing reports and molecular dating (spoligotyping)
55 studies it has been observed that the northern part of India is dominated by Lineage-3
56 whereas southern states show a prevalence of Lineage-1, central India has the presence of
57 CAS & EAI, eastern states show a mixed diversity of lineage and the north-eastern states
58 has a very dominant presence of Beijing lineage (6–9).

59 A genome-wide analysis study as a part of TB-ARC India project first time shed light on the
60 lesser-known genetic diversity of the Mtb strains in the southern part of India (5). This study
61 was conducted using 223 clinical isolates of which ~15% strains were resistant for at least
62 one drug when determined using phenotypic and genotypic drug susceptibility testing (DST)
63 methods. In other studies, also where a lesser number of drug-resistant strains were
64 investigated, the commercially available tests failed to detect the novel mutations present in
65 Indian strains thereby resulting in a false negative outcome (6). A whole-genome study from
66 The Foundation for Medical Research Mumbai also demonstrated the occurrence of novel

67 mutations associated with drug resistance development(8). Recently another study with 200
68 *Mycobacterium tuberculosis* (Mtb) whole-genome sequences from ICMR-Jalma consisting of
69 91 MDR samples collected from north India identified novel resistance-associated mutations
70 that are not used in any available molecular detection tests (7).
71 Whole-genome sequencing in recent times has provided us with a better understanding of
72 drug resistance phenotypes and genomic diversity of Mtb globally (10, 11). Prediction of
73 phenotype resistance utilizing genomic variants for some of the first-line drugs and widely
74 used second-line drugs have shown promising results (12–14). In the case of India, the
75 genomic surveillance of Mtb has been very limited and studies discussed earlier showed the
76 presence of novel resistance-associated genotypes that can impact therapeutic outcomes
77 (5, 7, 15). The current molecular tests used for rapid determination of resistance phenotypes
78 mostly represent genotypes observed in global Mtb strains and are mostly dominated by
79 European and American strains. In the case of India previous studies have shown that the
80 resistance phenotypes vary at the regional level based on the lineage prevalence. It has
81 been observed lineage 2 strains have a higher MDR rate in comparison to lineage 1 or 3
82 strains (11, 16, 17). Although whole genomic studies representing north and south India
83 have been undertaken in recent years, such detailed genomic profiles are missing for
84 eastern and north-eastern regions.

85
86 In our study, we have sequenced 118 culture-positive *M. tuberculosis* whole genomes which
87 include 7 follow-up samples collected from Sikkim, Meghalaya, and Odisha from 111
88 patients. The samples collected from the northeast were composed of MDR, XDR samples
89 as Fluoroquinolone resistance is highly prevalent in the region, whereas samples from
90 Odisha mostly consist of drug-susceptible samples. Through this study, we expanded the
91 understanding of the lineage diversity of the northeast and eastern region. We extensively
92 studied the genomic diversity of Fluoroquinolone resistant Lineage 2 strains along with
93 Lineage 1 and Lineage 3 strains of eastern India. We also examined the transmission
94 patterns of MDR strains in the northeast and their mutation acquisition patterns using follow-

95 up samples. Finally, we examined the performance of drug resistance phenotype prediction
96 using a known set of mutations.

97

98 **Results:**

99 In this study, we carried out whole-genome sequencing of 118 tuberculosis isolates cultured
100 using mycobacteria growth indicator tube (MGIT) from North East (n=70, including follow-
101 ups) and Odisha(n=48) representing north-east and east India collected between February
102 2017 to June 2020. Out of the 118 samples, there were 7 follow-up isolates belonging to five
103 patients collected between April 2018 and September 2019 from Sikkim and were removed
104 from the comparative analysis. Out of the remaining 111 samples, 9 samples didn't pass the
105 MTBC percentage filtration(>=80%) and pruned from further analysis.

106

107 The isolates collected from Sikkim and Meghalaya were all drug-resistant samples sent to
108 the RMRC-Bhubaneswar reference laboratory after the Cartridge Based Nucleic Acid
109 Amplification Test (CBNAAT) test showed presence of Rifampicin resistance associated
110 variant. Majority 51/54 of the isolates collected from the northeast were multi-drug resistant
111 (MDR) samples and 3/54 were classified as extensively drug-resistant (XDR). The samples
112 from Odisha were 39/48 drug-susceptible and except 9/48 samples out of which 2/48
113 classified as mono resistant and 7/48 as MDR.

114

115 The median age of patients enrolled in this study is 32.0 ± 17.01 years consisting of 57 Males,
116 44 Females sex, for 1 individual gender was not specified (Table -1, Supplementary Figure
117 1A). Information related to HIV status, smoking or alcohol consumption were not reported.
118 We further curated two major whole-genome datasets (n=423) published from India with
119 respective DST information (7, 18). For the South Indian samples, we collected the variants
120 call files and phenotypes from the PATRIC database, and North India cohort samples were
121 reanalysed following our sequence analysis criteria 471 (Sikkim-odisha n=102, curated
122 n=369) sequences from India were used for comparative analysis(19).

123

	<i>Count</i>	<i>Percentage (%)</i>
<i>Age (yr)</i>		
<30	41	40.20
>30	61	59.80
<i>Sex</i>		
<i>M</i>	57	55.88
<i>F</i>	44	43.14
<i>Undisclosed</i>	1	0.98
<i>Region</i>		
<i>North-East</i>	54	52.94
<i>East</i>	48	47.06
<i>Resistance (WGS)</i>		
<i>DS</i>	39	38.24
<i>MONO</i>	2	1.96
<i>MDR</i>	58	56.86
<i>XDR</i>	3	2.94
<i>Lineage</i>		
<i>Lineage-1</i>	31	30.39
<i>Lineage-2</i>	56	54.90
<i>Lineage-3</i>	12	11.76
<i>Lineage-4</i>	3	2.94

124

125 **Table 1:** Demographics of patients enrolled for the study, drug resistance phenotype of
126 sequenced samples (excluding follow-up samples) and lineage distribution (detected by
127 variant calls against published panels)

128 Using the QuantTB algorithm for mixed infection detection using whole-genome sequences
129 we observed similarity of NE and Odisha strains with other Indian or South East Asian
130 strains belonging to same lineages and only one strain detected per sample (20).

131

132 **Genome-wide variations shed light on *Mycobacterium tuberculosis* genetic diversity**
133 **in North-East and Odisha**

134 We detected a total of 12926 high-quality single nucleotide variants, 825 short indels with
135 lengths ranging from 29bp (max insertion) to 48bp (max deletion length) (Supplementary plot

136 1C) and a total of 2878 large deletions (Supplementary plot 1D). Effect of variants in
137 respective genes are predicted using snpEff tool (Figure 1D). Using the high-quality single
138 nucleotide variants we determined lineages for all the *Mtb* samples with Coll 2014 SNP
139 barcodes and fast lineage caller tool (21, 22). We observed a drastic difference between the
140 lineage distribution of *Mtb* in the northeast (NE) region and Odisha. Lineage-2 also known as
141 Beijing lineage has a very dominant presence in NE 62/64 whereas Lineage-1 also known
142 as Indo Oceanic lineage is more prominent 30/48 in Odisha (Figure 1B). The prevalence of
143 Beijing lineage in NE regions is coherent with a recent spoligotyping-based survey
144 conducted in Sikkim, the authors observed 62.41% occurrence of *Beijing* strains in the
145 region (23). Although lineage-1 is very dominant in Odisha the region also sees a prevalence
146 of lineage-3 11/48 also known as Central-Asian lineage followed by Beijing and lineage 4
147 (European/American) strain (Figure 1B). To further understand the most affected areas we
148 checked the distribution of lineages among different districts (Supplementary table1) of these
149 two states and found that most of the samples are from eastern region 51/64 in case of
150 Sikkim. The southern districts of Odisha are mostly affected by lineage-1 whereas central
151 and eastern districts show a mixed presence of all three lineages. As there were no previous
152 large scale genomic or spoligotyping surveys were not available from Odisha we assume
153 that the genetic diversity of *Mtb* in this region is heavily influenced by neighbouring regions a
154 high prevalence of lineage 1,2 and 4 (6, 7).
155 To understand the genetic diversity in context with other strains prevalent in India we
156 constructed phylogenetic tree of all 102 samples from our study along with two previously
157 published studies depicting the lineage diversity in south and north India.
158

159 **Phylogenetic classification and genetic similarity with other regional and global
160 strains**

161 Single nucleotide variants from 471 samples consisting of previously published and newly
162 sequenced *Mtb* whole genomes from India were used to create neighbor-joining
163 phylogenetic tree with 1000 bootstraps(24). The tree represents the genetic diversity of *Mtb*

164 across India represented by their geographical location and drug resistance phenotypes
165 (Figure 2). The drug resistance *Mtb* samples from Sikkim were divided into two sub-clusters
166 while coinciding with other lineage 2 samples from Tamil Nadu and Uttar Pradesh. One of
167 the Sikkim sub-clusters branches show very small branch lengths from the nearest node,
168 indicating presence of local transmission from a most recent common ancestor (MRCA) of
169 the strain (Figure 2).

