Title

Genomic evidence for prisons as amplifiers of community tuberculosis epidemics

Short title: Prison amplifiers of tuberculosis epidemics

Authors

Katharine S. Walter PhD[1]*, Paulo César Pereira dos Santos BS[2], Thais Oliveira Gonçalves BS[3], Brunna Oliveira da Silva BS[2], Andrea da Silva Santos PhD[2], Alessandra de Cássia Leite MSc[2], Alessandra Moura da Silva BA[4], Flora Martinez Figueira Moreira MSc[2], Roberto Dias de Oliveira PhD[5], Everton Ferreira Lemos PhD[4], Eunice Cunha PhD[3], Yiran E. Liu BS[1,6], Albert Ko MD[7], Caroline Colijn PhD[8], Ted Cohen MD[7], Barun Mathema PhD[9], Julio Croda MD[4,7,10]†, Jason R. Andrews MD[1]†

†These authors contributed equally to this work.

*Corresponding author. Email: kwalter@stanford.edu

Affiliations

[1]Division of Infectious Diseases and Geographic Medicine, Stanford University School of Medicine; Stanford, United States.

[2] Health Sciences Research Laboratory, Federal University of Grande Dourados; Dourados, Brazil.

[3] Laboratory of Bacteriology, Central Laboratory of Mato Grosso do Sul; Campo Grande, Brazil.

[4] School of Medicine, Federal University of Mato Grosso do Sul, School of Medicine; Campo Grande, Brazil.

[5] Nursing Course, State University of Mato Grosso do Sul; Dourados, Brazil.

[6]Cancer Biology Graduate Program, Stanford University School of Medicine; Stanford, United States.

[7]Department of Epidemiology of Microbial Diseases, Yale School of Public Health; New Haven, United States.

[8]Department of Mathematics, Simon Fraser University; Burnaby, Canada.

[9]Department of Epidemiology, Columbia University Mailman School of Public Health; New York, United States.

[10] Mato Grosso do Sul Office, Oswaldo Cruz Foundation; Campo Grande, Brazil.

Abstract

Prisons are high-incidence settings for tuberculosis around the world, yet the contribution of spillover from prisons in driving community epidemics has not been quantified. *We* whole genome sequenced 1,152 *M. tuberculosis* isolates from participants diagnosed with tuberculosis within prisons and in the community in *Mato Grosso do Sul, Brazil from 2014 to 2019*. By integrating timed phylogenies and detailed location data, we reconstructed probabilistic transmission histories. *M. tuberculosis* sequences from incarcerated and non-incarcerated people were closely phylogenetically related. We found that 57% of recent community-wide tuberculosis cases were attributable to transmission from individuals with an incarceration history, 2.6% of the population. Further, we find genomic evidence that the prison system disseminates *M. tuberculosis* genotypes through frequent transfers across the state. This population-wide genomic transmission reconstruction framework can be applied to identify key environments amplifying infectious disease transmission to prioritize public health interventions.

Teaser

Prison spillover plays an outsized role in community tuberculosis transmission.

MAIN TEXT

Introduction

Tuberculosis remains one of the leading causes of death by an infectious disease and led to 1.4 million deaths in 2019.(1) Locating where *M. tuberculosis* transmission occurs is critical for public health interventions, yet in high-incidence settings, where most tuberculosis is transmitted remains unknown.*(1–6)* Prisons, hospitals, or mines are known to be high-risk environments*(7–9)* and may have epidemiological impacts that transcend their walls, serving as institutional amplifiers of infection.*(10, 11)* Identifying institutional amplifiers of tuberculosis is of particular concern in Central and South America, where the incarcerated population has increased by over

200% since 2000 and 11% percent of tuberculosis cases now occur among incarcerated people, who comprise <1% of the total population.*(12)*

The full impact of increasing incarceration rates on the tuberculosis epidemic in the Americas and elsewhere has been difficult to quantify. Many people are infected in prison, but only diagnosed later.*(11, 13, 14)* Further, spillover may occur when incarcerated people, prison staff, or visitors are infected within high-incidence prison environments and transmit *M. tuberculosis* onwards in communities outside prison. Previous studies have found that the excess risk of tuberculosis within prisons extends to surrounding communities*(15, 16)* and that *M. tuberculosis* genotypes responsible for jail and prison outbreaks are also found in the surrounding communities.*(13, 15, 17, 18)*

Pathogen genomes can be powerfully harnessed to infer high-resolution transmission histories, including who-infected-whom.*(19)* However, studies have not yet combined *M. tuberculosis* genomic and epidemiologic data to reconstruct transmission linkages between potential institutional amplifiers and the community, nor have they estimated the fraction of community transmission attributable to such spillover. To quantify the role of prisons in population-wide tuberculosis transmission, we conducted prospective genomic surveillance for tuberculosis in the Brazilian state of Mato Grosso do Sul and used full *M. tuberculosis* genomes, combined with detailed meta-data on incarceration history, to reconstruct tuberculosis transmission chains spanning prisons and the community.

Results

Tuberculosis and incarceration trends in Mato Grosso do Sul

Brazil's incarcerated population grew by 107% (from 361,402 to 748,009) from 2005 to 2019, an increase closely paralleled in the Central West state of Mato Grosso do Sul, where the incarcerated population has grown by 115% (7,891 to 16,976) over the same period (Fig. 1a-c). In

81 2019, prisons in the state were at 203% occupancy. While the state's notification rate of new and
82 retreatment tuberculosis in the general population (38 per 100,000) is similar to Brazil's national
83 tuberculosis notification rate (40 per 100,000) in 2019,*(20)* the notification rate within prisons
84 was 43 times as high (1,666 per 100,000), again similar to Brazil's national notification rate
85 within prisons (1,596 per 100,000) in 2019 (Fig. 1d,e).

86 **Population-based genomic surveillance across Mato Grosso do Sul state**

87 From 2014 to 2019, 3,491 new and retreatment cases of tuberculosis were reported in the
88 Campo Grande and Dourados, the two largest cities in Mato Grosso do Sul, Brazil. Of these,
89 1,249 had positive cultures and we sequenced 787 *M. tuberculosis* isolates. We sequenced
90 additional isolates from three other cities in the state and from earlier years in Dourados for a total
91 of 1,090 isolates from unique tuberculosis notifications (Fig. 1b). Whole genome sequences
92 (WGS) for 1,043 isolates met our coverage and quality criteria (Methods). We excluded 10%
93 (108/1043) of isolates with mixed infection, resulting in 935 high-quality genomes from 935
94 unique tuberculosis episodes from 918 individuals.

95 Of the 935 isolates in our analyses, 50% (465/935) were incarcerated; 16% (150/935)
96 were from patients who were formerly incarcerated; and 34% (320/935) of the study population
97 did not have an incarceration history at the time of tuberculosis notification. Among those who
98 did not have an incarceration history, 32 people reported contact with incarcerated individuals.
99 Additional population characteristics are in Table S1. Isolates were largely from *M. tuberculosis*
100 lineage 4 and predominantly fell into three sublineages: lineage 4.1, 4.3, and 4.4 (Fig. 2). Overall,
101 we found a low prevalence of antibiotic resistance across isolates, with 93.7% of isolates
102 susceptible to all antibiotics (Fig. 2).

103 **Phylogenetic structure of *M. tuberculosis* from prisons and the community**

104 A maximum likelihood phylogeny (Fig. 2) demonstrates that *M. tuberculosis* isolates
105 sampled from incarcerated and non-incarcerated people do not form distinct clades and are

106 closely phylogenetically related. Terminal branch lengths, a proxy for the degree of recent
107 transmission,*(21)* are significantly longer for isolates from people with no incarceration history at
108 the time of tuberculosis notification (9.4×10^{-4} substitutions/site) than isolates from incarcerated
109 individuals (2.2×10^{-4}; two-sample t-test p-value = 1.14×10^{-7}) or formerly incarcerated
110 individuals (3.8×10^{-4}; $p = 1.6 \times 10^{-4}$), suggesting that individuals with an incarceration history
111 were more likely to be recently infected (Fig. S1).

