

1 **Full title:** Phylogeography and transmission of *M. tuberculosis* spanning prisons and surrounding  
2 communities in Paraguay

3  
4 **Short title:** *M. tuberculosis* transmission in Paraguay

5  
6 **Authors**

7 Gladys Estigarribia Sanabria<sup>1¶\*</sup>, Guillermo Sequera<sup>2,3¶</sup>, Sarita Aguirre<sup>3</sup>, Julieta Méndez<sup>1</sup>, Paulo César  
8 Pereira dos Santos<sup>4</sup>, Natalie Weiler Gustafson<sup>5</sup>, Analía Ortiz<sup>1</sup>, Cynthia Cespedes<sup>3</sup>, Gloria Martínez<sup>1</sup>,  
9 Alberto L. García-Basteiro<sup>6,7,8</sup>, Jason R. Andrews<sup>9</sup>, Julio Croda<sup>10,11,12&</sup>, Katharine S. Walter<sup>9&\*</sup>

10  
11 **Affiliations**

- 12 1. Instituto Regional de Investigación en Salud, Caaguazú, Paraguay
- 13 2. Instituto de Salud Global de Barcelona (ISGLOBAL), Barcelona, Spain
- 14 3. Programa Nacional de Control de la Tuberculosis, Ministerio de Salud Pública y Bienestar Social  
15 (MSPyBS), Asunción, Paraguay
- 16 4. Postgraduate Program in Infectious and Parasitic Diseases, Federal University of Mato Grosso do  
17 Sul, Mato Grosso do Sul, Brazil
- 18 5. Laboratorio Central de Salud Pública (LCSP), Ministerio de Salud Pública y Bienestar Social  
19 (MSPyBS), Asunción, Paraguay
- 20 6. Instituto de Salud Global de Barcelona (ISGLOBAL), Barcelona, Spain
- 21 7. Centro de Investigação em Saude de Manhica (CISM), Maputo, Mozambique
- 22 8. Centro de Investigación Biomédica en Red de Enfermedades Infecciosas (CIBERINFEC),  
23 Barcelona, Spain
- 24 9. Division of Infectious Diseases and Geographic Medicine, Stanford University School of  
25 Medicine, Stanford, CA, USA
- 26 10. Federal University of Mato Grosso do Sul - UFMS, Campo Grande, MS, Brazil

27 11. Oswaldo Cruz Foundation Mato Grosso do Sul, Mato Grosso do Sul, Brazil

28 12. Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven,

29 USA

30

31 ¶ These authors contributed equally to this work.

32 & These authors also contributed equally to this work.

33 \*Corresponding authors.

34

35

36

37

38

39 **Abstract (150 words)**

40

41 Recent rises in incident tuberculosis (TB) cases in Paraguay and the increasing concentration of TB  
42 within prisons highlight the urgency of targeting strategies to interrupt transmission and prevent new  
43 infections. However, whether specific cities or carceral institutions play a disproportionate role in  
44 transmission remains unknown. We conducted prospective genomic surveillance, sequencing 471 *M.*  
45 *tuberculosis* genomes, from inside and outside prisons in Paraguay's two largest urban areas, Asunción  
46 and Ciudad del Este, from 2016 to 2021. We found genomic evidence of frequent recent transmission  
47 within prisons and transmission linkages spanning prisons and surrounding populations. We identified a  
48 signal of frequent *M. tuberculosis* spread between urban areas and marked recent population size  
49 expansion of the three largest genomic transmission clusters. Together, our findings highlight the urgency  
50 of strengthening TB control programs to reduce transmission risk within prisons, where, in Paraguay,  
51 incidence was 70 times that outside prisons in 2021.

52

53 **Keywords:** tuberculosis, genomics, transmission, prisons, Paraguay

54

55 **Financing agencies:** National Institutes of Health grants R01 AI130058 (JRA) and R01 AI149620 (JRA  
56 and JC). Paraguay National Commission of Science and Technology grant CONACYT PIN 15-705 (GES  
57 and GS).

58

59

60

61

62

## 63 Introduction

64 Despite significant tuberculosis (TB) control efforts, the incidence rate of TB has declined only  
65 slowly in the World Health Organization Region of the Americas, and, alarmingly, has stagnated since  
66 2014<sup>1</sup>. The COVID-19 pandemic disrupted access to healthcare—including critical TB diagnostic and  
67 treatment programs—compounding the burden of TB and reversing decades of progress in TB control<sup>1</sup>.

68 New approaches to limit transmission are urgently needed in Paraguay, where TB control is  
69 chronically underfunded<sup>1</sup> and where TB incidence was 48 (41-56) per 100,000 people in 2020, higher  
70 than the mean incidence rate across the region<sup>1</sup>. More than a quarter of the country's population lives  
71 below the national poverty line<sup>2</sup> and are at heightened risk of TB infection and mortality. Further, recent  
72 dramatic increases in incarceration<sup>3,4</sup> put a rapidly growing population at high risk of infection and  
73 disease<sup>5-7</sup>. To guide interventions in Paraguay, there is a critical need to identify the populations at  
74 greatest risk of infection and locations and institutions where transmission most frequently occurs<sup>8</sup>.

75 Whole genome sequencing of *M. tuberculosis* has been powerfully applied to characterize recent  
76 transmission dynamics. Genomic approaches have dated introductions of *M. tuberculosis* and  
77 reconstructed patterns of historic spread across Central and South America<sup>9,10</sup>, estimated the contribution  
78 of recent transmission to incident TB cases<sup>11</sup>, reconstructed the emergence of resistance-associated  
79 mutations<sup>12</sup>, and inferred likely individual-level transmission events<sup>13</sup>. In Brazil<sup>14</sup> and Georgia<sup>13</sup>, for  
80 example, genomic approaches identified frequent transmission within prisons as well as evidence of  
81 spillover from prisons to surrounding communities. A single *M. tuberculosis* molecular study from  
82 Paraguay<sup>15</sup> on strains collected in 2003 reported that *M. tuberculosis* spoligotypes found across South  
83 America including the Latin-American (LAM) and Haarlem lineages were also common in Paraguay<sup>15,16</sup>.

84 Genomic approaches have not been applied to address major gaps in our understanding of TB  
85 transmission in Paraguay. First, the conditions of incarceration put people at high risk of many infectious  
86 diseases and globally, over the past twenty years, the incarcerated population in Central and South  
87 America has grown by 206%, the greatest increase in the world<sup>17</sup>. Escalating incarceration rates have been  
88 paralleled by an increasing concentration of notified TB among incarcerated individuals<sup>6</sup>. Yet the role of

89 prison environments on TB transmission both inside and beyond prisons, as sources of broader infection,  
90 has not yet been described in Paraguay. Second, while incidence of TB is heterogeneous across the  
91 country<sup>15</sup>, it remains unknown whether specific cities or regions function as hotspots, fueling transmission  
92 elsewhere. Finally, due to limited surveillance infrastructure, the prevalence of drug-resistance and multi-  
93 drug-resistance has not yet been systematically measured<sup>18–20</sup>. Only 56% of bacteriologically confirmed  
94 new cases of pulmonary TB were tested for rifampicin resistance in 2020<sup>1</sup>.

95 To characterize transmission dynamics and circulating diversity of *M. tuberculosis* in Paraguay,  
96 we conducted prospective genomic surveillance across the country from 2016 to 2021, including  
97 surveillance within and outside prisons, generating a genomic resource for continued surveillance in  
98 Paraguay. We estimated the role of likely recent *M. tuberculosis* transmission within prisons, the  
99 relatedness of prison and community transmission, and the frequent movement of *M. tuberculosis*  
100 between Paraguay's urban centers.

101

## 102 **Methods**

### 103 **Prospective population surveillance**

104 We conducted population-based genomic surveillance in three of Paraguay's departments:  
105 Central, Distrito Capital, and Alto Paraná, which together comprise approximately half (3,392,429  
106 people) of the country's 2021 population of 7.4 million. Sputum samples are routinely collected from all  
107 individuals presenting with symptoms of TB at primary health clinics and sent to the National Program  
108 for Tuberculosis Control (NPTC) of Paraguay reference laboratory for microbial diagnostics including  
109 culture and smear microscopy.

110 Study recruitment was done by study staff who visited patients at home and in prisons at the time  
111 patients began treatment (Directly Observed Therapy, DOT). At this time, the standard National TB  
112 Control Program questionnaire was conducted and patients who chose to enroll provided written consent  
113 for sequencing residual mycobacterial cultures for culture-positive samples. Study staff also collected  
114 additional demographic, clinical, residential, and epidemiological data, including information on history of

115 prior or current incarceration with a structured questionnaire. The study was approved by the ethics  
116 committee of the Central Laboratory of Public Health of the Ministry of Health and Social Welfare of  
117 Paraguay (International Certification FWA N° FWAOOO20088) with code CEI-LCSP 91/010217.