170 The samples from Odisha were divided between lineage 1 and lineage 3 clusters and
171 supporting our previous hypothesis they grouped with representative strains from North and
172 South India respectively (Figure 2). The lineage 4 samples from Odisha clustered with
173 similar strains from Uttar Pradesh and Rajasthan (Figure 2A).

174

175 To further extend the understanding of genetic diversity of the strains in global context we
176 performed principle component analysis(PCA) of the 471 samples along with a set of 2500
177 global *Mtb* samples curated from GMTV dataset (25). Using principal component 1 and 2
178 ~87% of variance was explained (Figure 3A). The strains from India grouped among the
179 global lineages based on their respective lineages but showed some deviation in case of
180 lineage 2 samples from Sikkim and lineage 3 samples from North India (Figure 3A).

181

182 The two clusters of drug-resistant *Mtb* samples from Sikkim observed in the phylogenetic
183 clustering of the samples, indicated the presence of local transmission clusters in the region
184 and to further investigate we utilized SNP distance based methods.

185

186 **Local transmission of drug resistance Beijing strain in Sikkim**
187

188 In accordance with our findings from phylogenetic tree we calculated pair-wise distance
189 among all the sequenced samples from our study (n=102) and plotted the distribution of
190 pairwise SNP distances. We observed two group of lineage 2 sequences (n=42,9) from
191 Sikkim showing pairwise SNP distance of less than 50 in contrast to other strains showing at

192 least 100 or more dissimilar SNPs among them (Figure 2B). We clustered the pairwise SNP
193 information using a hierarchical clustering method and observed two distinct clonal clusters of
194 drug-resistant samples collected from Sikkim (Supplementary Figure 2). To further resolve
195 the genetic similarity among the strains we created a phylogenetic tree of all 53 samples
196 from the region and included their district level information. The resultant phylogenetic tree
197 represented the main cluster (n=42) consisting strains mostly from east Sikkim with 3/42
198 strains each from north and west Sikkim. The smaller cluster (n=9) also represented
199 sequences from east district along with 3/9 strains from west district (Figure 3C).

200

201 **Drug resistance phenotype of North East Indian strains matches with global Beijing**
202 **strains with high Fluoroquinolone resistance burden**

203 One of the main goals of our study was to understand the genomic variations leading to high
204 fluoroquinolone (namely moxifloxacin(both low and high concentrations), levofloxacin, and
205 ofloxacin) in the state of Sikkim. The abundance of multi-drug resistance strains among
206 Lineage 2(Beijing) lineage is very high. Out of 53 drug-resistant samples collected from
207 Sikkim 51 (96.22%) belonged to Lineage 2 (sublineage 2.2.1) of which 48 (94.11%)
208 belonged to the multi-drug resistance category and 3 (5.88%) in the extensively drug-
209 resistant category. As per the scope of the national tuberculosis surveillance program, only a
210 limited number of drugs were tested for DST but with the help of whole-genome sequencing
211 data, we predicted the phenotypes for 11 drugs used for the treatment of tuberculosis. We
212 observed a high prevalence of rifampicin resistance followed by isoniazid, ethambutol,
213 streptomycin, and fluoroquinolone group of drugs in Sikkim. On the other hand in Odisha
214 isoniazid resistance is more prominent in comparison to other groups of drugs.

		Ethambutol	Isoniazid	Kanamycin	Moxifloxa cin	Ofloxacin	Rifampicin	Streptomycin
Meghalaya	R	1	1	0	1	1	1	1
Odisha		3	9	0	4	4	7	6
Sikkim		51	51	3	50	50	52	51

Meghalaya	S	0	0	1	0	0	0	0
Odisha		45	39	48	44	44	41	42
Sikkim		2	2	50	3	3	1	2

215

216 **Table 2:** Summary of drug resistance phenotypes for seven drugs predicted by Mikrobe
217 predictor tool(26), grouped by place of sample collection. R = Drug resistance & S = Drug
218 susceptible

219

220 We compared the predicted results with our first-line and second-line LPA assay data and
221 two culture-based DST results. In the case of rifampicin the prediction was able to explain
222 101/102 cases, but in the case of kanamycin and capreomycin resistance there were 6/99
223 unexplained cases. In the fluoroquinolone category the results overlapped in 95/98 cases,
224 thus mostly explained by known genomic changes. To get a better perspective of variant
225 accumulation loci we ranked the genes based on the number of non-synonymous mutations
226 present in the gene body followed by functional segregation. The top biological processes
227 came up as metabolism and respiration, information pathway, lipid metabolism, cell wall-
228 related process, and virulence (Figure 4A). All of these pathways are known to be directly
229 associated with drug resistance phenotypes and the changes in virulence-related genes
230 might have a role in the increased pathogenicity of Beijing lineage, which is overrepresented
231 in our dataset (Figure 4A)(27). When we looked into genes with highest number of non-
232 synonymous variations acquired, genes from RNA polymerase and DNA gyrase coding
233 family ranked in the top position (Figure 1B). Followed by our observation we checked for
234 the presence of known drug resistance associated loci and their frequencies (count), the
235 results were coherent with prediction outcomes (Figure 4C). Although the functional impact if
236 INDELS has been less studied in the context of drug resistance association and other
237 functional implication, number of INDELS detected in MDR samples were significantly higher
238 in contrast with drug susceptible sample and the XDR samples also show an increased

239 pattern (Figure 4D). The MONO resistant sequences only had slight increase in the INDEL
240 number in comparison with susceptible sequences.

241

242 **In-host pathogen dynamics in contrast to resistance-associated loci**

243 During the process of samples collection we also followed up with six patients from Sikkim.
244 We collected a total of 7 isolates from five patients in an interval of 10 days to 1.5 years to
245 examine if there is presence of in-host variations. Genetic variations fixed during the while
246 the treatment regimen is ongoing sheds light on clonal selection of drug resistance Mtb
247 strains and recently has been shown to have impact on host innate immunity and
248 metabolism(28). The pairwise SNP distance among the same follow-up patients ranged
249 between 69-105 (Median: 89.50) indicating that despite slow mutation rates (0.3-0.6
250 substitutions/per year) and lack of recombination some (22).To elucidate the presence of in-
251 host variation overtime we grouped the follow up sequences with the primary strain and
252 plotted the distribution of allele fractions (fraction of reads supporting an allele) against depth
253 of coverage of the region (Figure 4E). Trendline in the data showed increased fractional
254 representation of some of the variant positions although when we compared the allele
255 fractions of known resistance associated loci, they were fixed (AF = 1) throughout the
256 timeline (Figure 4E).

257

258 **Discussion:**

259 The focus of our study was to explore the genetic diversity of *M. tuberculosis* complex in
260 North East and Odisha where no prior whole-genome sequencing studies have been
261 undertaken. For the first time, our study shares insight about Mtb lineages in the state of
262 Odisha and their genetic diversity. In addition, our study provides insight into the genetic
263 basis of the high prevalence of multi-drug resistance Beijing strains, local transmission of
264 fluoroquinolones resistant strains, and acquisition of resistance-associated mutations in
265 North-Eastern states represented by Sikkim.