112 **Genetic clustering of *M. tuberculosis* from prisons and the community**

113 To identify potential clusters of recent transmission, we applied a commonly used 12-SNP
114 threshold,*(22)* including all isolates within the threshold distance of at least one other clustered
115 isolate. Eighty-three percent (777/935) of the isolates fell into 84 genomic clusters (each
116 including 2 to 170 isolates), providing evidence that notifications were largely attributable to
117 recent, local transmission rather than travel-associated importation or re-activation of genetically
118 distinct latent infections (Fig. 2). Isolates from incarcerated people were more frequently
119 clustered (93.3%, 434/465), than those from formerly incarcerated (86.0%, 129/150; $p < 0.0001$),
120 or never incarcerated people (66.9%, 214/320; $p < 0.0001$), again suggesting more recent
121 transmission within prisons.

122 We predicted that if prison and community-associated epidemics were distinct, isolates
123 from the community would be most closely related to and cluster with other isolates from the
124 community and vice versa. Of the 46 clusters with three or more isolates, including 703
125 participants, 40 included isolates from community members who had no reported incarceration
126 history. Of those potential transmission clusters, 85.0% (34/40) included isolates from people who
127 were currently or previously incarcerated.

128 We found a similar pattern of clustering of isolates from community members and
129 individuals with an incarceration history when using an alternative 5-SNP threshold for genomic
130 clustering*(23)*. With this threshold, of the 80 clusters with three or more isolates, 56 included

131 isolates from people with no incarceration history and 78.6% (44/56) also included isolates from
132 people with an incarceration history.

133 **A *M. tuberculosis* clone spans prisons and the community across the state**

134 The largest potential transmission cluster, including 170 isolates sampled from September
135 2010 to April 2019, was distributed across the state, including cases found across state's two
136 largest cities, Campo Grande and Dourados, as well as the smaller cities Corumbá (on the
137 Bolivian border) and Ponta Porã (on the Paraguayan border) (Fig. 1b, Fig. 3a). The cluster had a
138 most recent common ancestor in 1996 (95% HPD: 1989-2003; Fig. 3a), indicating that the drug-
139 susceptible clone had circulated for approximately 23 years before the most recent samples were
140 collected. 103 isolates were from people incarcerated at the time of diagnosis, 35 from people
141 formerly incarcerated, and 32 from people with no incarceration history at the time of tuberculosis
142 notification.

143 As observed in other clusters, isolates from people who are currently and formerly
144 incarcerated were closely genetically related—and often, identical—to isolates from people who
145 were not incarcerated at the time of notification, suggesting they are linked through recent
146 transmission (Fig. 3b).

147 We hypothesized that the criminal justice system's frequent transfer of people between
148 prisons and jails could disseminate TB across the state. Brazil's national criminal law defines
149 different "regimes" or stages of incarceration(30), which may facilitate the potential for spillover
150 and dissemination of infection. People are incarcerated under closed regimes, within prisons;
151 semi-open regimes, in which people may work outside prisons and return at night; and open
152 regimes, under which people serve sentences outside of prison, but are required to make periodic
153 court appearances.(30) We analyzed movements in the state's incarceration database and found
154 that the average duration of incarceration was 230 days in a prison and 25 days in a jail. The
155 incarceration database documents 7,982 mean yearly transfers between prisons (including closed

156 and semi-open prisons) from 2015-2018 in a prison population of 17,221 in August 2018,
157 including frequent transfers between cities across the state (Fig. 4a).

158 To investigate the role of the criminal justice system in disseminating infection, we
159 tracked the spread of the largest sampled *M. tuberculosis* cluster (Fig. 3) across prisons in Mato
160 Grosso do Sul (Fig. 4b). From early 2011 to 2014, sampled isolates from the cluster were
161 concentrated in the maximum security prison in Dourados, after which it was identified within
162 two prisons in Campo Grande and then exported to prisons and jails in other cities. The cluster
163 was not contained within the state's prison network; notified tuberculosis cases among
164 community members outside of prison occurred over the duration of the clusters' state-wide
165 spread.

166 **Transmission trees reveal who-infected-whom, where, and when transmission occurred**

167 We further investigated genomic clusters by integrating timed phylogenies with
168 epidemiological information to infer transmission trees, allowing us to reconstruct not only
169 transmission linkages of who-infected-whom, but also locate when and where transmission
170 occurred. Reconstructed transmission linkages revealed frequent transmission within prisons and
171 spillover from prisons to the community. For example, we inferred likely transmission (89%
172 posterior probability) between two individuals (Fig. 5a,b) and occurred between July 2014 and
173 July 2015 with 94% posterior probability (Fig. 5c). Both the recipient and predicted infector were
174 incarcerated in the same prison in Campo Grande at the time of the recipient's tuberculosis
175 notification and that transmission occurred within this prison with a posterior probability of 97%.

176 **Spillover from prisons plays a disproportionate role in community transmission**

177 To estimate the total contribution of different populations to community-wide tuberculosis
178 transmission, we inferred transmission trees for all clustered isolates. We identified the location of
179 both the infector and recipient over the distribution of likely infection times, allowing us to
180 determine location-specific transmission probabilities. Finally, we summed transmission

181 probabilities over all potential transmission pairs by the incarceration status of both the infector
182 and recipient at the time of transmission, giving a population-wide WAIFW matrix (Fig. 5d).
183 Because we conducted active surveillance inside prisons and not in the general population, we
184 corrected pairwise transmission probabilities to account for higher genomic sampling proportion
185 of incarcerated populations (Fig. S2, Methods).

186 The population-wide WAIFW matrix reveals that a small proportion of the state's
187 population, 0.84% (22,706 incarcerated people of a total population of 2,713,147 in 2017), were
188 associated with 18% of observed transmissions to people with no incarceration history (Fig. 5d).
189 When additionally including people who were previously incarcerated, 2.6% of the population
190 (71,703 of 2,713,147) were associated with 39% of the transmission to people with no
191 incarceration history (Fig. 5d), demonstrating the outsized impact of prisons on tuberculosis
192 transmission in the state. In total, we found that 57% of all infections captured in our study
193 (including incarcerated, formerly incarcerated and never incarcerated individuals), including 46%
194 of infections in the community (among formerly incarcerated and never incarcerated individuals),
195 were attributable to transmission from incarcerated or formerly incarcerated individuals, who
196 comprise a small fraction of the state's population.

197 **Transmission inferences are robust to epidemiological priors**

198 We conducted a sensitivity analysis to determine the effect of epidemiological priors
199 (Table S2) on transmission inferences for the largest genomic cluster (170 isolates). Transmission
200 probabilities inferred for pairs of individuals were closely correlated across generation time priors
201 (Pearson's r: 0.94 – 0.98, Fig. S3) and sampling time priors (Pearson's r: 0.84 – 0.98, Fig. S4),
202 although the total number of observed transmission events was strongly influenced by both priors
203 (Figure S5). This suggests that, as expected, epidemiological priors impact the magnitude of
204 inferred transmission probabilities rather than the identity of predicted transmission pairs. Further,
205 alternative priors did not strongly influence the population-wide WAIWF matrix for the largest

genomic cluster (Fig. S6). The fraction of transmission to those with no incarceration history attributable to spillover (from infectors who are currently or previously incarcerated) in the largest cluster ranges from 63.5% (medium sampling and generation times) to 72.3% (short generation time and long sampling time).

Discussion

Here, we find genomic evidence that spillover from a key, high-risk environment can drive tuberculosis transmission risk in the broader population. Controlling tuberculosis thus urgently requires reducing risk created by institutional amplifiers, which have outsized effects on the general population.*(11)* Our population-level transmission reconstruction framework can be applied to identify other high-risk environments or populations and quantify their contribution to community tuberculosis transmission, informing targeted public health interventions.*(11)*

The disproportionate role that prisons play in driving community tuberculosis transmission is not unique to Brazil. Incidence of tuberculosis in prisons is extremely high globally.*(7, 8)* The rapid growth in incarceration rates across Central and South America, in particular, puts an increasing population at heightened risk of tuberculosis—risk that we have found can extend into neighboring communities. Although the prevalence of antibiotic resistance is low in Central West, Brazil, prisons have also amplified drug-resistant tuberculosis in Russia*(24)* and elsewhere. Further, the role of prisons and other detention centers as disease reservoirs is not unique to tuberculosis. Previous studies have identified spillover of meningococcal disease*(25)* and SARS-CoV-2*(26)* from prisons to the community.