118

## 119 **Laboratory and sequencing methods**

120 Sputum samples were cultured in the Ogawa-Kudoh Method described previously<sup>21,22</sup>. Cultures  
121 were incubated at 37°C and observed for growth twice a week for 60 days. *M. tuberculosis* DNA was  
122 extracted using Cetyltrimethyl ammonium bromide (CTAB) method<sup>23</sup>.

123 Sequencing was conducted at the Laboratorio Central de Salud Pública (LCSP), Paraguay  
124 Ministry of Health; Centro para el Desarrollo a la investigación Científica (CEDIC), Paraguay; and the  
125 Translational Genomics Research Institute (TGen), Arizona, US. DNA sequencing libraries were  
126 prepared with the Illumina DNA Prep library kit and sequenced on an Illumina MiSeq in Paraguay and an  
127 Illumina NextSeq (2 × 151-bp), at TGen. Raw sequence reads for samples passing filters are available at  
128 the Sequence Read Archive (PRJNA870648).

129

## 130 **Variant identification**

131 We identified *M. tuberculosis* genomic variation from whole genome sequence data with a  
132 previously described pipeline available at [https://github.com/ksw9/mtb\\_pipeline](https://github.com/ksw9/mtb_pipeline)<sup>24</sup>. Briefly, we trimmed  
133 low-quality bases (Phred-scaled base quality < 20) and removed adapters with Trim Galore  
134 (stringency=3)<sup>25</sup>. We used CutAdapt to further filter reads (--nextseq-trim=20 --minimum-length=20 --  
135 pair-filter=any)<sup>26</sup>. To exclude potential contamination, we used Kraken2 to taxonomically classify reads  
136 and removed reads that were not assigned to the *Mycobacterium* genus or that were assigned to a  
137 *Mycobacterium* species other than *M. tuberculosis*<sup>27</sup>. We mapped reads with bwa v. 0.7.15 (bwa mem)<sup>28</sup>  
138 to the H37Rv reference genome (NCBI Accession: NC\_000962.3) and removed duplicates with  
139 sambamba<sup>29</sup>. We called variants with GATK 4.1 HaplotypeCaller<sup>30</sup>, setting sample ploidy to 1, and  
140 GenotypeGVCFs, including non-variant sites in output VCF files. We included variant sites with a

141 minimum depth of 10X and a minimum variant quality score 40 and constructed consensus sequences  
142 with bcftools consensus<sup>31</sup>, excluding indels. We excluded SNPs in previously defined repetitive regions  
143 (PPE and PE-PGRS genes, phages, insertion sequences and repeats longer than 50 bp)<sup>32</sup>. We identified  
144 sub-lineage and evidence of mixed infection with TBProfiler v.4.2.0<sup>33,34</sup>. We additionally used  
145 TBProfiler with the TBDB repository (<https://github.com/jodyphelan/tbdb>) to identify resistance-  
146 associated mutations<sup>33,34</sup>. We described overall patterns of genomic drug-resistance including mutations  
147 that are included in the World Health Organisation catalogue as independently associated with  
148 resistance<sup>35</sup>.

149

### 150 **Phylogenetic and Bayesian evolutionary analysis**

151 We constructed full-length consensus FASTA sequences from VCF files, setting missing  
152 genotypes to missing, and used SNP-sites to extract a multiple alignment of internal variant sites only<sup>36</sup>.  
153 We used the R package *ape* to measure pairwise differences between samples  
154 (`pairwise.deletion=TRUE`)<sup>37</sup>. We selected a best fit substitution model with ModelFinder<sup>38</sup>, implemented  
155 in IQ-TREE multicore version 2.2.0<sup>39</sup>, evaluating all models that included an ascertainment bias  
156 correction for the use of an alignment of SNPs only. The best fit model according to Bayesian  
157 Information Criterion was K3Pu+F+ASC+R5, a three substitution types model with unequal base  
158 frequencies, an ascertainment bias correction, and a FreeRate model of rate heterogeneity across sites,  
159 including four categories. We then fit a maximum likelihood tree with IQ-TREE, with 1000 ultrafast  
160 bootstrap replicates<sup>39,40</sup>.

161 Genomic clustering is often used as a proxy measure of recent *M. tuberculosis* transmission;  
162 isolates that are more closely genetically related are hypothesized to be more likely linked through recent  
163 transmission rather than travel-associated importation or re-activation of genetically distinct latent  
164 infections.<sup>41,42</sup> We applied a commonly used genetic distance thresholds of 12- and 5- or fewer SNPs to  
165 identify genomic clusters<sup>43,44</sup>.

166 To investigate transmission patterns in the three largest genomic clusters more closely, we fit  
167 timed Bayesian trees to multiple sequence alignments with BEAST 2.6.2<sup>45</sup>, using TB notification dates to  
168 calibrate tips. Because of the short sampling timeframe of our data, we fixed the substitution rate to  $1 \times 10^{-7}$   
169 mutations/site/year, as previously described<sup>46</sup>, and consistent with previous estimates for the *M.*  
170 *tuberculosis* lineage 4 substitution rate<sup>47</sup>. To examine population dynamics in the three largest clusters,  
171 we used a Coalescent Bayesian Skyline model<sup>48</sup> with 5 dimensions, allowing the effective population size  
172 to change 4 times over the tree. We additionally fit a Bayesian tree to sublineage 4.2.1.1 samples using a  
173 constant population size, fixed substitution rate model. Markov chain Monte Carlo chains were run for  
174 200 million iterations, or longer, if required for convergence, excluding 10% of samples as burn-in. We  
175 used *treeannotator* to produce maximum clade credibility trees. We used the R package *beautier* to  
176 construct XML files<sup>49</sup> and corrected XML files for the number of constant positions in SNP alignments.  
177 We visualized phylogenetic trees with the R package *ggtree*<sup>50,51</sup>.

178 We calculated time-scaled haplotype density from a matrix of pairwise SNP distances with the R  
179 package *thd* as previously described<sup>52</sup> and compared time-scaled haplotype density between individuals  
180 who were never, formerly, or currently incarcerated with t-tests. We set the *M. tuberculosis* substitution  
181 rate to  $1 \times 10^{-7}$  substitutions per site per year and included an effective genome length of 3,916,441  
182 basepairs (the length of the reference genome minus the PE/PPE regions excluded from variant calling)  
183 and used a short (20 year) and long (50 year) epidemic timescale. We compared time-scaled haplotype  
184 density by incarceration status with t-tests and used analysis of variance to test for the independent effect  
185 of incarceration status after controlling for *M. tuberculosis* population structure, as previously described<sup>52</sup>.

186 We conducted discrete ancestral state reconstruction for sampling location with the R package  
187 *ape* for the three largest sub-lineages in our collection<sup>37</sup>. We restricted samples to those from Asunción  
188 and Ciudad del Este because of the small sample size outside those cities. We compared symmetric and  
189 asymmetric rates models fit with the R package *diversitree* (*make.mk2*) and compared model fits with  
190 analysis of variance<sup>53</sup>. We used stochastic character mapping<sup>54</sup> in the R package *diversitree*<sup>53</sup> to sample



191 500 location histories for each sublineage tree; we summarized these as the number of average  
192 movements between cities over the tree.

193

## 194 **Results**

### 195 **Population-based genomic surveillance**

196 From 2016 to 2021, 16,734 TB cases were notified in Paraguay, with the majority of cases (60%;  
197 10,095/16,734) occurring in the urban departments Central, Distrito Capital, and Alto Paraná, where we  
198 conducted prospective genomic surveillance (Fig. 1a). In 2021, the TB notification rate was 70 times  
199 higher in prisons than outside (3,378 cases per 100,000 in prisons/49 cases per 100,000 in the general  
200 population) (Fig. 1b). Therefore, we included focused genomic surveillance in the two largest prisons in  
201 the country, Tacumbú Prison and the Prison of Ciudad del Este, which together hold 36% (4,950/13,821)  
202 of Paraguay's incarcerated population, notification rates are 2,000 and 3,500 per 100,000 people,  
203 respectively.

204 Whole genome sequences (WGS) for a total of 532 isolates met our coverage and quality criteria  
205 (Methods), including 488 from unique TB notifications. Of the samples passing filters, 158 were from  
206 individuals diagnosed with TB while in prison and 330 were from people diagnosed in the community.  
207 TB isolates were collected in Asunción (274/488) and in Ciudad del Este (214/488). We excluded 17  
208 isolates with evidence of mixed infection with more than one sub-lineage detected, resulting in 471 *M.*  
209 *tuberculosis* isolates for following analyses.