266 We observed a high occurrence of a localized patient-to-patient transmission pattern of MDR
267 TB Beijing strains (<35 SNPs) in North East India. The contact-based transmission of MDR
268 strains might be one of the reasons for the high abundance of fluoroquinolone resistance
269 along with rifampicin resistant cases in the region. The cluster in East Sikkim indicates the
270 probable presence of nosocomial transmission of Tuberculosis in the region, which might
271 pose challenges to the ongoing mitigation programs and required to be taken into account
272 for prevention programs (17). But Due to the lack of sufficient metadata and number of
273 samples we were not able to determine whether the transmissions events took place in a
274 healthcare centre or region-specific manner.

275 The North East region is known to have a very high burden of drug-resistant tuberculosis
276 cases and a recent study accessing molecular diversity of MTBC in the regions showed a
277 high prevalence of Beijing strains (~62%) and higher occurrence of MDR phenotypes in the
278 strain(29.9%) in comparison to other MTBC lineages(4.7%) (12, 18). Our observation is
279 consistent with the previous findings additionally we observed an increased burden of
280 fluoroquinolone resistance in the region as well (12). Although our isolates from Sikkim were
281 enriched for rifampicin (CBNAAT) resistance phenotype the concurrent findings suggest our
282 results are representative of NE region the dominance of Beijing lineage and multi-drug
283 resistance (12).

284 Follow up samples collected from Sikkim in a period of ~1.5 years depicted change in allele
285 frequency of strains accumulated over the period of sample collection. With the limited
286 number of samples we were not able to capture any resistance associated changes. But
287 such studies can provide insight in to the initial acquisition of fluroquinolone resistance
288 burden in the region as amplification of fluroquinolones and second-line injectables occurred
289 in a very recent time(29).

290 On the other hand in Odisha we observe a very diverse representation of Mtb strains, all
291 three lineages prevalent in India observed in this region. Lineage 1 is the dominant followed
292 by lineage 3 and 2. The distribution of lineage has a similar pattern as the study conducted
293 in Tamil Nadu, the cause might be the frequent migration of people between regions in

294 search of livelihood. Although Odisha has less burden of drug resistance tuberculosis the
295 number of reported cases are also less due to lack of health facilities in close vicinity, and
296 poor lifestyle choices (30). Odisha is a tribal population dominated region and the community
297 is scattered across dense forest areas or mountains making it very difficult to estimate the
298 actual burden of tuberculosis. Studies performed in nearby state of Madhya Pradesh shared
299 insight into high tuberculosis burden Saharia tribe(20.4%) a geographically isolated
300 population of agriculture labours(31).

301 Detection of drug-resistant phenotype in a very early stage of diagnosis helps in the
302 determination of treatment regimen, which is a major contributor in determining the treatment
303 outcome. Due to the huge burden of new cases, sometimes confirmatory tests for second-
304 line drugs resistance take up to 20-30 days after initial detection of Rifampicin resistance
305 using rapid PCR-based techniques. Recent studies have demonstrated the reduced efficacy
306 of GeneXpert MTB/RIF, Hain MTBDRplus, and Hain MTBDRsl in the detection of drug
307 resistance phenotypes due to region-specific mutations that are not incorporated in the tests
308 (19).

309 In the newly sequenced samples we were able to explain most of the drug resistance ceases
310 with genotype-based predictions but for a small fraction of sequences, the prediction-based
311 approach failed to account for outcomes. Majority of the phenotypic resistance has been
312 explained by mutations in gyrase protein coding genes and rna polymerase subunit coding
313 genes. A large proportion of Rifampicin and Fluroquinolone resistant(MDR) samples
314 harboured gyrrA E21Q, D94A, G668D and A90V along with rpoB S450W, rpoC G332S
315 amino acid substitutions. For the cases where Mykrobe predictor failed to associate
316 phenotype based on genotype information, depicts a lack of region specific resistance
317 conferring features (26).

318

319 In summary large scale genomic studies are required to understand the dynamics of highly
320 diverse Mtb population India to device better region specific diagnostics methods. Also large
321 scale genome sequencing undertakings will help improving already available low cost

322 molecular diagnostic tests. Using large scale association studies targeted panels can be
323 developed for prediction of drug resistance phenotypes directly from sputum samples in
324 short period of time using massively parallel sequencing approach.

325

326 **Methodology**

327 **Sample collection and MGIT culture:**

328 The samples sequenced/used in this study were initially processed and cultured in National
329 Reference Laboratory, Regional Medical Research Centre, Odisha. The study was
330 conducted in accordance with recommended guidelines and safety procedures of ICMR-
331 Regional Medical Research Centre, Odisha. All the subjects gave written informed consent
332 in accordance with the Declaration of Helsinki.

333

334 The samples used in this study were randomly selected from a pool of CBNAAT tested
335 known Rifampicin resistance samples from North East India and sputum samples collected
336 from Odisha submitted for testing of the presence of *M. tuberculosis*. All the sputum samples
337 were initially treated with N-acetyl-l-cysteine-sodium hydroxide (NALC-NaOH) for
338 decontamination and inoculated on Lowenstein-Jensen (LJ) slants for primary culture and
339 mycobacteria growth indicator tube (MGIT) subsequently. Once the MGIT tubes showed
340 detectable growth levels of the bacteria they were kept in an incubator for another 2-3 weeks
341 to achieve sufficient numbers of bacteria for DNA isolation.

342

343 **Drug susceptibility testing:**

344 The sputum samples were tested for drug resistance using four different WHO-
345 recommended testing methods. The primary test was CBNAAT for checking the presence of
346 Rifampicin resistance in primary collection centers, the results were further validated using
347 drug resistance first line LPA testing for Rifampicin and Isoniazid (Genotype MTBDRplus
348 version 2.0 Hain Lifescience GmbH), samples showing resistance were further tested for
349 second-line LPA consisting of Fluoroquinolone resistance second-line injectables and

350 Kanamycin (low) mutations (GenoType MTBDRsl by Hain Lifescience GmbH). Samples
351 showing Fluoroquinolone and MDR phenotype were further tested for second-line drug
352 susceptibility testing using 1% proportion DST method by liquid BD BACTEC MGIT
353 automated culture challenged with Kanamycin (2.5 µg/ml), Amikacin (1.0 µg/ml),
354 Capreomycin (2.5 µg/ml), Ofloxacin (2.0 µg/ml), Levofloxacin (1.5 µg/ml), Moxifloxacin (0.5
355 and 2.0 µg/ml) in a BD BACTEC™ system to check the growth of bacteria sufficient for
356 conferring resistance. Samples showing Rifampin resistance were directly tested for second-
357 line resistance using LPA as per NTEP diagnostic recommendations.

358

359 **DNA isolation and library preparation:**

360

361 The MGIT cultured samples were kept for another 2-3 weeks to get sufficient growth for DNA
362 isolation. DNA from all the samples were isolated using a bead beating method using
363 HiPurA™ (catalogue number: MB545) from HIMEDIA labs inside the Biosafety level 3 (BSL-
364 3) facility of ICMR-RMRC, Bhubaneswar. The DNA quality of individual tuberculosis samples
365 were checked using NanoDrop™ spectrophotometers and samples having a 260/280 ratio
366 between 1.6-2.0 were selected for preparation of whole genome sequencing libraries. All
367 sequencing libraries were prepared using Nextera XT DNA library preparation kit from
368 Illumina, Inc with 1ng of input DNA in 0.2 ng/µl concentration. The prepared whole-genome
369 sequencing libraries were pooled and sequenced using NextSeq 550 system with a 2 ×
370 150bp paired-end layout.

371

372 **Quality control and classification of MTBC percent:**

373 To check the percent of Mycobacterium tuberculosis complex in each clinical isolates we ran
374 Kraken2 (PMID) against a database containing archaea, bacteria, viral, plasmid, vector in
375 RefSeq and human reference assembly (minikraken_8GB_20200312). The proportion of
376 reads falling under MTBC were summarised using a recently published method (32).