Our results emphasize that reducing the extraordinarily high transmission risk created by prison environments is an urgent priority. Public health programs need to work to reduce the extreme overcrowding within prisons; improve the unsanitary, often inhumane conditions of incarceration; and expand access to primary healthcare and nutrition. Tuberculosis control

programs should expand routine active screenings during*(11, 27)* and following incarceration so that cases can be diagnosed early and linked to treatment. Individuals leaving prison are at heightened risk of tuberculosis and may be a particularly vulnerable population. Further, preliminary evidence suggests that isoniazid protective therapy is protective against tuberculosis infection and should be considered as a protective intervention in high-risk prison environments.*(28)*

Reducing this excess burden of disease will require work that extends beyond biomedical interventions. The most direct way to mitigate the excess tuberculosis risk created by prisons is for governments to identify alternatives to detention so that incarceration is used only as a last resort. Prison releases have already begun as a means of reducing infectious disease transmission risk:*(29)* in an attempt to reduce the risk of COVID-19 in prisons, Brazil released more than thirty thousand people early by mid-2020.*(30)*

Our study has several limitations. First, while we were able to enroll a majority of *M. tuberculosis* isolates from incarcerated people with tuberculosis in the city of Dourados, overall sampling of the state's ongoing tuberculosis epidemic was incomplete. To minimize sampling bias, we determined genomic sampling probabilities and adjusted transmission probabilities to account for over-sampling of prisons. Our probabilistic approach allows us to account for uncertainty in transmission linkages; however, we still cannot infer the identity or incarceration history of unsampled people who contribute to transmission. Terminal branch lengths may similarly be affected by sampling bias or pathogen population size and are thus an imperfect proxy of recency of transmission.

Second, transmission trees reflect information from timed *M. tuberculosis* phylogenies and epidemiological priors. It is difficult to distinguish genomic sampling proportion from generation time and sampling time distributions. Because we conducted transmission inference in a tuberculosis-endemic setting, rather than in a discrete outbreak, our rate of genomic sampling

256 was lower than in some previous tuberculosis transmission studies. To reduce uncertainty in
257 sampling proportion, we conducted transmission inference on recent subtrees (MRCA > 2012) for
258 which we had sampled a greater proportion of isolates. To investigate the influence of
259 epidemiological priors on transmission inferences, we conducted a sensitivity analysis (Table S2)
260 and found that our conclusions were largely unchanged (Supplementary Text, Figs. S3-S6).

261 Third, we conducted whole genome sequencing of *M. tuberculosis* isolates cultured from
262 sputum and generated single consensus genomes from each participant. Excluding within-host
263 diversity could lead to some incorrect transmission inferences. For example, if we sampled an *M.*
264 *tuberculosis* genome that was genetically distant from the genome that was transmitted, we could
265 incorrectly exclude transmission.

266 Finally, we employed an existing bioinformatic approach to identify *M. tuberculosis*
267 variants and excluded repetitive genomic regions as is common practice.*(31, 32)* Future studies
268 that leverage within-host *M. tuberculosis* diversity and/or include the diverse PE/PPE genes could
269 provide greater resolution and reduce the uncertainty of transmission inferences.*(33)*

270 In this study, we present genomic evidence that prisons act as tuberculosis reservoirs in
271 Central West, Brazil. The dramatic expansion of incarceration in recent decades has put an
272 increasing population at extremely high risk of tuberculosis; this risk extends to surrounding
273 communities. Reducing the excess tuberculosis transmission risk within prisons and other
274 detention centers is an urgent public health priority. Our findings additionally highlight the
275 potential for pathogen genomic surveillance to identify other key environments or core
276 populations with a disproportionate role in pathogen transmission and guide public health
277 priorities.

278

279 **Materials and Methods**

280 **Study design**

281 We conducted population-based tuberculosis surveillance in the state of Mato Grosso do
282 Sul in Central West Brazil from January 2014 through May 2019. Surveillance included active
283 screening in three of the largest prisons in the state as well as ongoing passive surveillance
284 focused on the two largest cities in the state, Campo Grande and Dourados, and three cities at the
285 state's border with Paraguay and Bolivia (Fig. 1a,b; **Supplementary Methods**). All participants
286 provided written consent, and this study was conducted with the approval of the Research Ethics
287 Committee from the Federal University of Grande Dourados, Federal University of Mato Grosso
288 do Sul and National Research Ethics Committee (CONEP) (CAAE 37237814.4.0000.5160,
289 2676613.3.1001.5160, and 26620619.6.0000.0021) and Stanford University Institutional Review
290 Board (IRB-40285).

291 **Incarceration history**

292 To investigate the incarceration history of tuberculosis patients more closely, we obtained
293 permission from the Mato Grosso do Sul state prison administration agency, Agência Estadual de
294 Administração do Sistema Penitenciário, to access a database of all prison entries, exits, and
295 transfers within the criminal justice record system, Sistema Integrado de Gestão Operacional,
296 from January 1, 2005 through December 31, 2018. Brazil's national criminal law defines different
297 "regimes" or stages of incarceration*(34)*, which may facilitate the potential for spillover of
298 infection. People are incarcerated under closed regimes, within prisons; semi-open regimes, in
299 which people may work outside prisons and return at night; and open regimes, under which
300 people serve sentences outside of prison, but are required to make periodic court appearances.*(34)*

301 **Whole genome sequencing**

302 We sequenced whole genomes from cultured isolates on an Illumina NextSeq (2 x 151-
303 bp). Sequence data is available on the Sequence Read Archive (SRA), in BioProject
304 PRJNA671770. We applied variant calling methods closely following those described in Menardo

305 et al.*(31)* to be consistent with the methods used for molecular clock estimation (**Supplementary**
306 **Methods**).

307 **Phylogenetic and Bayesian evolutionary analysis**

308 We fit maximum likelihood trees with RAxML-ng 1.0.1.*(35)* For each cluster including
309 three or more isolates, we fit Bayesian trees with BEAST 2.6.2*(36)* with a strict clock, constant
310 coalescent population size model using tuberculosis notification dates to calibrate tips
311 (**Supplementary Methods**).

312 **Transmission inference**

313 *M. tuberculosis* phylogenetic trees represent patterns of evolutionary relatedness between
314 the consensus bacterial genomes sampled from different individuals. Because most outbreaks are
315 incompletely sampled (i.e. an *M. tuberculosis* sequence is not available for every case) and
316 individuals may be infected with diverse populations of *M. tuberculosis*, phylogenies do not
317 represent the underlying history of transmission.*(19)* We used TransPhylo*(19)* to infer
318 transmission linkages—including unsampled hosts—that were consistent with the underlying
319 timed phylogenies and transmission process (i.e. generation time and sampling intervals). To
320 focus on densely sampled trees that would not be dominated by unsampled hosts, we sliced
321 phylogenies at 2012 in order to generate subtrees with a most recent common ancestor of 2012 or
322 later for transmission inference, including 122 subtrees each with 2 to 23 tips and comprising
323 56% of isolates (528/935).

324 By sampling from the posterior distribution of transmission trees, we estimated the
325 posterior probability that each individual within a transmission tree was infected by each other
326 individual in the tree. We summarized the posterior set of transmission trees in TransPhylo with
327 who acquired infection from whom (WAIFW) matrices, W, where $W_{i,j}$ is the posterior probability
328 of transmission from individual j to individual i. Many transmission events are unobserved
329 because of incomplete sampling; therefore, the sum of transmission probabilities to any individual

is often less than one. We combined transmission trees with incarceration histories to quantify the proportion of observed transmission attributable to people with a history of incarceration. An individual's location at the time of their tuberculosis notification may not reflect where transmission occurred. To understand where infector-recipient pairs were at the time transmission occurred, we integrated transmission trees with incarceration information (Fig. 4c). First, we used the transmission trees to infer the posterior distribution of infection times for each recipient i. We then partitioned the infector-recipient transmission probability $W_{i,j}$ across the recipient's infection time distribution, resulting in W_{i_u,j_v}, the probability of transmission from individual j to individual i while j has incarceration status v and i has incarceration status u. Incarceration status was trichotomized as incarcerated, formerly incarcerated, and no incarceration history.