210

### 211 **Genotypic resistance**

212 The majority, 96% (454/471) of sampled *M. tuberculosis*, were drug sensitive; 3% (15/471) were  
213 resistant to at least one drug; and 0.42% (2/471) were multi-drug resistant, resistant to both isoniazid and  
214 rifampin. Resistance was not associated with sub-lineage ( $X^2(11) = 7.7, p = 0.74$ ). We identified three  
215 unique isoniazid resistance-conferring mutations on the genes *fabG1*, *katG*, or both among the 10 isolates

216 with any isoniazid resistance; the three rifampicin-conferring mutations in *rpoB* (two on multi-drug  
217 resistant isolates) were unique (Fig. 2).

218

### 219 **Stable genomic diversity of *M. tuberculosis* in Paraguay**

220 After excluding mixed infections, all *M. tuberculosis* isolates were from *M. tuberculosis* lineage  
221 4. The single mixed infection was co-infected with both lineages 1 and 4. Samples predominantly fell into  
222 four sub-lineages: 4.3.3/LAM (42.5%; 200/471), 4.1.2 /Haarlem (38.2%; 180/471), 4.4.1 /S (12.3%;  
223 58/471), and 4.3.4/LAM (3.2%; 15/471) (Fig. 2). The distribution of these sublineages was stable and did  
224 not change significantly from a collection of 173 *M. tuberculosis* isolates collected in 2003<sup>15</sup> (Fig. S1).

225

### 226 **Geographic structure despite frequent migration across *M. tuberculosis* sub-lineages**

227 We found a strong pattern of geographic structure (Fig. 3), with sub-lineage 4.1.2.1 dominant in  
228 Asunción (54.1%, 142/262 samples) and sub-lineage 4.3.3 dominant in Ciudad del Este (60.8%, 127/209)  
229 ( $\chi^2(11)=81, p<0.001$ ) (Fig. 3). While we observed geographically distinct patterns of *M. tuberculosis*  
230 diversity in Asunción and Ciudad del Este, reconstruction of the ancestral locations for the three most  
231 prevalent sub-lineages revealed frequent movement of *M. tuberculosis*. Ancestors of sampled sub-lineage  
232 4.1.2.1 isolates likely moved between the two major cities at least 150 times, since the most recent  
233 common ancestor (MCRA) of the sub-lineage in Paraguay; sub-lineages 4.4.1.1 and 4.3.3 similarly  
234 moved more than 100 times between the urban centers (Fig. 3).

235 To test whether Asunción and Ciudad del Este served as sources for *M. tuberculosis*, exporting  
236 infection elsewhere, we compared rates of arrival and export of each sub-lineage. Sub-lineage 4.1.2.1  
237 moved more frequently Asunción to Ciudad del Este (mean: 75 transitions) compared to vice versa  
238 (mean: 70 transitions), and a model for asymmetric rates was supported ( $\chi^2(2)=4.1, p=0.04$ ) (Fig. 3).  
239 Both sub-lineages 4.4.1.1 (with the prevalent *ahpC* mutation) ( $\chi^2(2)=0.16, p=0.69$ ) and 4.3.3  
240 ( $\chi^2(2)=0.56, p=0.46$ ) had similar rates of migration to and from Ciudad del Este to Asunción. Despite the

241 geographic structure observed, there was not sufficient signal to infer a likely geographic source for any  
242 of the dominant sub-lineages.

243

#### 244 ***M. tuberculosis* genomic clusters span prisons and the general population**

245 We next explored evidence of recent *M. tuberculosis* transmission in Paraguay, including  
246 relatedness between *M. tuberculosis* infections within and outside of prisons. In a maximum likelihood  
247 phylogeny (Fig. 2), *M. tuberculosis* isolates sampled from incarcerated and non-incarcerated people are  
248 distributed across the tree and did not form distinct clades, indicating recent shared evolutionary history  
249 of isolates sampled from prisons and the community. However, sub-lineage was associated with  
250 incarceration status ( $\chi^2(22)=52.3, p<0.001$ ).

251 TB notifications were often attributable to recent transmission, with 78% percent (369/471) of all  
252 isolates falling within 26 genomic clusters (each including 2 to 159 isolates) defined by a 12-SNP  
253 threshold<sup>42</sup>. Isolates from incarcerated people were more frequently clustered (92.6%, 138/149), than  
254 those from formerly incarcerated (71.0%, 22/31,  $\chi^2(1)=10.1, p=0.001$ ) or never incarcerated people  
255 (71.8%, 209/291;  $\chi^2(1)=24.2, p <0.0001$ ), likely reflecting more recent transmission within prisons. We  
256 found a similar trend with a stricter threshold of 5 SNPs, with 45.4% (214/471) of all isolates in genomic  
257 transmission clusters, including 58.3% (87/149) isolates from incarcerated individuals, 45.2% (14/31)  
258 from those formerly incarcerated, and 38.8% (113/291) from never incarcerated individuals.

259 We predicted that if prison and community-associated epidemics were distinct, isolates from the  
260 community would be most closely related to and cluster with other isolates from the community and vice  
261 versa. Approximately half (48.0%; 12/25) of genomic clusters including people with no incarceration  
262 history also included individuals with a recent history of incarceration. The consequence is that 85.2%  
263 (178/209) of individuals with evidence of recent transmission and no recent incarceration were within  
264 transmission clusters including individuals with prior incarceration.

265 We additionally quantified *M. tuberculosis* recent transmission with time-scaled haplotype  
266 diversity, a measure of the centrality of a single tip isolate to all other isolates on the tree<sup>52</sup>. Individuals

267 who were incarcerated at the time of TB notification had a higher time-scaled haplotype index for a short  
268 epidemic timescale (median: 0.59, IQR: 0.24-0.72) than did formerly (median: 0.18, IQR: -0.37-0.71;  
269  $t(36)=1.7$ ,  $p=0.03$ ) or never incarcerated individuals (median: 0.20, IQR: -0.71-0.66;  $t(360)=5.9$ ,  
270  $p<0.001$ ) (Fig. S3). This finding was consistent across epidemic timescales considered (Fig. S3). After  
271 adjusting for population structure, we found that incarceration status was significantly associated with  
272 time-scaled haplotype diversity (one-way ANOVA:  $F(2,85)=85$ ,  $p<0.001$ ), evidence that the association  
273 was independent of TB lineage.

274

### 275 **Recent growth in *M. tuberculosis* transmission clusters**

276 We additionally reconstructed population size dynamics of the three largest genomic clusters with  
277 a Bayesian coalescent population size model. The three largest genomic clusters, including 159, 91, and  
278 15 samples, increased in effective population size by 200, 90, and 40-fold respectively. Cluster growth  
279 was relatively recent, with cluster most recent common ancestors (MRCA) occurring in 1998 (95% HDI:  
280 1994-2001), 1996 (95% HDI: 1991-2000), and 1998 (95% HDI: 1992-2003) respectively, to 2021, when  
281 the most recent samples were collected (Fig. 4).

282

### 283 **Emergence of a putative resistance-associated *ahpC* promoter mutation**

284 Eleven percent of samples (50/471) shared a mutation in *ahpC* promoter (G>A, 74 bases  
285 upstream of the 5' start codon), previously considered a location for compensatory mutations co-  
286 occurring with *katG* mutations in isoniazid resistant isolates<sup>35,55</sup>. While *ahpC* promoter mutations are not  
287 included as an independent resistance-conferring mutation in the WHO resistance catalogue<sup>35</sup>, in our  
288 collection, *ahpC* mutations occurred on otherwise susceptible genomic background within sub-lineage  
289 4.4.1.1. The *ahpC* mutation occurred in a monophyletic clade of 49 samples in sublineage 4.4.1.1 (Fig. 5),  
290 which shared a most recent common ancestor in 1903 (95% HDI: 1888-1916), likely reflecting a single  
291 emergence event. Among the basal group of nine samples without a fixed *ahpC* promoter locus (*ahpC*-74)  
292 mutation, one sample was polymorphic, with 16% (13/79) of reads representing the *ahpC* mutation.

293 Among the samples sharing the *ahpC* mutation, a single isolate had a co-occurring rifampicin resistance-  
294 conferring mutation in *rpoB* (His445Leu) (Fig. 5).

295 We tested whether the success of the *ahpC* mutation in the absence of a *katG* mutation (i.e.  
296 outside of a compensatory context) we observed in Paraguay could be explained by an increase in  
297 associated transmissibility. The *ahpC* mutation was not associated with an increased time-scaled  
298 haplotype density (*ahpC* mutants, median 0.19, IQR: 0.09- 0.22; *ahpC* non-mutants, median: 0.56, IQR: -  
299 0.44-0.71,  $p = 0.93$ ). Further, individuals with an incarceration history (currently or formerly  
300 incarcerated) were no more likely to be infected with a *M. tuberculosis* isolate with the *ahpC* mutation  
301 than were individuals with no incarceration history ( $\chi^2(1)=0.25, p=0.62$ ).