377 Samples having less than 80% of reads classified ad MTBC were removed from further
378 analysis.

379

380 **Alignment and variant calling:**

381 The samples passing MTBC filtration criteria were mapped to the H37Rv (NC_000962.3)
382 reference genome obtained from NCBI using BWA-MEM(33) . Duplicate reads from all the
383 samples were removed using Picard before calling variants. Single nucleotide variants and
384 short insertion and deletions were called using Pilon with --fix all, breaks option. From pilon
385 output sites with “PASS” filter were extracted and divided in SNV and INDEL calls (34). We
386 further filtered all variants with minimum map quality (MQ) score of 30, minimum base quality
387 (BQ) score of 20, minimum read depth of 10 or more than 80% reads supporting alternate
388 genotype, and no insertion or deletions called in the region (IC=0 and DC =60). The filtered
389 variants were then merged using bcftools merge and annotated using SnpEff(Supplementary
390 plot 1D) with H37rv reference database(35, 36). For further analysis genotype matrix and
391 annotations were extracted using bcftools and SnpSift(37).

392 Unfiltered variant call files for South India samples were directly downloaded from Broad
393 Institute portal (https://olive.broadinstitute.org/collections/tb_india.1) and the variant positions
394 were updated to NCBI H37Rv reference assembly using Picard LiftOver utility and
395 CP003248.2 to NC_000962.3 conversion chain file generated by UCSC toolkit.

396

397 **Lineage calling**

398 Lineage determination of all the samples were performed using Fast-line-caller from Farhat-
399 lab(22). The tool takes vcf files as input and assigns lineages based on presence of pre-
400 defined lineage determining mutations reported by multiple sources. We also predicted
401 digital spoligotyping using SpoTyping tool and the predicted types were annotated by
402 querying SITVIT database, both of the methods provided similar lineage calls and lineage
403 nomenclature provided in Fast lineage caller with Colls set were considered in final analysis
404 (21, 38).

405 **Phylogeny reconstruction**

406 Phylogenetic analyses were performed using only the SNVs and variants from
407 PE/PPE/PGRS were removed using bedtools intersect using mask regions provided in
408 Snippy (39). Merged multi-sample vcf files were converted to a fasta file containing the
409 genotypes using vcf2fasta tool. Construction of a phylogenetic tree was performed using IQ-
410 TREE2 (24). We have restricted model selection to GTR models and automatic model
411 selection was done with ModelFinder Plus implementation. SH-aLRT test and 1000
412 bootstrap replicates were computed using UFBoot2 implementation in the tool (40, 41).

413 **PCA of Indian and global tuberculosis isolates**

414 For PCA of tuberculosis samples we downloaded a collection of more than 2500 GMTV
415 database and merged all the files with all the Indian tuberculosis genomic variant datasets
416 and an aligned SNV fasta file was generated. Then for pairwise distance among the samples
417 were calculated using.snp-dist tool. Principal components were calculated and plotted using
418 PCAtools R package.

419

420 **Statistical analysis:**

421 Except for the mentioned cases, the rest of the statistical tests were performed using base R
422 functions and plotted using the GGPUBR wrapper of the ggplot2 package.

423

424 **Data & code availability:**

425 All the whole genome sequencing raw datasets will be uploaded to NCBI SRA repository
426 upon acceptance of this manuscript for publication. The code used for analysis and
427 statistical tests will be made available after publication of this manuscript.

428

429 **Author contributions**

430 AG, SKR planned the study and co-ordinated the data analysis and sample collection with
431 ICMR-RMRC team. AG, VKB performed the DNA extraction and library preparation of Mtb

432 samples. HB, DD provided access to BSL-3 facility, Mtb cultures and respective metadata.

433 AG wrote the manuscript SKR, DD and SP edited the manuscript.

434

435 **Declaration of interests**

436 The authors declared no conflict of interests.

437

438 **Acknowledgments**

439 AG is supported by ICMR-SRF fellowship, VKB is supported by GenomeIndia fellowship. To

440 perform the study SKR received the funding from DBT-ILS intramural grant.

441

442 **References:**

443 1. World Health Organization. 2021. Global Tuberculosis Report 2021Global
444 Tuberculosis Report.

445 2. World Health Organization (WHO). 2020. WHO report on TB 2020Who.

446 3. Ndjeka N, Schnippel K, Master I, Meintjes G, Maartens G, Romero R, Padanilam X,
447 Enwerem M, Chotoo S, Singh N, Hughes J, Variava E, Ferreira H, Te Riele J, Ismail
448 N, Mohr E, Bantubani N, Conradie F. 2018. High treatment success rate for multidrug-
449 resistant and extensively drug-resistant tuberculosis using a bedaquiline-containing
450 treatment regimen. Eur Respir J 52.

451 4. Ryu YJ. 2015. Diagnosis of pulmonary tuberculosis: recent advances and diagnostic
452 algorithms. Tuberc Respir Dis (Seoul) 78:64–71.

453 5. Manson AL, Abeel T, Galagan JE, Sundaramurthi JC, Salazar A, Gehrmann T,
454 Shanmugam SK, Palaniyandi K, Narayanan S, Swaminathan S, Earl AM. 2017.
455 Mycobacterium tuberculosis whole genome sequences from Southern India suggest
456 novel resistance mechanisms and the need for region-specific diagnostics. Clin Infect
457 Dis 64:1494–1501.

458 6. Manson AL, Abeel T, Galagan JE, Sundaramurthi JC, Salazar A, Gehrmann T,
459 Shanmugam SK, Palaniyandi K, Narayanan S, Swaminathan S, Earl AM. 2017.

- 460 Mycobacterium tuberculosis Whole Genome Sequences From Southern India
461 Suggest Novel Resistance Mechanisms and the Need for Region-Specific
462 Diagnostics. *Clin Infect Dis* 64:1494–1501.
463 7. Advani J, Verma R, Chatterjee O, Pachouri PK, Upadhyay P, Singh R, Yadav J, Naaz
464 F, Ravikumar R, Buggi S, Suar M, Gupta UD, Pandey A, Chauhan DS, Tripathy SP,
465 Gowda H, Prasad TSK. 2019. Whole genome sequencing of Mycobacterium
466 tuberculosis clinical isolates from India reveals genetic heterogeneity and region-
467 specific variations that might affect drug susceptibility. *Front Microbiol* 10:1–15.
468 8. Chatterjee A, Nilgiriwala K, Saranath D, Rodrigues C, Mistry N. 2017. Whole genome
469 sequencing of clinical strains of Mycobacterium tuberculosis from Mumbai, India: A
470 potential tool for determining drug-resistance and strain lineage. *Tuberculosis* 107:63–
471 72.
472 9. Singh J, Sankar MM, Kumar P, Couvin D, Rastogi N, Singh S, Katoch VM, Chauhan
473 DS, Katoch K, Rodrigues C, Lakshmi V, Taori GM, Dagnawala HF, Singh R,
474 Bhattacharya B, Choudhury B, Singh N, Devi U, Swaminathan S. 2015. Genetic
475 diversity and drug susceptibility profile of Mycobacterium tuberculosis isolated from
476 different regions of India. *J Infect* 71:207–219.
477 10. Coll F, Phelan J, Hill-Cawthorne GA, Nair MB, Mallard K, Ali S, Abdallah AM,
478 Alghamdi S, Alsomali M, Ahmed AO, Portelli S, Oppong Y, Alves A, Bessa TB,
479 Campino S, Caws M, Chatterjee A, Crampin AC, Dheda K, Furnham N, Glynn JR,
480 Grandjean L, Minh Ha D, Hasan R, Hasan Z, Hibberd ML, Joloba M, Jones-López EC,
481 Matsumoto T, Miranda A, Moore DJ, Mocillo N, Panaiotov S, Parkhill J, Penha C,
482 Perdigão J, Portugal I, Rchiad Z, Robledo J, Sheen P, Shesha NT, Sirgel FA, Sola C,
483 Oliveira Sousa E, Streicher EM, Helden P Van, Viveiros M, Warren RM, McNerney R,
484 Pain A, Clark TG. 2018. Genome-wide analysis of multi- and extensively drug-
485 resistant Mycobacterium tuberculosis. *Nat Genet* 50:307–316.
486 11. Comas I, Coscolla M, Luo T, Borrell S, Holt KE, Kato-Maeda M, Parkhill J, Malla B,
487 Berg S, Thwaites G, Yeboah-Manu D, Bothamley G, Mei J, Wei L, Bentley S, Harris