We then summed pairwise, location-specific transmission probabilities over all individual subtrees with respect to the incarceration status of both infector j and recipient i. We defined the population wide transmission matrix, $T_{u,v} = \sum_i \sum_j W_{i_u,j_v}$, where $T_{u,v}$ is the sum of the posterior probabilities of observed transmission events to individuals in population u from individuals in population v. We normalized $T_{u,v}$ by dividing by the sum of posterior transmission probabilities to each recipient population u, $N_u = \sum_v T_{u,v}$. We then defined the uncorrected proportion of the total posterior transmission probabilities to population u from population v, or the transmission fraction, as $p_{u,v} = \frac{T_{u,v}}{N_u}$.

Genomic sampling varied across populations and specifically, we predicted that incarcerated people might have been sampled at a higher rate due to our active case finding and greater culture coverage in three prisons in the state. To correct for sampling bias, we determined the genomic sampling proportion r_k, which we defined as the proportion of notified tuberculosis cases for which a high-quality genomic sequence was available, for population k at the time of tuberculosis notification. We calculated genomic sampling proportion as the number of sequenced genomes for each group divided by the number of tuberculosis notifications for each group from

2014-2019 in Campo Grande and Dourados, the two major cities in our prospective study (Fig. S2). We adjusted each pairwise transmission probability by the sampling rate of the infector, based on the incarceration status of the infector at the time of tuberculosis notification, for $j \in k$: $W'_{i,j} = W_{i,j} \frac{1}{r_k}$ and calculated the sampling-adjusted transmission fractions as described above.

Role of the funding source

The funding source had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Statistical methods

We inferred transmission trees with TransPhylo*(19)* as described above. We compared phylogenetic tree terminal branch lengths between populations with two-sample t-tests and compared the rate of clustering between populations with two-sample proportion tests.

References

1. G. Churchyard, P. Kim, N. S. Shah, R. Rustomjee, N. Gandhi, B. Mathema, D. Dowdy, A. Kasmar, V. Cardenas, What We Know about Tuberculosis Transmission: An Overview, *J. Infect. Dis.* **216**, S629–S635 (2017).

2. S. C. Auld, N. Sarita Shah, B. Mathema, T. S. Brown, N. Ismail, S. V. Omar, J. C. M. Brust, K. N. Nelson, S. Allana, A. Campbell, K. Mlisana, P. Moodley, N. R. Gandhi, Extensively drug-resistant tuberculosis in South Africa: Genomic evidence supporting transmission in communities, *Eur. Respir. J.* **52** (2018), doi:10.1183/13993003.00246-2018.

3. K. Middelkoop, B. Mathema, L. Myer, E. Shashkina, A. Whitelaw, G. Kaplan, B. Kreiswirth, R. Wood, L. G. Bekker, Transmission of tuberculosis in a south African community with a high prevalence of HIV infection, *J. Infect. Dis.* **211**, 53–61 (2015).

4. J. R. Andrews, C. Morrow, R. P. Walensky, R. Wood, Integrating social contact and

380 environmental data in evaluating tuberculosis transmission in a South African township., *J. Infect.*
381 *Dis.* **210**, 597–603 (2014).
382 5. T. A. Yates, P. Y. Khan, G. M. Knight, J. G. Taylor, T. D. McHugh, M. Lipman, R. G. White,
383 T. Cohen, F. G. Cobelens, R. Wood, D. A. J. Moore, I. Abubakar, The transmission of
384 Mycobacterium tuberculosis in high burden settings, *Lancet Infect. Dis.* **16**, 227–238 (2016).
385 6. J. R. Andrews, C. Morrow, R. Wood, Modeling the role of public transportation in sustaining
386 tuberculosis transmission in South Africa, *Am. J. Epidemiol.* **177**, 556–561 (2013).
387 7. K. Dolan, A. L. Wirtz, B. Moazen, M. Ndeffo-mbah, A. Galvani, S. A. Kinner, R. Courtney,
388 M. McKee, J. J. Amon, L. Maher, M. Hellard, C. Beyrer, F. L. Altice, Global burden of HIV,
389 viral hepatitis, and tuberculosis in prisoners and detainees, *Lancet* **388**, 1089–1102 (2016).
390 8. O. Cords, L. Martinez, J. Warren, J. O'Marr, K. S. Walter, T. Cohen, J. Zheng, A. I. Ko, J.
391 Croda, J. R. Andrews, Incidence and Prevalence of Tuberculosis in Incarcerated Populations: A
392 Systematic Review and Meta-analysis, *Lancet Public Heal.* .
393 9. J. Urrego, A. I. Ko, A. Da Silva Santos Carbone, D. S. G. Paião, R. V. E. Sgarbi, C. W. Yeckel,
394 J. R. Andrews, J. Croda, The impact of ventilation and early diagnosis on tuberculosis
395 transmission in Brazilian prisons, *Am. J. Trop. Med. Hyg.* **93**, 739–746 (2015).
396 10. S. Basu, D. Stuckler, M. McKee, Addressing institutional amplifiers in the dynamics and
397 control of tuberculosis epidemics, *Am. J. Trop. Med. Hyg.* **84**, 30–37 (2011).
398 11. T. S. Mabud, M. de Lourdes Delgado Alves, A. I. Ko, S. Basu, K. S. Walter, T. Cohen, B.
399 Mathema, C. Colijn, E. Lemos, J. Croda, J. R. Andrews, J. Z. Metcalfe, Ed. Evaluating strategies
400 for control of tuberculosis in prisons and prevention of spillover into communities: An
401 observational and modeling study from Brazil, *PLoS Med.* **16**, e1002737 (2019).
402 12. K. Walter, L. Martinez, Arakaki-Sanchez, V. Sequera, S. G. Estigarribia, T. Cohen, A. Ko, A.
403 García-Basteiro, Z. Rueda, O. R. Lopez, M. Espinal, J. Croda, J. Andrews, The escalating
404 tuberculosis crisis in Central and South American prisons. In press., *Lancet* (2020),

405 doi:10.1016/S0140-6736(20)32578-2.

406 13. F. P. C. Sacchi, R. M. Praça, M. B. Tatara, V. Simonsen, L. Ferrazoli, M. G. Croda, P. N.
407 Suffys, A. I. Ko, J. R. Andrews, J. Croda, Prisons as reservoir for community transmission of
408 tuberculosis, brazil, *Emerg. Infect. Dis.* **21**, 452–455 (2015).

409 14. R. Coker, M. McKee, R. Atun, B. Dimitrova, E. Dodonova, S. Kuznetsov, F. Drobniewski,
410 Risk factors for pulmonary tuberculosis in Russia: Case-control study, *Br. Med. J.* **332**, 85–87
411 (2006).

412 15. J. L. Warren, L. Grandjean, D. A. J. Moore, A. Lithgow, J. Coronel, P. Sheen, J. L. Zelner, J.
413 R. Andrews, T. Cohen, Investigating spillover of multidrug-resistant tuberculosis from a prison: a
414 spatial and molecular epidemiological analysis, *BMC Med.* **16**, 122 (2018).

415 16. H. E. Jenkins, V. Plesca, A. Ciobanu, V. Crudu, I. Galusca, V. Soltan, A. Serbulenco, M.
416 Zignol, A. Dadu, M. Dara, T. Cohen, Assessing spatial heterogeneity of multidrug-resistant
417 tuberculosis in a high-burden country, *Eur. Respir. J.* **42**, 1291–1301 (2013).

418 17. M. N. Séraphin, X. Didelot, D. J. Nolan, J. R. May, M. S. R. Khan, E. R. Murray, M. Salemi,
419 J. G. Morris, M. Lauzardo, Genomic investigation of a Mycobacterium tuberculosis outbreak
420 involving prison and community cases in Florida, United States, *Am. J. Trop. Med. Hyg.* **99**, 867–
421 874 (2018).

422 18. T. F. Jones, C. L. Woodley, F. F. Fountain, W. Schaffner, Increased incidence of the outbreak
423 strain of Mycobacterium tuberculosis in the surrounding community after an outbreak in a jail.
424 (Original Article), *South. Med. J.* **96**, 155–158 (2003).

425 19. X. Didelot, C. Fraser, J. Gardy, C. Colijn, H. Malik, Genomic infectious disease epidemiology
426 in partially sampled and ongoing outbreaks, *Mol. Biol. Evol.* **34**, 997–1007 (2017).

427 20. World Health Organization, *Global tuberculosis report 2020* (Geneva, 2020).