302

### 303 Discussion

304 We generated, to our knowledge, the first genomic portrait of circulating *M. tuberculosis*  
305 diversity and transmission dynamics to directly inform Paraguay's national TB control program priorities.  
306 We found the majority of TB cases included in our study were attributable to recent transmission,  
307 including elevated rates of recent transmission within prisons and transmission clusters spanning prisons  
308 and surrounding communities that have expanded over the past twenty years. While *M. tuberculosis* is  
309 geographically structured in Paraguay, we identified a signal of continuous spread of *M. tuberculosis*  
310 between Paraguay's major urban centers.

311 We found that most sampled infections were likely attributable to recent transmission rather than  
312 long-distance migration or activation of latent disease, similar to what has been reported in other medium-  
313 incidence countries<sup>14</sup>. Consistent with expectations that clustering rates may correlate with incidence,  
314 when applying a 5-SNP threshold, we found that isolates from Paraguay were more frequently clustered  
315 (45%) than those from a low-incidence setting in Spain (23%) and less frequently clustered than in a  
316 high-incidence setting in Mozambique (58%). Interestingly, we found a higher rates of clustering  
317 compared to what was reported in Malawi (36%), a high-incidence setting<sup>11</sup>. This could reflect the

318 shorter, one year sampling timeframe of the Malawi study<sup>11</sup>, resulting in different genomic sampling  
319 rates, the use of different genomic sequencing pipelines, or true differences in transmission.

320 Paraguay's incarceration rate has dramatically increased, from 60 per 100,000 people in 2000 to  
321 194 per 100,000 in 2020<sup>3,6,56</sup>. More than seventy percent of the incarcerated population are pre-trial  
322 detainees, the highest proportion in South America<sup>3</sup>. The inhumane and unhealthy conditions of prison  
323 environments put people at heightened risk of disease and mortality; this risk translates into an increasing  
324 concentration of TB within prisons, with 18% (537/2,593) of notified TB cases in Paraguay occurring  
325 among incarcerated individuals in 2020<sup>6</sup>. Paraguay's TB Control Program has worked in prisons since  
326 2004 to provide trainings for healthcare providers and all diagnostic and treatment supplies, including  
327 laboratory capacity for microbiological testing in four prisons.

328 Our findings highlight the critical need to expand and strengthen existing programs to detect and  
329 treat TB early and to expand awareness and knowledge of the risks associated with prison environments.  
330 Isolates sampled from prisons were more frequently found in genomic transmission clusters and had a  
331 higher time-scaled haplotype density than did isolates from outside prisons, phylogenetic evidence that  
332 recent transmission was more frequent in prisons than in communities outside prisons. Further, *M.*  
333 *tuberculosis* sampled from prisons and the community were closely evolutionarily related and the  
334 majority of putative transmission clusters including individuals who were never incarcerated also  
335 included people who had a recent incarceration history, indicating that reducing transmission risk within  
336 prisons is an urgent public health priority with consequences both within and outside prisons.

337 While rates of drug resistance were relatively low, we found several phylogenetically unique  
338 mutations associated with both isoniazid and rifampicin resistance. These unique mutations could reflect  
339 either the *de novo* acquisition of a resistance mutation or the importation of a resistance mutation from  
340 outside Paraguay. Regardless, there is a critical need for expanding drug-susceptibility testing including  
341 both rapid testing for rifampicin resistance in addition to isoniazid monoresistance are critical to ensure  
342 patients are put on correct treatment courses and to reduce the risk of further resistance acquisition<sup>19,20</sup>.

343           The emergence of an independent *ahpC* mutation within a single sublineage opens questions  
344 about its phenotypic consequences. Previous studies in laboratory strains have reported that *ahpC*  
345 mutations are compensatory in the context of *katG* isoniazid resistance-conferring mutations, by  
346 recovering the bacterium’s ability to detoxify organic peroxides, but did not find measurable isoniazid  
347 resistance conferred by independent *ahpC* mutations<sup>55</sup>. Genome-wide association studies of clinical *M.*  
348 *tuberculosis* isolates confirmed the compensatory role of *ahpC* mutations<sup>57</sup>. *ahpC* mutations did not meet  
349 the criteria for being included in the 2021 WHO catalogue of resistance-conferring mutations because  
350 they were either too rare or had a low positive predictive value for isoniazid resistance as an independent  
351 mutation<sup>35</sup>. However, a study of isoniazid resistant isolates from Brazil reported that while *ahpC*  
352 mutations often co-occurred with *katG* mutations, they were also found in the absence of known  
353 resistance mutations in *katG* or *inhA*<sup>58</sup>.

354           A previous genome-wide survival analysis identified lineages and specific mutations associated  
355 with pre-resistance, genomic backgrounds that had a heightened likelihood of acquiring resistance-  
356 conferring mutations<sup>59</sup>. Whether *ahpC* acts as in a similar way, generating a “pre-compensated” genomic  
357 background, increasing the likelihood of future *katG* mutations, remains unknown.

358           Our study has several limitations. First, while we sequenced all available *M. tuberculosis*  
359 cultures, our final sample size of *M. tuberculosis* genomes was small relative to the number of notified  
360 TB cases in our study departments over the study period. Some locally circulating genotypes may  
361 therefore not be included in our sample and may lead to an underestimate of the contribution of recent  
362 transmission to incident TB. However, we sampled over a moderately long timeframe (five years) and  
363 included samples from high-incidence prisons and neighboring communities, providing greater  
364 opportunity to recover transmission events. Second, surveillance focused on Paraguay’s urban centers,  
365 where the majority of TB notifications occur. Future *M. tuberculosis* genomic surveillance in the Chaco,  
366 western Paraguay, where incidence is three times higher than in eastern Paraguay<sup>15</sup>, is needed.  
367 Additionally, further analysis at the regional-level will be critical for understanding transmission between  
368 Paraguay and neighboring countries. Third, we sampled TB infections from prisons at a higher rate than

369 infections outside prisons, potentially biasing upwards estimates of the rate of genomic clustering within  
370 prisons compared to outside prisons. Finally, we sequenced isolates from cultured sputum, as is routinely  
371 done for *M. tuberculosis* genomic epidemiology, but which limits the within-host diversity recovered  
372 from an individual's infection. Future research is needed to develop sequencing approaches to recover  
373 within-host *M. tuberculosis* variation and incorporate this level of variation into transmission and  
374 ancestral state reconstruction.

375         Together, our results underscore an urgent need for TB control measures to interrupt ongoing  
376 transmission in Paraguay, particularly in high-incidence prison settings, which have an outsized role in  
377 broader transmission. Further, the connectivity of Paraguay's urban centers indicates that TB control  
378 needs to be coordinated country-wide.

379

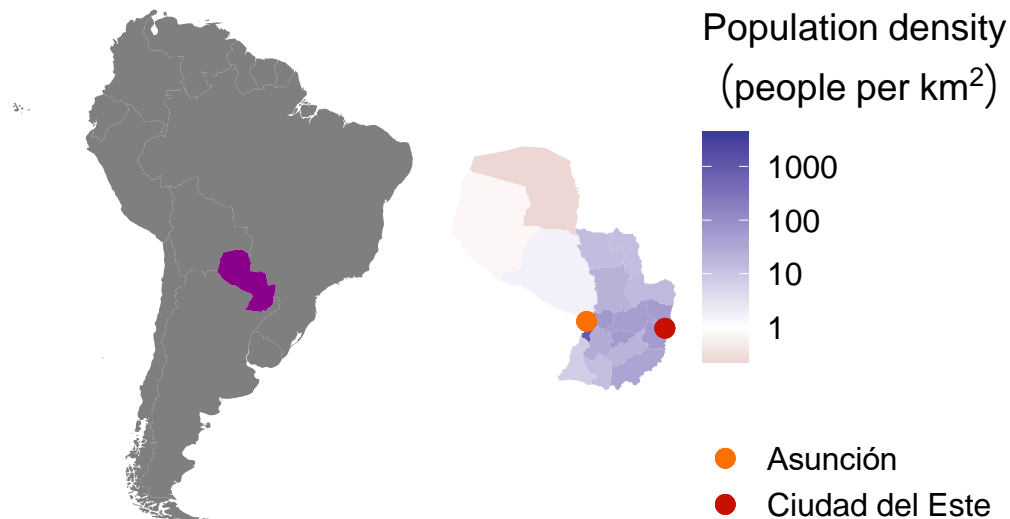
### 380 **Acknowledgments**

381         The authors thank the following health staff who facilitated the field work laboratory analyses  
382 and clinical follow up: Ruth Martinez, Nestor Moreno, Natalia Sosa. We also would like to express our  
383 gratitude to Pilar Muñoz, Johana Monteserin, CEDIC, LCSP and the Penitentiary Health Department  
384 from the Ministry of Justice for its support and advice for this project. The study was supported by the  
385 grant PIN15-705 from National Commission of Science and Technology (CONACYT) of Paraguay.