- 488 SR, Niemann S, Diel R, Aseffa A, Gao Q, Young D, Gagneux S. 2013. Out-of-Africa
489 migration and Neolithic coexpansion of *Mycobacterium tuberculosis* with modern
490 humans. *Nat Genet* 45:1176–1182.
- 491 12. Chen ML, Doddi A, Royer J, Freschi L, Schito M, Ezewudo M, Kohane IS, Beam A,
492 Farhat M. 2019. Beyond multidrug resistance: Leveraging rare variants with machine
493 and statistical learning models in *Mycobacterium tuberculosis* resistance prediction.
494 *EBioMedicine* 43:356–369.
- 495 13. Yang Y, Niehaus KE, Walker TM, Iqbal Z, Walker AS, Wilson DJ, Peto TEA, Crook
496 DW, Smith EG, Zhu T, Clifton DA. 2018. Machine learning for classifying tuberculosis
497 drug-resistance from DNA sequencing data. *Bioinformatics* 34:1666–1671.
- 498 14. Yang Y, Walker TM, Walker AS, Wilson DJ, Peto TEA, Crook DW, Shamout F,
499 CRyPTIC Consortium, Zhu T, Clifton DA. 2019. DeepAMR for predicting co-occurrent
500 resistance of *Mycobacterium tuberculosis*. *Bioinformatics* 35:3240–3249.
- 501 15. Chatterjee A, Nilgiriwala K, Saranath D, Rodrigues C, Mistry N. 2017. Whole genome
502 sequencing of clinical strains of *Mycobacterium tuberculosis* from Mumbai, India: A
503 potential tool for determining drug-resistance and strain lineage. *Tuberculosis* 107:63–
504 72.
- 505 16. Gutierrez MC, Ahmed N, Willery E, Narayanan S, Hasnain SE, Chauhan DS, Katoch
506 VM, Vincent V, Locht C, Supply P. 2006. Predominance of ancestral lineages of
507 *Mycobacterium tuberculosis* in India. *Emerg Infect Dis* 12:1367–74.
- 508 17. Merker M, Barbier M, Cox H, Rasigade JP, Feuerriegel S, Kohl TA, Diel R, Borrell S,
509 Gagneux S, Nikolayevskyy V, Andres S, Nübel U, Supply P, Wirth T, Niemann S.
510 2018. Compensatory evolution drives multidrug-resistant tuberculosis in central Asia.
511 *Elife* 7:1–31.
- 512 18. Manson AL, Cohen KA, Abeel T, Desjardins CA, Armstrong DT, Barry CE, Brand J,
513 Chapman SB, Cho SN, Gabrielian A, Gomez J, Jodals AM, Joloba M, Jureen P, Lee
514 JS, Malinga L, Maiga M, Nordenberg D, Noroc E, Romancenco E, Salazar A,
515 Ssengooba W, Velayati AA, Winglee K, Zalutskaya A, Via LE, Cassell GH, Dorman

- 516 SE, Ellner J, Farnia P, Galagan JE, Rosenthal A, Crudu V, Homorodean D, Hsueh
517 PR, Narayanan S, Pym AS, Skrahina A, Swaminathan S, Van Der Walt M, Alland D,
518 Bishai WR, Cohen T, Hoffner S, Birren BW, Earl AM. 2017. Genomic analysis of
519 globally diverse *Mycobacterium tuberculosis* strains provides insights into the
520 emergence and spread of multidrug resistance. *Nat Genet* 49:395–402.
- 521 19. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, Gillespie JJ,
522 Gough R, Hix D, Kenyon R, Machi D, Mao C, Nordberg EK, Olson R, Overbeek R,
523 Pusch GD, Shukla M, Schulman J, Stevens RL, Sullivan DE, Vonstein V, Warren A,
524 Will R, Wilson MJC, Yoo HS, Zhang C, Zhang Y, Sobral BW. 2014. PATRIC, the
525 bacterial bioinformatics database and analysis resource. *Nucleic Acids Res* 42:D581-
526 91.
- 527 20. Anyansi C, Keo A, Walker BJ, Straub TJ, Manson AL, Earl AM, Abeel T. 2020.
528 QuantTB- A method to classify mixed *Mycobacterium tuberculosis* infections within
529 whole genome sequencing data. *BMC Genomics* 21.
- 530 21. Coll F, McNerney R, Guerra-Assunção JA, Glynn JR, Perdigão J, Viveiros M, Portugal
531 I, Pain A, Martin N, Clark TG. 2014. A robust SNP barcode for typing *Mycobacterium*
532 tuberculosis complex strains. *Nat Commun* 5:4–8.
- 533 22. Freschi L, Vargas R, Husain A, Kamal SMM, Skrahina A, Tahseen S, Ismail N,
534 Barbova A, Niemann S, Cirillo DM, Dean AS, Zignol M, Farhat MR. 2021. Population
535 structure, biogeography and transmissibility of *Mycobacterium tuberculosis*. *Nat
536 Commun* 12.
- 537 23. Devi KR, Pradhan J, Bhutia R, Dadul P, Sarkar A, Gohain N, Narain K. 2021.
538 Molecular diversity of *Mycobacterium tuberculosis* complex in Sikkim, India and
539 prediction of dominant spoligotypes using artificial intelligence. *Sci Rep* 11:7365.
- 540 24. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A,
541 Lanfear R, Teeling E. 2020. IQ-TREE 2: New Models and Efficient Methods for
542 Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* 37:1530–1534.
- 543 25. Chernyaeva EN, Shulgina M V., Rotkevich MS, Dobrynin P V., Simonov SA, Shitikov

- 544 EA, Ischenko DS, Karpova IY, Kostryukova ES, Ilina EN, Govorun VM, Zhuravlev VY,
545 Manicheva OA, Yablonsky PK, Isaeva YD, Nosova EY, Mokrousov I V., Vyazovaya
546 AA, Narvskaya O V., Lapidus AL, O'Brien SJ. 2014. Genome-wide Mycobacterium
547 tuberculosis variation (GMTV) database: A new tool for integrating sequence
548 variations and epidemiology. BMC Genomics 15.
- 549 26. Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, Earle S, Pankhurst LJ,
550 Anson L, De Cesare M, Piazza P, Votintseva AA, Golubchik T, Wilson DJ, Wyllie DH,
551 Diel R, Niemann S, Feuerriegel S, Kohl TA, Ismail N, Omar S V., Smith EG, Buck D,
552 McVean G, Walker AS, Peto TEA, Crook DW, Iqbal Z. 2015. Rapid antibiotic-
553 resistance predictions from genome sequence data for *Staphylococcus aureus* and
554 *Mycobacterium tuberculosis*. Nat Commun 6.
- 555 27. Farhat MR, Jesse Shapiro B, Kieser KJ, Sultana R, Jacobson KR, Victor TC, Warren
556 RM, Streicher EM, Calver A, Sloutsky A, Kaur D, Posey JE, Plikaytis B, Oggioni MR,
557 Gardy JL, Johnston JC, Rodrigues M, C Tang PK, Kato-Maeda M, Borowsky ML,
558 Muddukrishna B, Kreiswirth BN, Kurepina N, Galagan J, Gagneux S, Birren B, Rubin
559 EJ, Lander ES, Sabeti PC, Murray M. 2013. Genomic analysis identifies targets of
560 convergent positive selection in drug-resistant *Mycobacterium tuberculosis*. Nat Publ
561 Gr 45.
- 562 28. Vargas R, Freschi L, Marin M, Epperson LE, Smith M, Oussenko I, Durbin D, Strong
563 M, Salfinger M, Farhat MR. 2021. In-host population dynamics of mycobacterium
564 tuberculosis complex during active disease. Elife 10:1–37.
- 565 29. Ektefaie Y, Dixit A, Freschi L, Farhat MR. 2021. Globally diverse *Mycobacterium*
566 tuberculosis resistance acquisition: a retrospective geographical and temporal
567 analysis of whole genome sequences. The Lancet Microbe 2:e96–e104.
- 568 30. Hussain T, Tripathy SS, Das S, Satapathy P, Das D, Thomas B, Pati S. 2020.
569 Prevalence, risk factors and health seeking behaviour of pulmonary tuberculosis in
570 four tribal dominated districts of Odisha: Comparison with studies in other regions of
571 India. PLoS One 15.