428 21. K. E. Holt, P. Mcadam, P. Vuong, K. Thai, N. Thuy, T. Thuong, D. Thi, M. Ha, N. N. Lan, N.
429 H. Lan, N. Thi, Q. Nhu, H. T. Hai, V. Thi, N. Ha, G. Thwaites, D. J. Edwards, A. P. Nath, K.

430 Pham, D. B. Ascher, J. Farrar, C. C. Khor, Y. Y. Teo, M. Inouye, Frequent transmission of the
431 Mycobacterium tuberculosis Beijing lineage and positive selection for the EsxW Beijing variant
432 in Vietnam, *Nat. Genet.* (2018), doi:10.1038/s41588-018-0117-9.
433 22. T. M. Walker, T. A. Kohl, S. V. Omar, J. Hedge, C. Del Ojo Elias, P. Bradley, Z. Iqbal, S.
434 Feuerriegel, K. E. Niehaus, D. J. Wilson, D. A. Clifton, G. Kapatai, C. L. C. Ip, R. Bowden, F. A.
435 Drobniewski, C. Allix-Béguec, C. Gaudin, J. Parkhill, R. Diel, P. Supply, D. W. Crook, E. G.
436 Smith, A. S. Walker, N. Ismail, S. Niemann, T. E. A. Peto, J. Davies, C. Crichton, M. Acharya, L.
437 Madrid-Marquez, D. Eyre, D. Wyllie, T. Golubchik, M. Munang, Whole-genome sequencing for
438 prediction of Mycobacterium tuberculosis drug susceptibility and resistance: A retrospective
439 cohort study, *Lancet Infect. Dis.* **15**, 1193–1202 (2015).
440 23. T. M. Walker, C. L. C. Ip, R. H. Harrell, J. T. Evans, G. Kapatai, M. J. Dedicoat, D. W. Eyre,
441 D. J. Wilson, P. M. Hawkey, D. W. Crook, J. Parkhill, D. Harris, A. S. Walker, R. Bowden, P.
442 Monk, E. G. Smith, T. E. A. Peto, Whole-genome sequencing to delineate Mycobacterium
443 tuberculosis outbreaks: A retrospective observational study, *Lancet Infect. Dis.* **13**, 137–146
444 (2013).
445 24. F. Drobniewski, Y. Balabanova, V. Nikolayevsky, M. Ruddy, S. Kuznetzov, S. Zakharova, A.
446 Melentyev, I. Fedorin, Drug-resistant tuberculosis, clinical virulence, and the dominance of the
447 Beijing strain family in Russia, *J. Am. Med. Assoc.* **293**, 2726–2731 (2005).
448 25. J. W. Tappero, R. Reporter, J. D. Wenger, B. A. Ward, M. W. Reeves, T. S. Missbach, B. D.
449 Plikaytis, L. Mascola, A. Schuchat, Meningococcal Disease in Los Angeles County, California,
450 and among Men in the County Jails, *N. Engl. J. Med.* **335**, 833–841 (1996).
451 26. E. Reinhart, D. L. Chen, Incarceration And Its Disseminations: COVID-19 Pandemic Lessons
452 From Chicago's Cook County Jail, *Health Aff.* **39**, 1412–1418 (2020).
453 27. A. da S. Santos, R. D. de Oliveira, E. F. Lemos, F. Lima, T. Cohen, O. Cords, L. Martinez, C.
454 Gonçalves, A. I. Ko, J. R. Andrews, J. Croda, Yield, Efficiency, and Costs of Mass Screening

455 Algorithms for Tuberculosis in Brazilian Prisons, *Clin. Infect. Dis.* (2020),
456 doi:10.1093/cid/ciaa135.
457 28. R. D. de Oliveira, A. da Silva Santos, C. B. Reis, A. de Cássia Leite, F. P. C. Sacchi, R. C. P.
458 de Araujo, P. C. P. dos Santos, V. C. Rolla, L. Martinez, J. Andrews, J. Croda, Primary
459 prophylaxis to prevent tuberculosis infection in prison inmates: A randomized, double-blind,
460 placebo-controlled trial, *Am. J. Trop. Med. Hyg.* **103**, 1466–1472 (2020).
461 29. E. A. Wang, B. Western, D. M. Berwick, COVID-19, Decarceration, and the Role of
462 Clinicians, Health Systems, and Payers: A Report From the National Academy of Sciences,
463 Engineering, and Medicine., *Jama* **06511** (2020), doi:10.1001/jama.2020.22109.
464 30. R. De Oliveira Andrade, Covid-19: Prisons exposed in Brazil's crisis*BMJ* **370** (2020),
465 doi:10.1136/bmj.m2884.
466 31. F. Menardo, S. Duchêne, D. Brites, S. Gagneux, The molecular clock of mycobacterium
467 tuberculosis, *PLoS Pathog.* **15** (2019), doi:10.1371/journal.ppat.1008067.
468 32. C. J. Meehan, G. A. Goig, T. A. Kohl, L. Verboven, A. Dippenaar, M. Ezewudo, M. R.
469 Farhat, J. L. Guthrie, K. Laukens, P. Miotto, B. Ofori-Anyinam, V. Dreyer, P. Supply, A. Suresh,
470 C. Utpatel, D. van Soolingen, Y. Zhou, P. M. Ashton, D. Brites, A. M. Cabibbe, B. C. de Jong, M.
471 de Vos, F. Menardo, S. Gagneux, Q. Gao, T. H. Heupink, Q. Liu, C. Loiseau, L. Rigouts, T. C.
472 Rodwell, E. Tagliani, T. M. Walker, R. M. Warren, Y. Zhao, M. Zignol, M. Schito, J. Gardy, D.
473 M. Cirillo, S. Niemann, I. Comas, A. Van Rie, Whole genome sequencing of Mycobacterium
474 tuberculosis: current standards and open issues, *Nat. Rev. Microbiol.* **17**, 533–545 (2019).
475 33. K. S. Walter, C. Colijn, T. Cohen, B. Mathema, Q. Liu, J. Bowers, D. M. Engelthaler, A.
476 Narechania, D. Lemmer, J. Croda, J. R. Andrews, Genomic variant-identification methods may
477 alter mycobacterium tuberculosis transmission inferences, *Microb. Genomics* **6**, 1–16 (2020).
478 34. J. Mariner, J. Cavallaro, *Behind Bars in Brazil* (Human Rights Watch, New York, 1998).
479 35. A. M. Kozlov, D. Darriba, T. Flouri, B. Morel, A. Stamatakis, J. Wren, Ed. RAxML-NG: A

480 fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference,
481 *Bioinformatics* (2019), doi:10.1093/bioinformatics/btz305.
482 36. R. Bouckaert, T. G. Vaughan, J. Barido-Sottani, S. Duchêne, M. Fourment, A. Gavryushkina,
483 J. Heled, G. Jones, D. Kühnert, N. De Maio, M. Matschiner, F. K. Mendes, N. F. Müller, H. A.
484 Ogilvie, L. Du Plessis, A. Popinga, A. Rambaut, D. Rasmussen, I. Siveroni, M. A. Suchard, C. H.
485 Wu, D. Xie, C. Zhang, T. Stadler, A. J. Drummond, BEAST 2.5: An advanced software platform
486 for Bayesian evolutionary analysis, *PLoS Comput. Biol.* **15**, e1006650 (2019).
487 37. A. da S. Santos, R. D. de Oliveira, E. F. Lemos, F. Lima, T. Cohen, O. Cords, L. Martinez, C.
488 Gonçalves, A. Ko, J. R. Andrews, J. Croda, Yield, Efficiency and Costs of Mass Screening
489 Algorithms for Tuberculosis in Brazilian Prisons., *Clin. Infect. Dis.* (2020),
490 doi:10.1093/cid/ciaa135.
491 38. D. Robinson, fuzzyjoin: Join Tables Together on Inexact Matching (2020) (available at
492 https://github.com/dgrtwo/fuzzyjoin).
493 39. F. Krueger, Trim Galore (2019) (available at
494 https://github.com/FelixKrueger/TrimGalore/blob/master/Docs/Trim_Galore_User_Guide.md).
495 40. M. Martin, Cutadapt Removes Adapter Sequences From High-Throughput Sequencing Reads,
496 *EMBnet.journal* **17** (2011), doi:https://doi.org/10.14806/ej.17.1.200.
497 41. D. E. Wood, S. L. Salzberg, Kraken: Ultrafast metagenomic sequence classification using
498 exact alignments, *Genome Biol.* **15**, R46 (2014).
499 42. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform,
500 *Bioinformatics* **25**, 1754–1760 (2009).
501 43. D. Brites, C. Loiseau, F. Menardo, S. Borrell, M. B. Boniotti, R. Warren, A. Dippenaar, S. D.
502 C. Parsons, C. Beisel, M. A. Behr, J. A. Fyfe, M. Coscolla, S. Gagneux, A new phylogenetic
503 framework for the animal-adapted mycobacterium tuberculosis complex, *Front. Microbiol.* **9**,
504 2820 (2018).