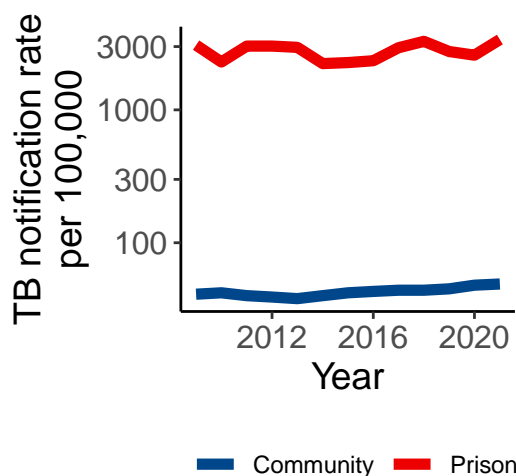
386



**A**



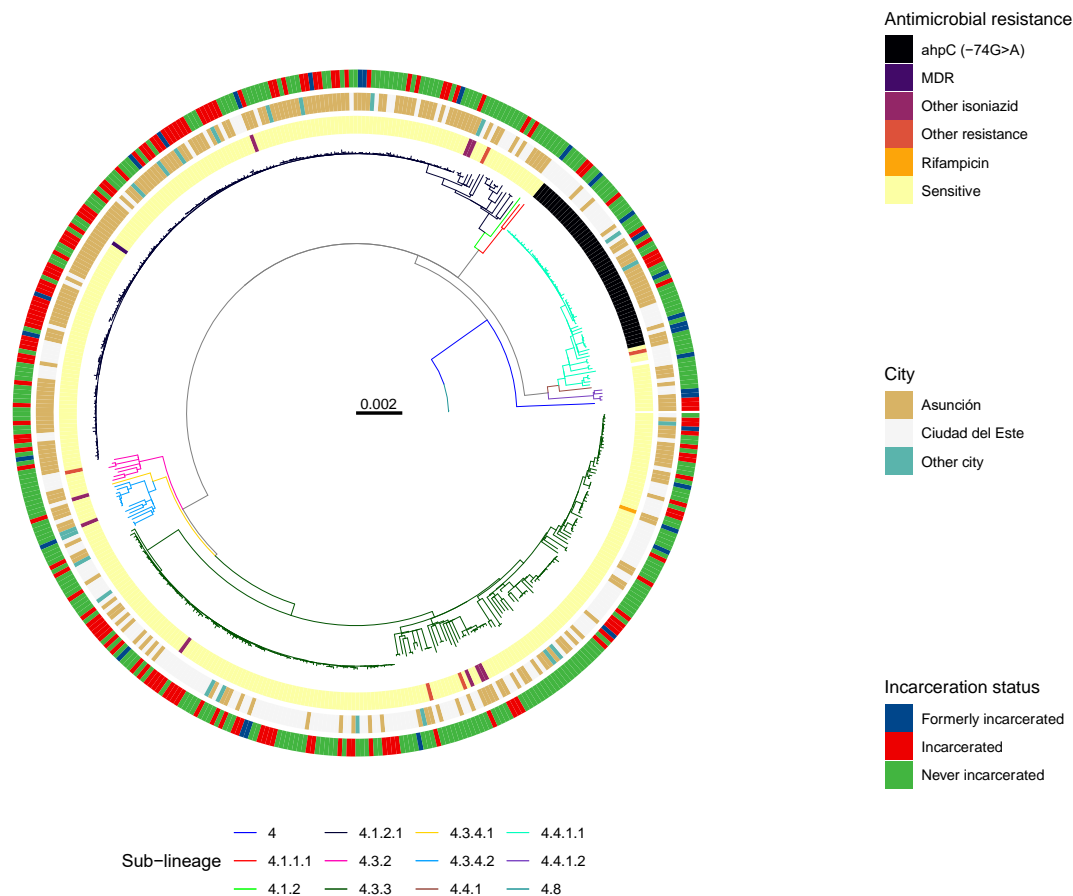
**B**



387

388 **Figure 1. Genomic surveillance within and outside prisons in Paraguay's urban centers.** a) Map of  
389 South America, with Paraguay highlighted, and as an inset. Paraguay's departments are colored by  
390 population size density and points indicate the two largest urban centers in Paraguay, where we conducted  
391 focused genomic surveillance. (b) Notification rate of TB per 100,000 people in prisons and in the general  
392 population from 2009 to 2020.

393



394

395 **Figure 2. *M. tuberculosis* isolates from incarcerated and non-incarcerated people are closely related**

396 **across Paraguay.** A maximum likelihood phylogeny of 471 tuberculosis isolates from Lineage 4 inferred

397 from a multiple sequence alignment of 9,966 SNPs and rooted on three sub-lineage 4.8 samples from this

398 study. Branch lengths are in units of substitutions per site. Branches are colored by sub-lineage. From the

399 inside, rings are colored by antimicrobial resistance; city of sampling; and incarceration status at the time

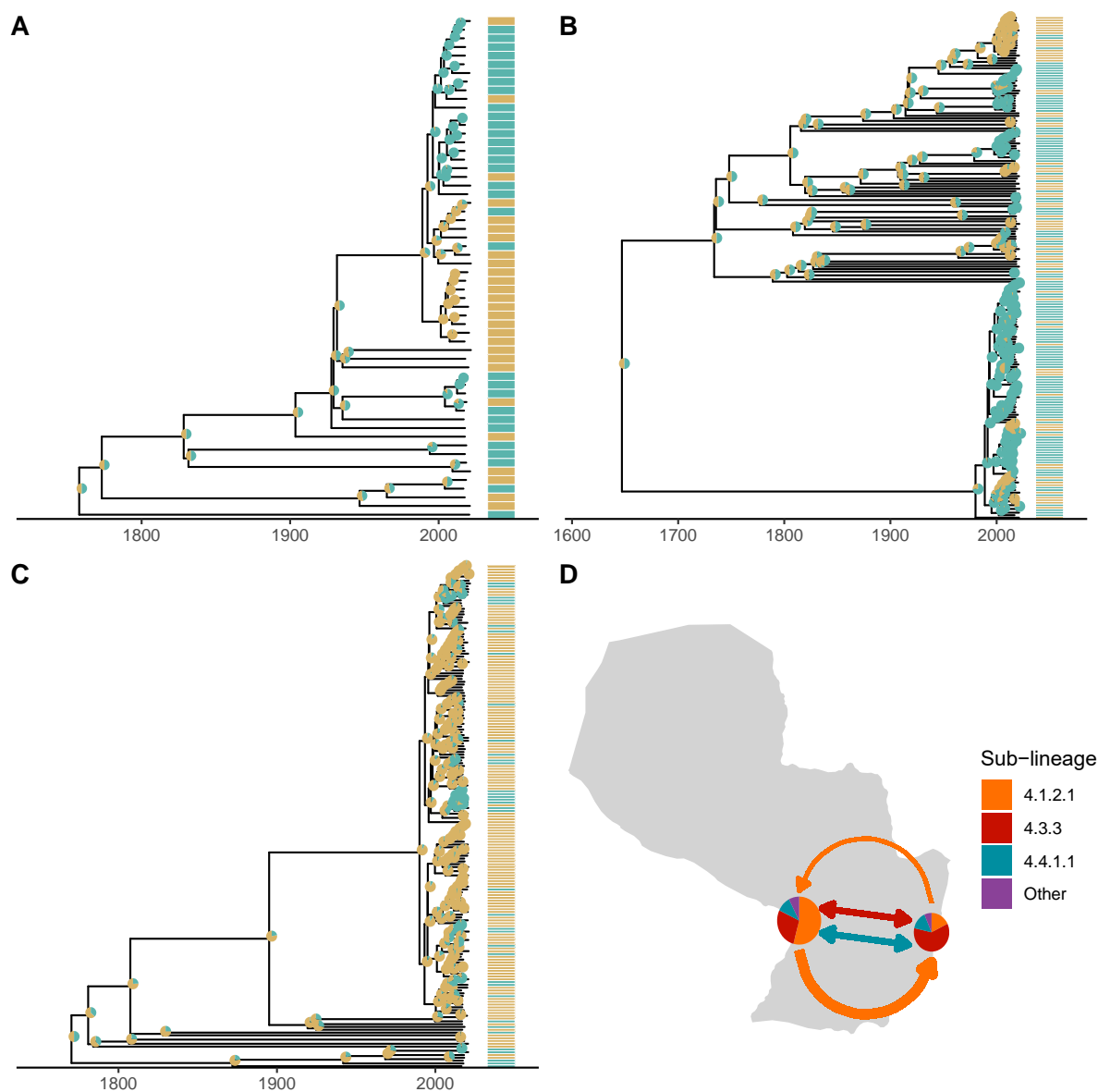
400 of TB notification. Other isoniazid includes isoniazid mono-resistant isolates without an *ahpC* promoter