- 572 31. Rao VG, Bhat J, Yadav R, Sharma RK, Muniyandi M. 2019. Declining tuberculosis
573 prevalence in Saharia, a particularly vulnerable tribal community in Central India:
574 evidences for action. *BMC Infect Dis* 19:180.
- 575 32. Vargas R, Freschi L, Marin M, Epperson LE, Smith M, Oussenko I, Durbin D, Strong
576 M, Salfinger M, Farhat MR. 2021. In-host population dynamics of *Mycobacterium*
577 tuberculosis complex during active disease. *Elife* 10.
- 578 33. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with
579 BWA-MEM.
- 580 34. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng
581 Q, Wortman J, Young SK, Earl AM. 2014. Pilon: An integrated tool for comprehensive
582 microbial variant detection and genome assembly improvement. *PLoS One* 9.
- 583 35. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A,
584 Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and
585 BCFtools. *Gigascience* 10.
- 586 36. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden
587 DM. 2012. A program for annotating and predicting the effects of single nucleotide
588 polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain
589 w1118; iso-2; iso-3. *Fly (Austin)* 6:80–92.
- 590 37. Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, Lu X. 2012. Using
591 *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a
592 new program, SnpSift. *Front Genet* 3.
- 593 38. Xia E, Teo Y-Y, Ong RT-H. 2016. SpoTyping: fast and accurate in silico
594 *Mycobacterium* spoligotyping from sequence reads. *Genome Med* 8:19.
- 595 39. Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing
596 genomic features. *Bioinformatics* 26:841–842.
- 597 40. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017.
598 ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods*
599 14:587–589.

- 600 41. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2:
601 Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol* 35:518–522.
602
603
604

605 **Figure 1** A) *M. tuberculosis* whole genome sequence samples from our study and collected
606 from other studies for comparative analysis. B) Distribution of Mtb lineages across different
607 regions with number of samples. C) Drug resistance phenotype of samples from our study
608 represented by stacked bars and groups by respective regions. D) Effect of variants on
609 corresponding gene predicted using snpEFF.

610

611 **Figure 2** Neighbor joining tree representing *M. tuberculosis* isolates from North, South,
612 North-East and East India isolates, with 1000 bootstraps (Outgroup *M. canetti* not shown).
613 The node colour represents isolate lineage, inner circle shows corresponding drug
614 resistance phenotype and outer circle coloured based on the isolates geographical location
615 (divisions/states).

616

617 **Figure 3** A) Principle component analysis of Indian tuberculosis strains with GMTV strains,
618 shape of the dots are used to represent the lineage and colour represents the dataset/strains
619 source. B) Distribution of pair-wise SNP distances of North East and Odisha sequences,
620 dashed line represents 35 SNP differences among the pairs. C) Phylogenetic tree
621 representing the Sikkim sample local transmission clusters with their drug resistance
622 phenotypes, lineage and collection district.

623

624 **Figure 4** A) Functional categories of highly mutated genes (ranked based on number of non-
625 synonymous mutations detected) . B-C) Drug resistance associated genes and amino acid
626 changes with respective mutation density and frequency in sequenced samples. D) Number
627 of INDELs detected per phenotypic categories compared with susceptible samples
628 compared using non-parametric test. E) Allele fractions of follow up isolates plotted against
629 respective depth of coverage coloured by days post initial collection.

630

631 **Supplementary Figure 1** A) Age distribution of patients enrolled for the study. B) Fractional
632 distribution of MTBC detected in sequenced samples. C) INDEL length and count distribution

633 , where negative lengths represents deletions and positive lengths insertions. D) Distribution
634 of deletions detected in using Delly2 pipeline. E) Functional annotation of single nucleotide
635 variants.

636

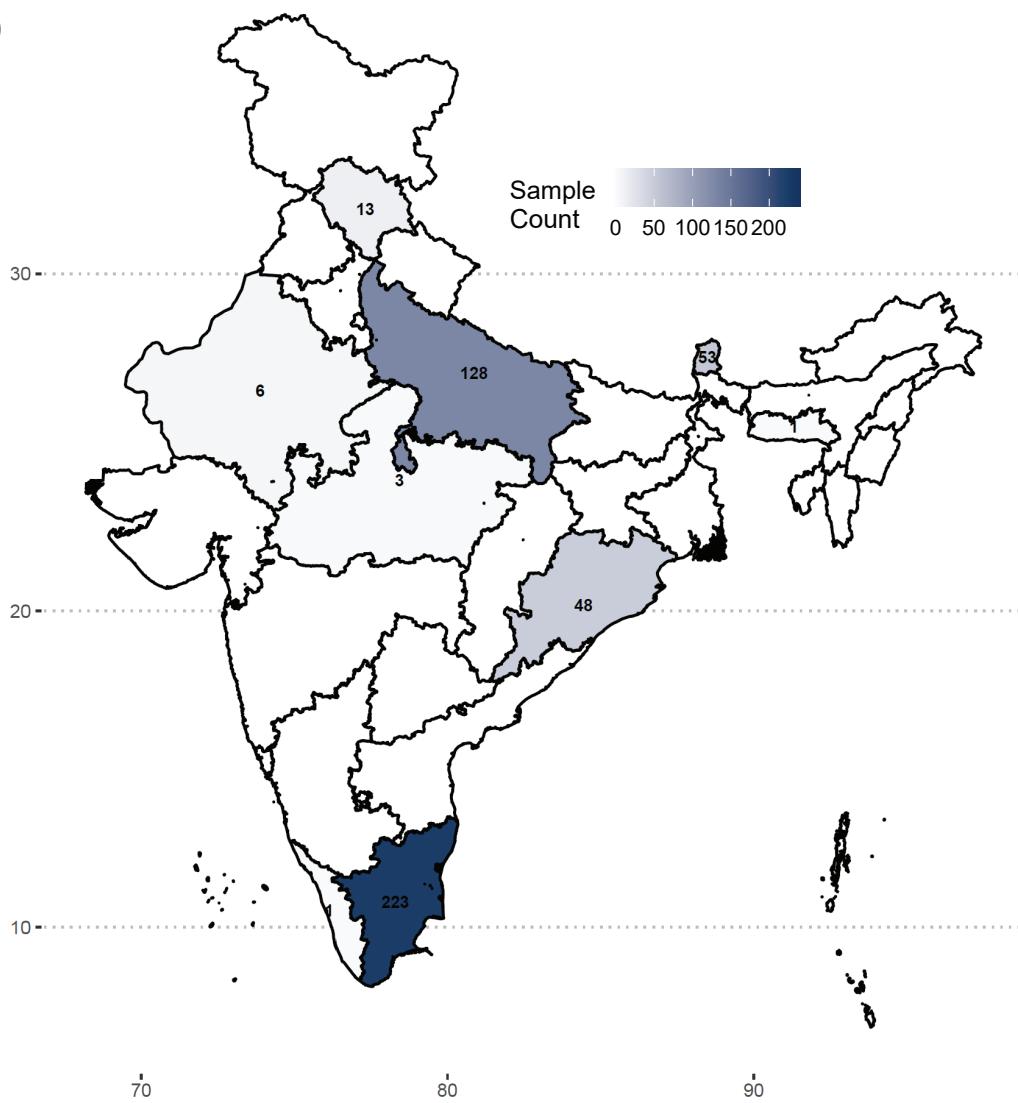
637 **Supplementary Figure 2** A) Hierarchical clustering of pair-wise SNP distances annotated
638 with lineage state and drug resistance phenotypes.