505 44. C. Anyansi, A. Keo, B. J. Walker, T. J. Straub, A. L. Manson, A. M. Earl, T. Abeel,
506 QuantTB- A method to classify mixed Mycobacterium tuberculosis infections within whole
507 genome sequencing data, *BMC Genomics* **21**, 80 (2020).
508 45. A. J. Page, S. R. Harris, T. Seemann, B. Taylor, A. J. Delaney, J. A. Keane, J. Soares, SNP-
509 sites: rapid efficient extraction of SNPs from multi-FASTA alignments, *Microb. Genomics* **2**, 1–5
510 (2016).
511 46. P. Bradley, N. C. Gordon, T. M. Walker, L. Dunn, S. Heys, B. Huang, S. Earle, L. J.
512 Pankhurst, L. Anson, M. De Cesare, P. Piazza, A. A. Votintseva, T. Golubchik, D. J. Wilson, D.
513 H. Wyllie, R. Diel, S. Niemann, S. Feuerriegel, T. A. Kohl, N. Ismail, S. V Omar, E. G. Smith, D.
514 Buck, G. McVean, A. S. Walker, T. E. A. Peto, D. W. Crook, Z. Iqbal, Rapid antibiotic-resistance
515 predictions from genome sequence data for Staphylococcus aureus and Mycobacterium
516 tuberculosis, *Nat. Commun.* **6** (2015), doi:10.1038/ncomms10063.
517 47. C. Allix-Béguec, I. Arandjelovic, L. Bi, P. Beckert, M. Bonnet, P. Bradley, A. M. Cabibbe, I.
518 Cancino-Muñoz, M. J. Caulfield, A. Chaiprasert, D. M. Cirillo, D. A. Clifton, I. Comas, D. W.
519 Crook, M. R. De Filippo, H. de Neeling, R. Diel, F. A. Drobniewski, K. Faksri, M. R. Farhat, J.
520 Fleming, P. Fowler, T. A. Fowler, Q. Gao, J. Gardy, D. Gascoyne-Binzi, A. L. Gibertoni-Cruz, A.
521 Gil-Brusola, T. Golubchik, X. Gonzalo, L. Grandjean, G. He, J. L. Guthrie, S. Hoosdally, M.
522 Hunt, Z. Iqbal, N. Ismail, J. Johnston, F. M. Khanzada, C. C. Khor, T. A. Kohl, C. Kong, S.
523 Lipworth, Q. Liu, G. Maphalala, E. Martinez, V. Mathys, M. Merker, P. Miotto, N. Mistry, D. A.
524 J. Moore, M. Murray, S. Niemann, R. T. H. Ong, T. E. A. Peto, J. E. Posey, T. Prammananan, A.
525 Pym, C. Rodrigues, M. Rodrigues, T. Rodwell, G. M. Rossolini, E. S. Padilla, M. Schito, X.
526 Shen, J. Shendure, V. Sintchenko, A. Sloutsky, E. G. Smith, M. Snyder, K. Soetaert, A. M.
527 Starks, P. Supply, P. Suriyapol, S. Tahseen, P. Tang, Y. Y. Teo, T. N. T. Thuong, G. Thwaites, E.
528 Tortoli, S. V. Omar, D. van Soolingen, A. S. Walker, T. M. Walker, M. Wilcox, D. J. Wilson, D.
529 Wyllie, Y. Yang, H. Zhang, Y. Zhao, B. Zhu, Prediction of susceptibility to first-line tuberculosis

530 drugs by DNA sequencing, *N. Engl. J. Med.* **379**, 1403–1415 (2018).
531 48. J. E. Phelan, D. M. O'Sullivan, D. Machado, J. Ramos, Y. E. A. Oppong, S. Campino, J.
532 O'Grady, R. McNerney, M. L. Hibberd, M. Viveiros, J. F. Huggett, T. G. Clark, Integrating
533 informatics tools and portable sequencing technology for rapid detection of resistance to anti-
534 tuberculous drugs, *Genome Med.* **11**, 41 (2019).
535 49. E. Paradis, K. Schliep, R. Schwartz, Ed. ape 5.0: an environment for modern phylogenetics
536 and evolutionary analyses in R, *Bioinformatics* **35**, 526–528 (2019).
537 50. A. D. Leaché, B. L. Banbury, J. Felsenstein, A. Nieto-Montes De Oca, A. Stamatakis, New
538 Acquisition Bias Corrections for Inferring SNP Phylogenies, *Syst. Biol* **64**, 1032–1047 (2015).
539 51. A. Rambaut, A. J. Drummond, D. Xie, G. Baele, M. A. Suchard, Posterior summarization in
540 Bayesian phylogenetics using Tracer 1.7, *Syst. Biol.* **67**, 901–904 (2018).
541 52. A. Rieux, C. E. Khatchikian, tipdatingbeast: an r package to assist the implementation of
542 phylogenetic tip-dating tests using beast, *Mol. Ecol. Resour.* **17**, 608–613 (2017).
543 53. Y. I. Xu, I. Cancino-Muñoz, M. I. Torres-Puente, L. M. Villamayor, R. I. Borrá, M. Borrá s-
544 Má ñez, M. Bosque, J. J. Camarena ID, E. Colomer-Roig, J. I. Colomina, I. Escribano, O. I.
545 Esparcia-Rodríguez, A. I. Gil-Brusola, R. Moreno-Muñoz, D. NavarroID, M. Navarro, N. Orta, E.
546 Pé rez, J. Prat, J. I. Carlos Rodríguez, M. Montserrat Ruiz-García, H. Vanaclocha, C. ColijnID, I.
547 I. Comas, I. Cancino-Munoz, M. I. Torres-Puente, L. M. Villamayor, R. Borrás, M. Borrás-
548 Máñez, M. Bosque, J. J. Camarena, E. Colomer-Roig, J. I. Colomina, I. Escribano, O. I. Esparcia-
549 Rodríguez, A. I. Gil-Brusola, C. Gimeno, A. Gimeno-Gascón, B. Gomila-Sard, D. González-
550 Granda, N. Gonzalo-Jiménez, M. R. Guna-Serrano, J. L. López-Hontangas, C. Martín-González,
551 R. Moreno-Munoz, D. Navarro, M. Navarro, N. Orta, E. Pérez, J. Prat, J. C. Rodríguez, M. M.
552 Ruiz-García, H. Vanaclocha, C. Colijn, I. I. Comas, High-resolution mapping of tuberculosis
553 transmission: Whole genome sequencing and phylogenetic modelling of a cohort from Valencia
554 Region, Spain, *PLoS Med.* **16** (2019), doi:10.1371/journal.pmed.1002961.

555 54. D. Ayabina, J. O. Ronning, K. Alfsnes, N. Debech, O. B. Brynildsrud, T. Arnesen, G.
556 Norheim, A.-T. T. Mengshoel, R. Rykkvin, U. R. Dahle, C. Colijn, V. Eldholm, Genome-based
557 transmission modelling separates imported tuberculosis from recent transmission within an
558 immigrant population, *Microb. Genomics* **4**, 1–13 (2018).
559

560 **Acknowledgments**

561

562 We would like to acknowledge the contributions of AGEPEN, LACEN, and the Mato
563 Grosso do Sul State Health Department. All authors had full access to all the data in the
564 study and had final responsibility for the decision to submit

565

566 **Funding:**

567 National Institutes of Health grant R01AI130058 (JRA)

568 Brazil's National Council for Scientific and Technological Development
569 grant 404237/2012-6 (JC)

570

571 **Author contributions:**

572 Conceptualization: JRA, JC, KSW, CC, TC, BM, AIK
573 Methodology: JRA, JC, KSW, CC, TC, BM, AIK
574 Investigation: PCPS, TOG, BOS, ASS, ACL, AMS, FMFM, RDO, EFL, EC, YL
575 Visualization: KSW, JRA
576 Funding acquisition: JRA, JC
577 Project administration: JRA, JC
578 Supervision: JRA, JC
579 Writing – original draft: KSW, JRA

Writing – review & editing: all authors

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: Raw Illumina sequence data are available on the Sequence Read Archive under BioProject PRJNA671770.