401 mutation. Rifampicin indicates rifampicin mono-resistant isolates. Other resistance includes isolates with

402 mutations associated with resistance to pyrazinamide, streptomycin, or fluoroquinolones and not isoniazid

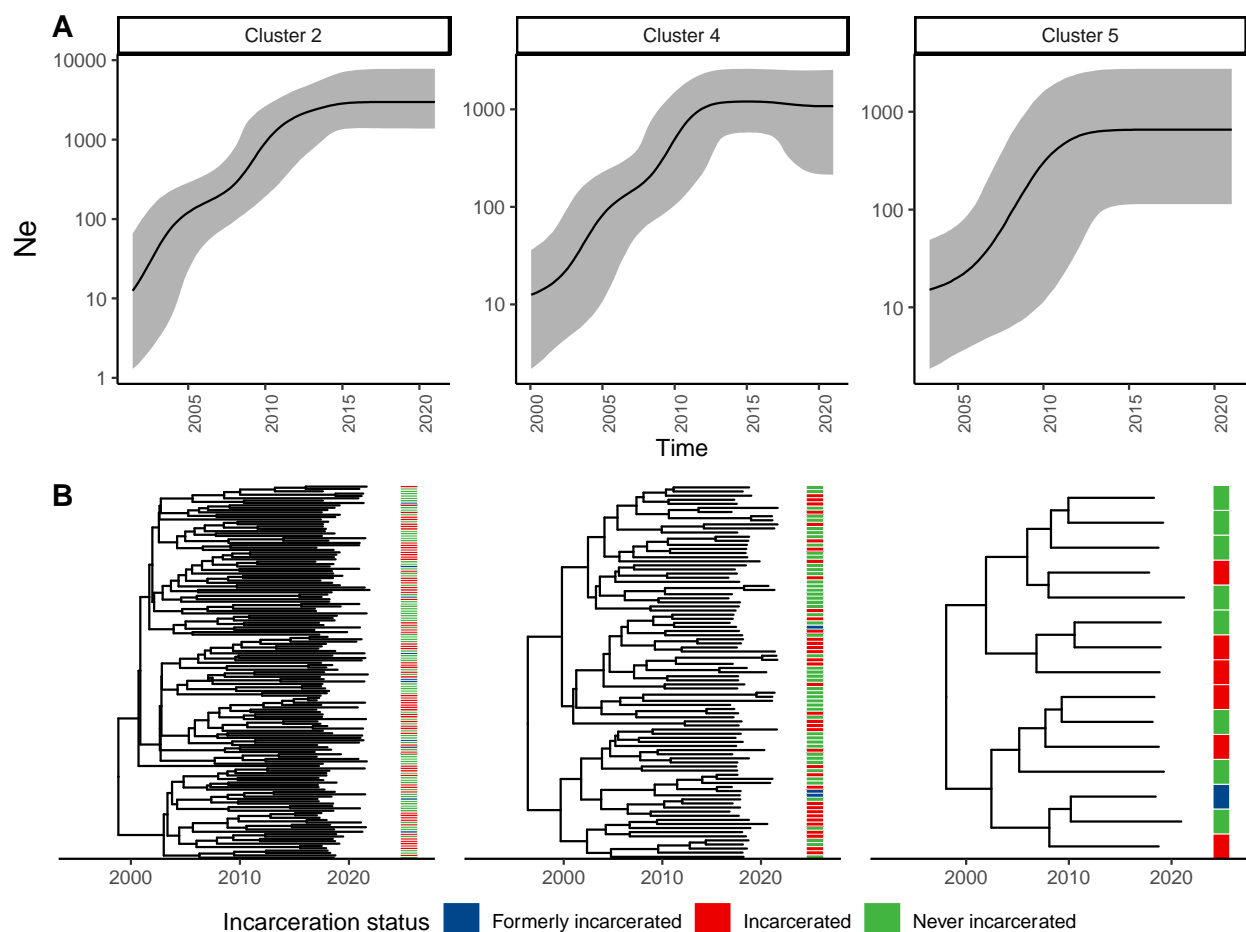
403 or rifampicin.

404

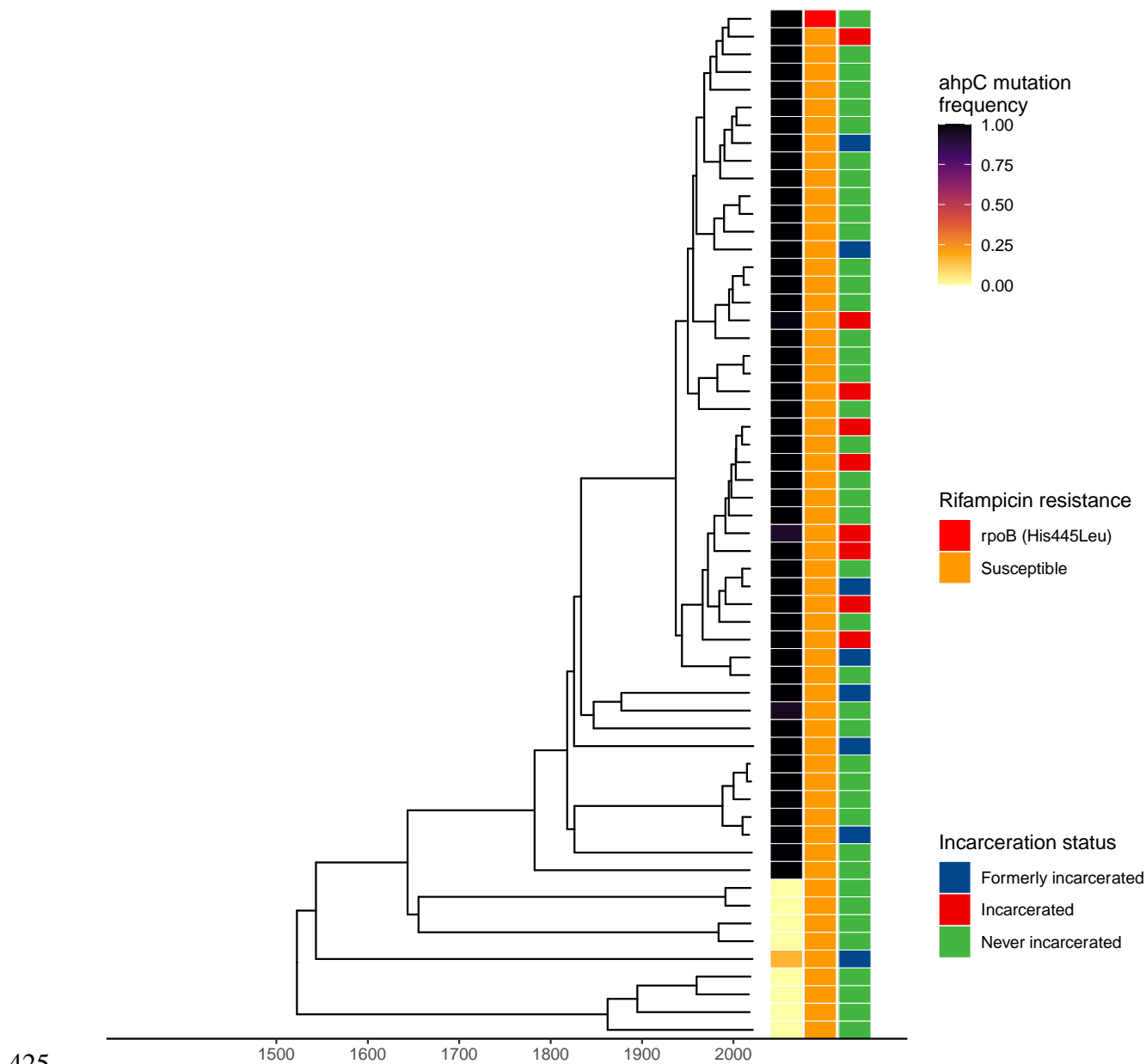


405  
406 **Figure 3. Frequent gene flow of *M. tuberculosis* connects Paraguay's major urban centers.** (a-c) We  
407 used discrete ancestral state reconstruction to reconstruct migration between the two cities for the three  
408 dominant sub-lineages in our sample; lineages 4.1.2.1, 4.3.3, and 4.4.1.1. Bayesian maximum clade  
409 credibility trees of samples in the three dominant sub-lineages with tip points colored by city of sampling  
410 and pie charts at nodes indicating the inferred ancestral location. Branch lengths are in years and grey bars  
411 indicate 95% high posterior density estimates of node date. (d) Map of Paraguay with pie charts  
412 indicating the genomic diversity sampled in Asunción (at Paraguay's western border) and Ciudad del Este