639

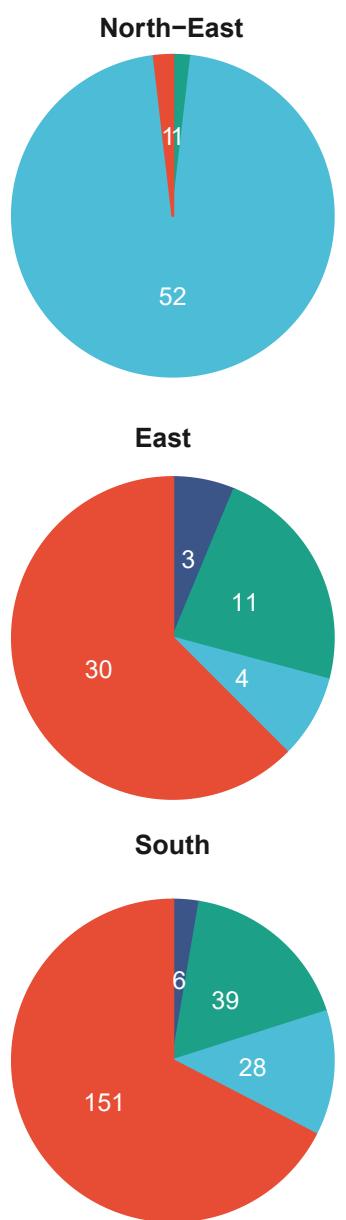
Figure 1

bioRxiv preprint doi: <https://doi.org/10.1101/2022.04.11.487831>; this version posted April 11, 2022. 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-NC-ND 4.0 International license.

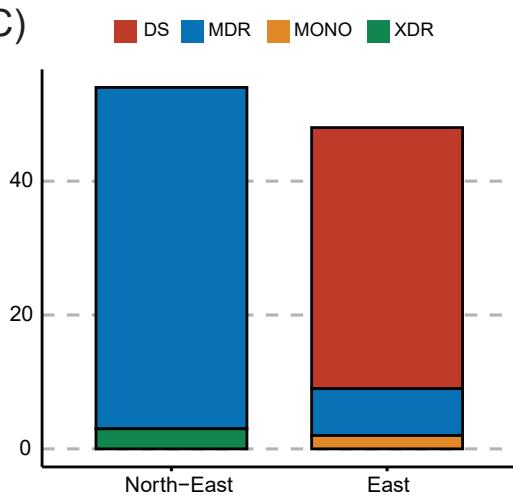
A)



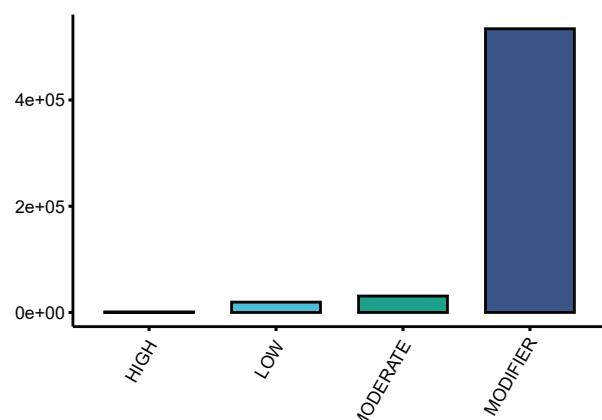
B)



C)



D)



North

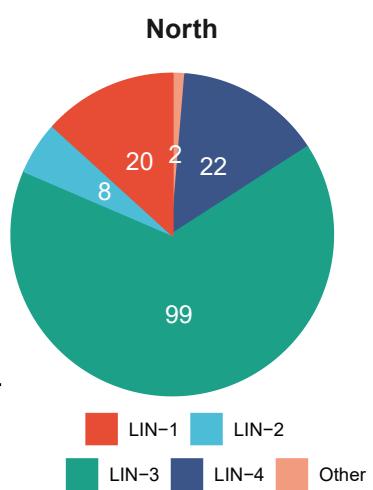


Figure 2

bioRxiv preprint doi: <https://doi.org/10.1101/2022.04.11.487831>; this version posted April 11, 2022. 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-NC-ND 4.0 International license.