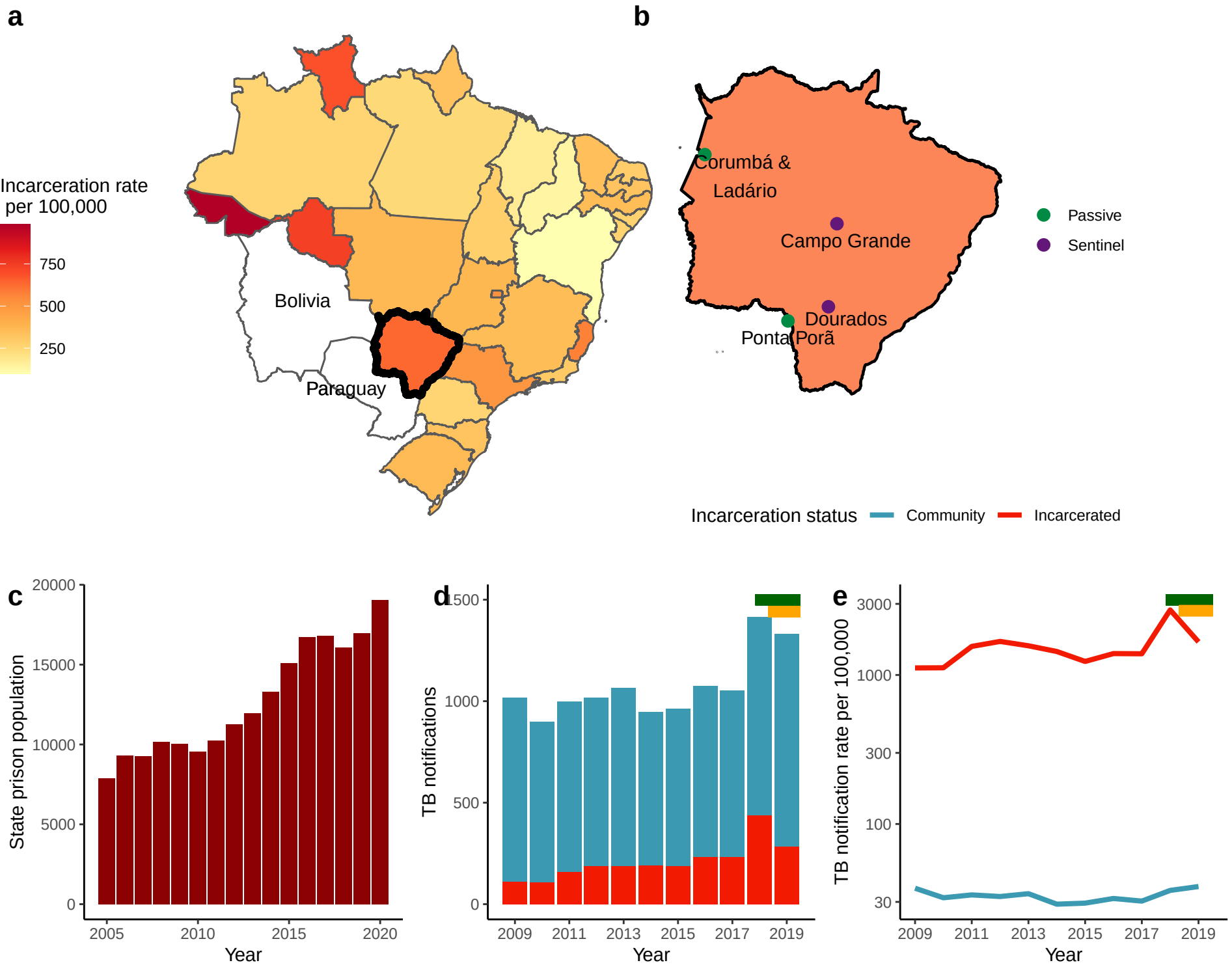
Figures and Tables

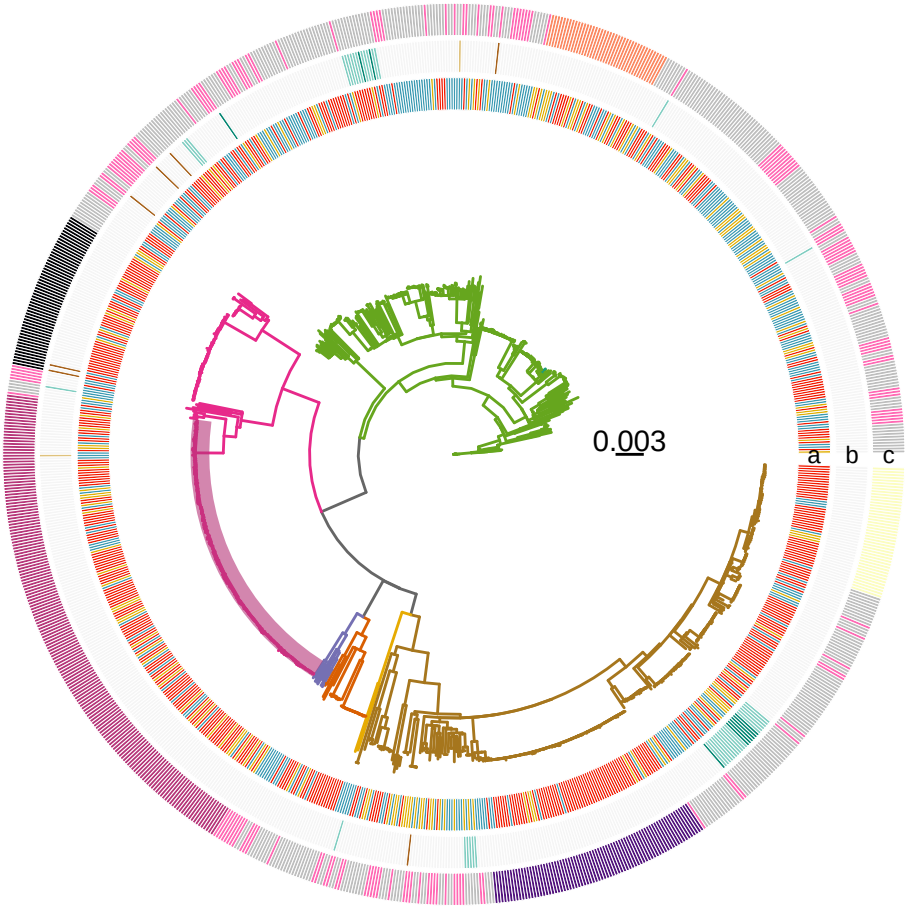
Figure 1. Tuberculosis is increasingly concentrated within prisons in Mato Grosso do Sul, Brazil. (a) Map of Brazil with states colored by the 2019 incarceration rate with Mato Grosso do Sul outlined in black. (b) Map of Mato Grosso do Sul state with the cities with passive surveillance only and where active surveillance was additionally conducted within prisons. (c) The incarcerated population size in Mato Grosso do Sul grew by 142% from 2005 to 2020, increasing from 7,891 to 19,065 people. (d) Mato Grosso do Sul state's annual new and retreatment tuberculosis notifications, colored by incarceration status and (e) the notification rate per 100,000 people for the incarcerated population and non-incarcerated populations from 2009 to 2019. The y-axis is log-scaled. The mean tuberculosis notification rate was more than 46 times greater among the incarcerated population compared to the non-incarcerated population. The green and yellow bars (d) and (e) indicate the period of active surveillance in prisons in Campo Grande and Dourados respectively.