413 (eastern border). Arrows are colored by sub-lineage and are weighted by the relative rate of migration  
414 between cities. Bi-directional arrows indicate equal rates of migration in each direction.  
415  
416



417  
418 **Figure 4. Genomic transmission clusters spanning prisons and neighboring communities have**  
419 **recently expanded.** (a) Effective population size ( $N_e$ ) estimates for the three largest genomic clusters in  
420 our sample over time. Black lines indicate  $N_e$  inferred in a Bayesian Skyline Coalescent model and grey  
421 shading indicates 95% high posterior density estimates. (b) Median clade credibility trees inferred from  
422 the Bayesian Skyline Coalescent model. Branch lengths are in years and grey bars indicate 95% high  
423 posterior density estimates of node date. The heatmap to the right of the phylogeny indicates patient  
424 incarceration status at the time of TB notification.



**Figure 5. Emergence of a putative resistance-associated *ahpC* promoter mutation.** Time-scaled Bayesian maximum clade credibility tree for 58 samples in sublineage 4.4.1.1. Branch lengths are in years and grey bars indicate 95% high posterior density estimates of node date. The heatmap to the right of the phylogeny indicates patient incarceration status at the time of TB notification, *ahpC* mutation frequency within an individual's infection, and the occurrence of a rifampicin resistance-conferring mutation (*rpoB* His445Leu).

- 433 1. The World Health Organization. *Global Tuberculosis Report 2021*.  
434 <https://www.who.int/publications/digital/global-tuberculosis-report-2021> (2021).
- 435 2. World Bank: Poverty and Inequality Platform (Institution/Organization). *Poverty headcount ratio*  
436 *at national poverty lines (% of population) - Paraguay*. (2022).
- 437 3. Institute for Criminal Policy Research. *World Prison Brief*. <http://www.prisonstudies.org/> (2019).
- 438 4. Walter, K. S. et al. The escalating tuberculosis crisis in central and South American prisons. *The*  
439 *Lancet* vol. 397 1591–1596 (2021).
- 440 5. Cords, O. et al. Incidence and prevalence of tuberculosis in incarcerated populations: a systematic  
441 review and meta-analysis. *Lancet Public Heal.* **6**, 300–308 (2021).
- 442 6. Sequera, V. G. et al. Increased incarceration rates drive growing tuberculosis burden in prisons  
443 and jeopardize overall tuberculosis control in Paraguay. *Sci. Reports 2020 101* **10**, 1–7 (2020).
- 444 7. Organización Panamericana de la Salud Programa Regional de Tuberculosis  
445 (Institution/Organisation). *VI Reunión Regional: Avances y desafíos del control de la TB en*  
446 *Poblaciones Privadas de Libertad (PPL)*. (2013).
- 447 8. Churchyard, G. et al. What We Know about Tuberculosis Transmission: An Overview. *J. Infect.*  
448 *Dis.* **216**, S629–S635 (2017).
- 449 9. Stucki, D. et al. Mycobacterium tuberculosis lineage 4 comprises globally distributed and  
450 geographically restricted sublineages. *Nat. Genet.* **48**, 1535–1543 (2016).
- 451 10. Brynildsrud, O. B. et al. Global expansion of Mycobacterium tuberculosis lineage 4 shaped by  
452 colonial migration and local adaptation. *Sci. Adv.* **4**, 5869–5886 (2018).
- 453 11. Guerra-Assunção, J. et al. Large-scale whole genome sequencing of M. tuberculosis provides  
454 insights into transmission in a high prevalence area. *Elife* **4**, 1–17 (2015).
- 455 12. Ektefaie, Y., Dixit, A., Freschi, L. & Farhat, M. R. Globally diverse Mycobacterium tuberculosis  
456 resistance acquisition: a retrospective geographical and temporal analysis of whole genome  
457 sequences. *The Lancet Microbe* (2021) doi:10.1016/S2666-5247(20)30195-6.
- 458 13. Gygli, S. M. et al. Prisons as ecological drivers of fitness-compensated multidrug-resistant

- 459 Mycobacterium tuberculosis. *Nat. Med.* **27**, 1171–1177 (2021).
- 460 14. Walter, K. S. et al. The role of prisons in disseminating tuberculosis in Brazil: A genomic  
461 epidemiology study. *Lancet Reg. Heal. - Am.* **9**, 100186 (2022).
- 462 15. Candia, N. et al. First insight into Mycobacterium tuberculosis genetic diversity in Paraguay. *BMC*  
463 *Microbiol.* **7**, 1–10 (2007).
- 464 16. Woodman, M., Haeusler, I. L. & Grandjean, L. Tuberculosis genetic epidemiology: A latin  
465 american perspective. *Genes (Basel)*. **10**, (2019).
- 466 17. Walter, K. et al. The escalating tuberculosis crisis in Central and South American prisons. *Lancet*.
- 467 18. Atima, F. et al. Prevalence of multidrug-resistant tuberculosis in Latin America and the Caribbean:  
468 a systematic review and meta-analysis. (2020) doi:10.1111/tmi.13453.
- 469 19. Dean, A. S. et al. Prevalence and genetic profiles of isoniazid resistance in tuberculosis patients: A  
470 multicountry analysis of cross-sectional data. *PLOS Med.* **17**, e1003008 (2020).
- 471 20. Sulisid, G. & Paiid, M. Isoniazid-resistant tuberculosis: A problem we can no longer ignore.  
472 (2020) doi:10.1371/journal.pmed.1003023.
- 473 21. Kudoh, S. & Kudoh, T. A simple technique for culturing tubercle bacilli. *Bull. World Health*  
474 *Organ.* **51**, 71–82 (1974).
- 475 22. Palaci, M. et al. Contribution of the Ogawa-Kudoh swab culture method to the diagnosis of  
476 pulmonary tuberculosis in Brazil. *Int. J. Tuberc. Lung Dis.* **17**, 782–786 (2013).
- 477 23. Schiebelhut, L. M., Abboud, S. S., Omez Daglio, L. E. G. & Swift, H. F. A comparison of DNA  
478 extraction methods for high-throughput DNA analyses. (2016) doi:10.1111/1755-0998.12620.
- 479 24. Verma, R. et al. Detection of M. tuberculosis in the environment as a tool for identifying high-risk  
480 locations for tuberculosis transmission. *Sci. Total Environ.* 156970 (2022)  
481 doi:10.1016/J.SCITOTENV.2022.156970.
- 482 25. Krueger, F. Trim Galore. (2019).
- 483 26. Martin, M. Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads.  
484 *EMBnet.journal* **17**, (2011).

- 485 27. Wood, D. E. & Salzberg, S. L. Kraken: Ultrafast metagenomic sequence classification using exact  
486 alignments. *Genome Biol.* **15**, R46 (2014).
- 487 28. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform.  
488 *Bioinformatics* **25**, 1754–1760 (2009).
- 489 29. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Sambamba: fast processing of  
490 NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
- 491 30. Van der Auwera, G. A. & O’Connor, B. *Genomics in the cloud : using Docker, GATK, and WDL*  
492 *in Terra. Genomics in the cloud : using Docker, GATK, and WDL in Terra* (O’Reilly Media,  
493 2020).
- 494 31. Danecek, P. et al. Twelve years of SAMtools and BCFtools. *Gigascience* **10**, 1–4 (2021).
- 495 32. Brites, D. et al. A new phylogenetic framework for the animal-adapted mycobacterium  
496 tuberculosis complex. *Front. Microbiol.* **9**, 2820 (2018).
- 497 33. Phelan, J. E. et al. Integrating informatics tools and portable sequencing technology for rapid  
498 detection of resistance to anti-tuberculous drugs. *Genome Med.* **11**, 41 (2019).
- 499 34. Coll, F. et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome  
500 sequences. *Genome Med.* **7**, (2015).
- 501 35. World Health Organization. *Catalogue of mutations in Mycobacterium tuberculosis complex and*  
502 *their association with drug resistanc.* <https://www.who.int/publications/i/item/9789240028173>  
503 (2021).
- 504 36. Page, A. J. et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments.  
505 *Microb. Genomics* **2**, 1–5 (2016).
- 506 37. Paradis, E. & Schliep, K. Ape 5.0: An environment for modern phylogenetics and evolutionary  
507 analyses in R. *Bioinformatics* **35**, 526–528 (2019).
- 508 38. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., Von Haeseler, A. & Jermin, L. S.  
509 modelfinder: fast model selection for accurate phylogenetic estimates. **14**, (2017).
- 510 39. Minh, B. Q. et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in