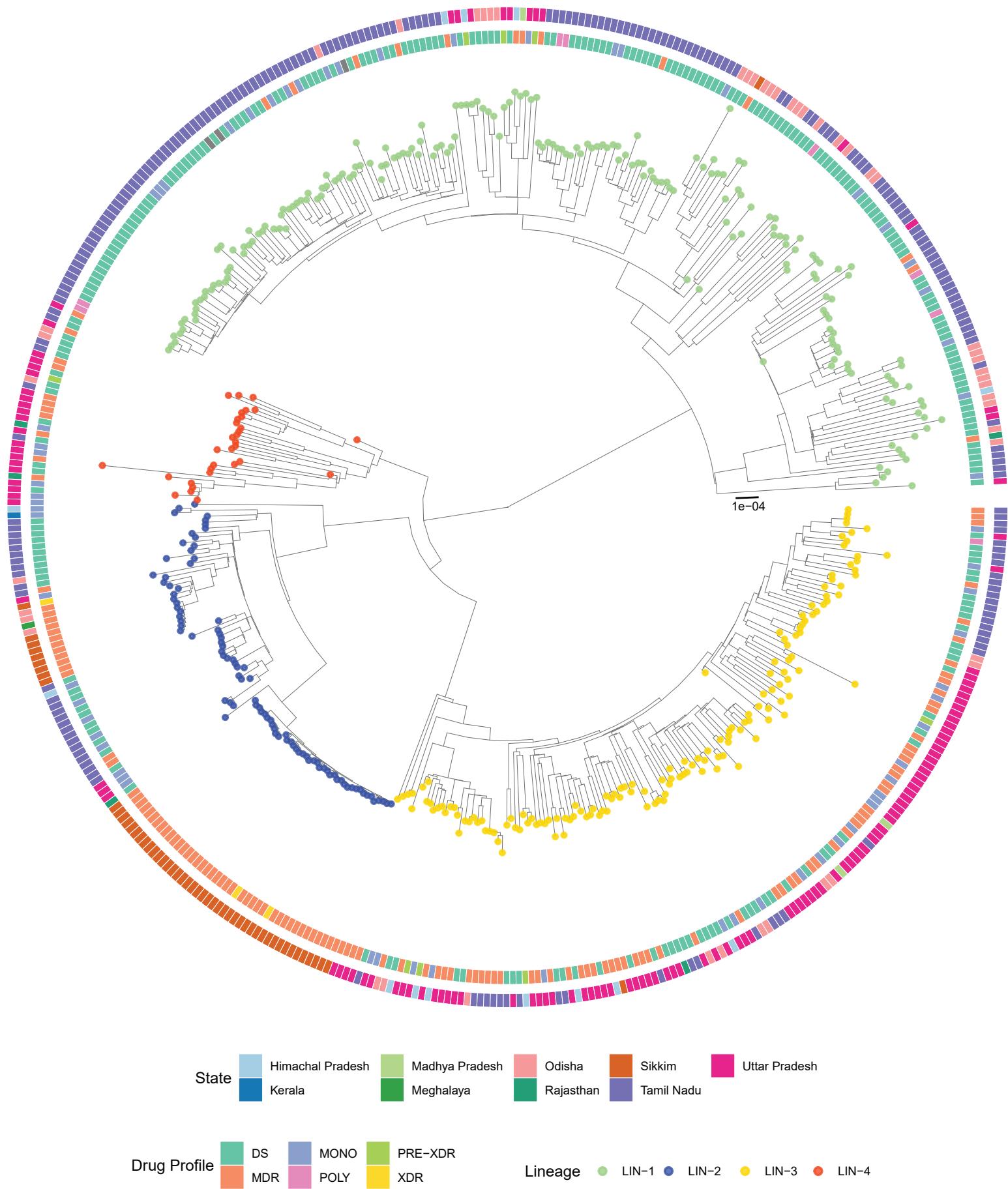


Figure 3

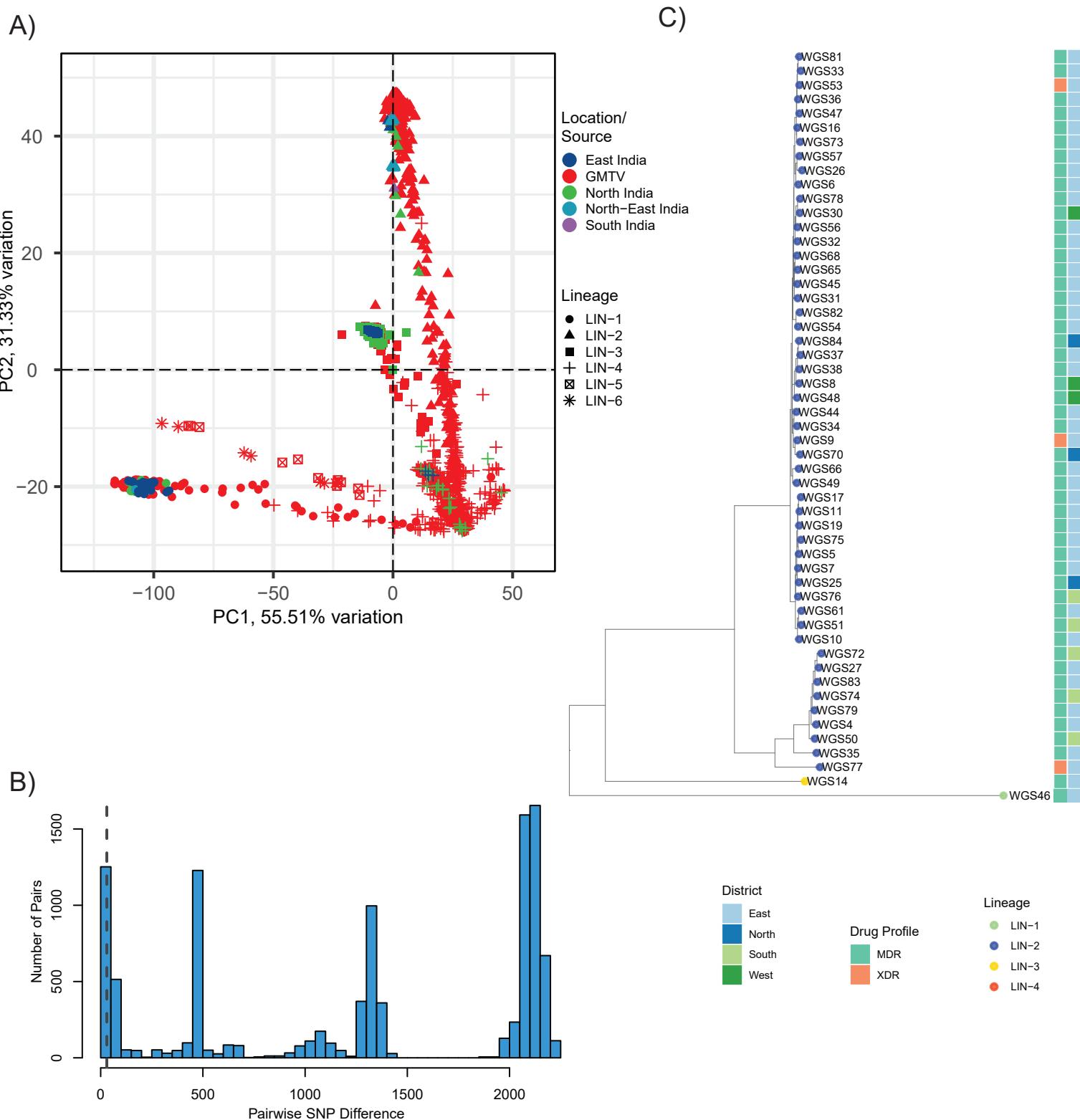
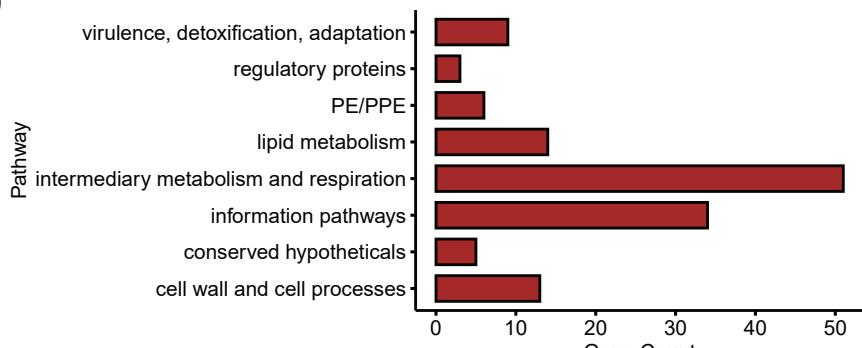


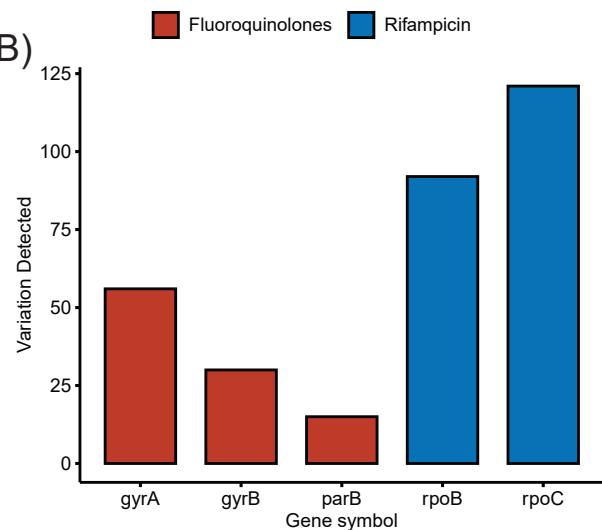
Figure 4

bioRxiv preprint doi: <https://doi.org/10.1101/2022.04.11.487831>; this version posted April 11, 2022. 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-NC-ND 4.0 International license.

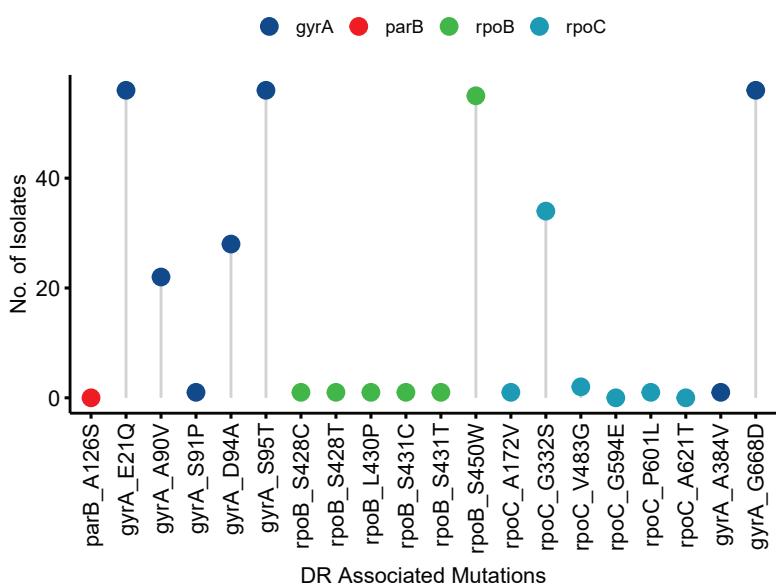
A)



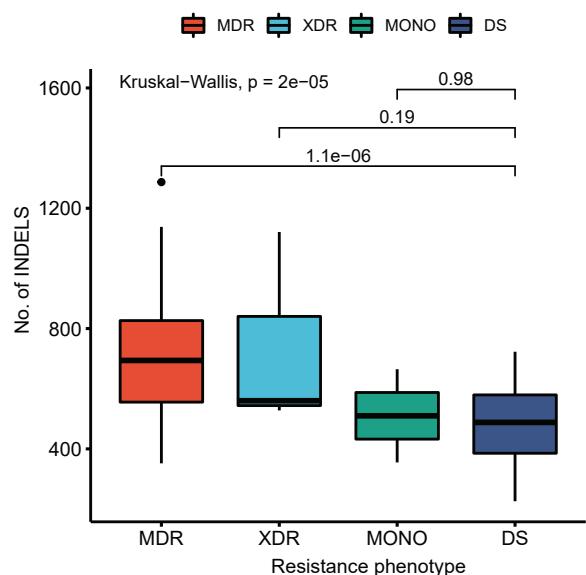
B)



C)



D)



E)