- 511 the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
- 512 40. Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: Improving  
513 the Ultrafast Bootstrap Approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
- 514 41. Walker, T. M. et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis  
515 outbreaks: A retrospective observational study. *Lancet Infect. Dis.* **13**, 137–146 (2013).
- 516 42. Borgdorff, M. W. & Van Soolingen, D. The re-emergence of tuberculosis: what have we learnt  
517 from molecular epidemiology? *Clin. Microbiol. Infect.* **19**, 889–901 (2013).
- 518 43. Walker, T. M. et al. Assessment of Mycobacterium tuberculosis transmission in Oxfordshire, UK,  
519 2007–12, with whole pathogen genome sequences: An observational study. *Lancet Respir. Med.* **2**,  
520 285–292 (2014).
- 521 44. Cervera, B. S., López, M. G., Oms, Á. C.-, García, A. M. & Muñoz, Irving Cancino-, Torres-  
522 Puentes, Manuela, Villamayor, Luis, Madrazo-Moya C, Mambuque, Edson, Sequera, Victor,  
523 Respeito, Durval, Blanco, Silvia, A. Fine- grain population structure and transmission patterns of  
524 Mycobacterium tuberculosis in southern Mozambique, a high TB/ HIV burden area. *Microb.*  
525 *Genomics* **accepted**, 844 (2022).
- 526 45. Bouckaert, R. et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary  
527 analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).
- 528 46. Yang, C. et al. Phylogeography and transmission of M. tuberculosis in Moldova: A prospective  
529 genomic analysis. *PLOS Med.* **19**, e1003933 (2022).
- 530 47. Menardo, F., Duchêne, S., Brites, D. & Gagneux, S. The molecular clock of mycobacterium  
531 tuberculosis. *PLoS Pathog.* **15**, (2019).
- 532 48. Drummond, A. J., Rambaut, A., Shapiro, B. & Pybus, O. G. Bayesian Coalescent Inference of Past  
533 Population Dynamics from Molecular Sequences. *Mol. Biol. Evol.* **22**, 1185–1192 (2005).
- 534 49. Bilderbeek, R. J. C. & Etienne, R. S. babette: BEAUti 2, BEAST2 and Tracer for R. *Methods*  
535 *Ecol. Evol.* **9**, 2034–2040 (2018).
- 536 50. Yu, G., Lam, T. T. Y., Zhu, H. & Guan, Y. Two Methods for Mapping and Visualizing Associated

- 537 Data on Phylogeny Using Ggtree. *Mol. Biol. Evol.* **35**, 3041–3043 (2018).
- 538 51. Yu, G. Using ggtree to Visualize Data on Tree-Like Structures. *Curr. Protoc. Bioinforma.* **69**, e96  
539 (2020).
- 540 52. Rasigade, J. P. et al. Strain-specific estimation of epidemic success provides insights into the  
541 transmission dynamics of tuberculosis. *Sci. Reports 2017 71* **7**, 1–12 (2017).
- 542 53. Fitzjohn, R. G. Diversitree: comparative phylogenetic analyses of diversification in R. *Methods*  
543 *Ecol. Evol.* **3**, 1084–1092 (2012).
- 544 54. Huelsenbeck, J. P., Nielsen, R. & Bollback, J. P. Stochastic Mapping of Morphological  
545 Characters. *Syst. Biol.* **52**, 131–158 (2003).
- 546 55. Sherman, D. R. et al. Compensatory *ahpC* Gene Expression in Isoniazid-Resistant *Mycobacterium*  
547 tuberculosis. *Science (80-. )*. **272**, 1641–1643 (1996).
- 548 56. Walter, K. S. et al. The escalating tuberculosis crisis in central and South American prisons.  
549 *Lancet* **397**, 1591–1596 (2021).
- 550 57. Coll, F. et al. Genome-wide analysis of multi- and extensively drug-resistant *Mycobacterium*  
551 tuberculosis. *Nat. Genet.* **50**, 307–316 (2018).
- 552 58. Silva, M. S. N. et al. Mutations in *katG*, *inhA*, and *ahpC* Genes of Brazilian Isoniazid-Resistant  
553 Isolates of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **41**, 4471–4474 (2003).
- 554 59. Torres Ortiz, A. et al. Genomic signatures of pre-resistance in *Mycobacterium tuberculosis*. *Nat.*  
555 *Commun.* **12**, (2021).
- 556
- 557
- 558

559 **Supporting information captions**

560

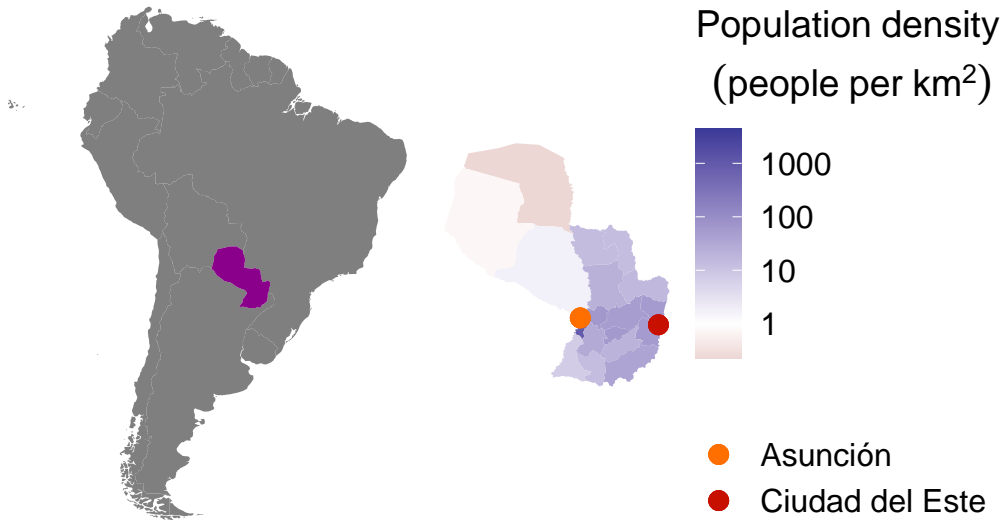
561 **Figure S1. Genomic *M. tuberculosis* surveillance in Paraguay.** Flowchart indicates the total number of  
562 notified cases of TB in Paraguay, from 2016 to 2021; TB cases in the three major urban departments of  
563 Paraguay; TB cases in urban departments stratified by incarceration status at the time of TB notification;  
564 and number of cases for which sequenced *M. tuberculosis* passed all quality filters.

565

566 **Figure S2. Longitudinal changes in sampled *M. tuberculosis* genomic diversity.** We  
567 compared genomic diversity in our study (2016-2021) with that sampled in the only previous  
568 genetic study of *M. tuberculosis* in Paraguay (Candia et al. 2003). Stacked bar plots indicate the  
569 proportion of samples falling in each clade. From left to right, panels indicate the total diversity  
570 sampled, samples from Asunción, samples from Ciudad del Este, and samples from the 2003  
571 study. Panels from the current study are stratified by incarceration status at the time of TB  
572 notification; the 2003 study did not present data stratified by incarceration status.

573

574 **Figure S3. Time-scaled haplotype density by incarceration status at the time of TB notification.** We  
575 measured time-scaled haplotype diversity, a measure of the centrality of a single tip isolate to all other  
576 isolates on the tree, an alternative proxy for recent transmission that considers not only the nearest  
577 phylogenetic neighbor, but all tree trips. We calculated time-scaled haplotype density from a matrix of  
578 pairwise SNP distances with the R package *thd* as previously described<sup>52</sup>. We set the *M. tuberculosis*  
579 substitution rate to  $1 \times 10^{-7}$  substitutions per site per year and included an effective genome length of  
580 3,916,441 basepairs (the length of the reference genome minus the PE/PPE regions excluded from variant  
581 calling) and used short (10 year) and long (20 year) epidemic timescales. Boxplots are colored by  
582 incarceration status at the time of TB notification. Boxes indicate the interquartile range, lines indicate  
583 median values, and whiskers indicate the range of the data. Points outside the whiskers indicate outliers.

**A****B**