Figure 2. *M. tuberculosis* isolates from incarcerated and non-incarcerated people are closely related in Mato Grosso do Sul, Brazil. A maximum likelihood phylogeny of 932 tuberculosis isolates from Lineage 4 inferred from a multiple sequence alignment of 19,753 SNPs. Branch

605 lengths are in units of substitutions per site. From the inside, rings are colored by incarceration
606 status, antimicrobial resistance prediction, and genomic cluster (Methods). "Other" resistance
607 indicates resistance to at least one antibiotic other than isoniazid or rifampicin. Branches are
608 colored by sublineage. The clade containing the largest predicted genomic cluster (Cluster 5),
609 including 170 isolates, is highlighted in purple.
610
611 **Figure 3. A single, recently emerged *M. tuberculosis* clone spans prisons and the community**
612 **across the major cities Mato Grosso do Sul state.** (a) A Bayesian time-calibrated phylogeny of
613 the largest sampled genomic cluster of 170 isolates. The cluster emerged approximately 23 years
614 before the most recently sampled isolate, with a most recent common ancestor in 1996 (95%
615 HPD: 1989-2003). Tip point color indicates patient's incarceration status at the time of
616 tuberculosis notification. Annotation bar colors indicate city. Grey error bars indicate the 95%
617 Bayesian highest posterior density intervals for node date. Clade posterior support values are
618 shown on the middle of branches for clades with posterior support > 0.5. (b) A haplotype network
619 of the single largest genomic cluster, including 170 isolates. Nodes represent unique haplotypes
620 and are scaled to size. Points along branches indicate SNP distances between haplotypes. Node
621 color indicates incarceration status at the time of diagnosis.
622
623 **Figure 4. Prisons disseminate tuberculosis through the frequent movement of incarcerated**
624 **people between prisons.** (a) Incarceration locates people within highly connected contact
625 networks. Map of Mato Grosso do Sul state with orange points indicating the 21 prisons which
626 most frequently transfer people between prisons. Edges are colored by the mean number of yearly
627 transfers made between prisons from 2015-2018, including transfers in both directions. Edges
628 were drawn between prison pairs with more than 10 transfers in the study period. Transfers
629 between prisons within a city are not shown. (b) An *M. tuberculosis* clone (depicted in Figure 3)

630 spreads throughout multiple prisons, jails, and the community. Each line represents an individual
631 patient infected with an isolate in the largest genomic cluster, ordered by notification date. Line
632 color indicates patient location, determined by matching patient names with the state incarceration
633 database (Methods). Points indicate TB notification dates; point color indicates patient location at
634 the time of TB notification. Individuals with notification dates after December 31, 2018 are not
635 shown.
636
637 **Figure 5. Frequent transmission within prisons and spillover of infection from prisons**
638 **perpetuates the tuberculosis epidemic in Mato Grosso do Sul, Brazil.** We integrated *M.*
639 *tuberculosis* phylogenetic trees with epidemiologic data to infer probable transmission linkages
640 among sampled patients as demonstrated by an example cluster (a-c). (a) Transmission histories
641 can be represented by a phylogenetic trees with branches colored by host in which tips represent
642 sampled isolates and transmission events are indicated by asterisks. (b) We summarized sampled
643 transmission trees into who-acquired-infection-from-whom matrices in which rows represent
644 individual recipients, columns represent individual infectors, and each cell is colored by pairwise
645 transmission probability. (c) The posterior distribution of infection times (grey histogram) for
646 individual d with bars below representing location information for individual d and individual c,
647 the predicted infector. Bars are colored by location over time and circles indicate date of
648 tuberculosis notification. (d) The population-wide WAIFW matrix describing the proportions of
649 observed transmissions to each recipient population from each infector population, corrected for
650 genomic sampling (Methods). Rows represent recipient populations, columns represent infector
651 populations, and cells indicate the proportion of total infections to each recipient population from
652 each infector population defined by their incarceration status at the time of transmission and
653 corrected for sampling rate.
654
655


a
Incarceration rate per 100,000
750
500
250
Bolivia
Paraguay
b
Corumbá & Ladário
Campo Grande
Dourados
Ponta Porã
Passive
Sentinel
Incarceration status
Community
Incarcerated
c
State prison population
20000
15000
10000
5000
0
2005
2010
2015
2020
Year
d
TB notifications
1500
1000
500
0
2009
2011
2013
2015
2017
2019
Year
e
TB notification rate per 100,000
3000
1000
300
100
30
2009
2011
2013
2015
2017
2019
Year



a) Incarceration status

- Formerly incarcerated
- Incarcerated
- No incarceration history

b) Antimicrobial resistance

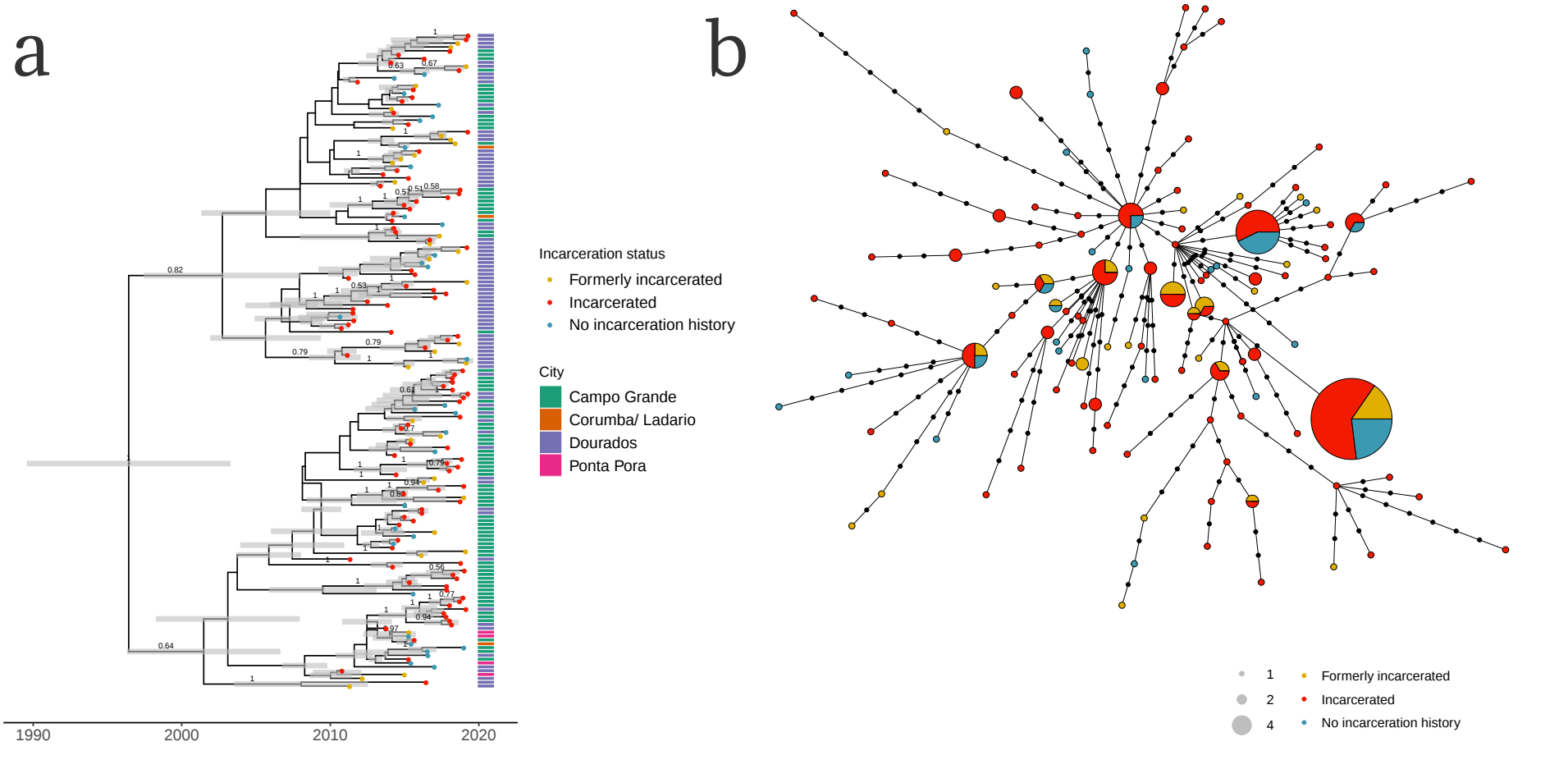
- Isoniazid
- MDR
- Other
- Rifampicin
- Susceptible

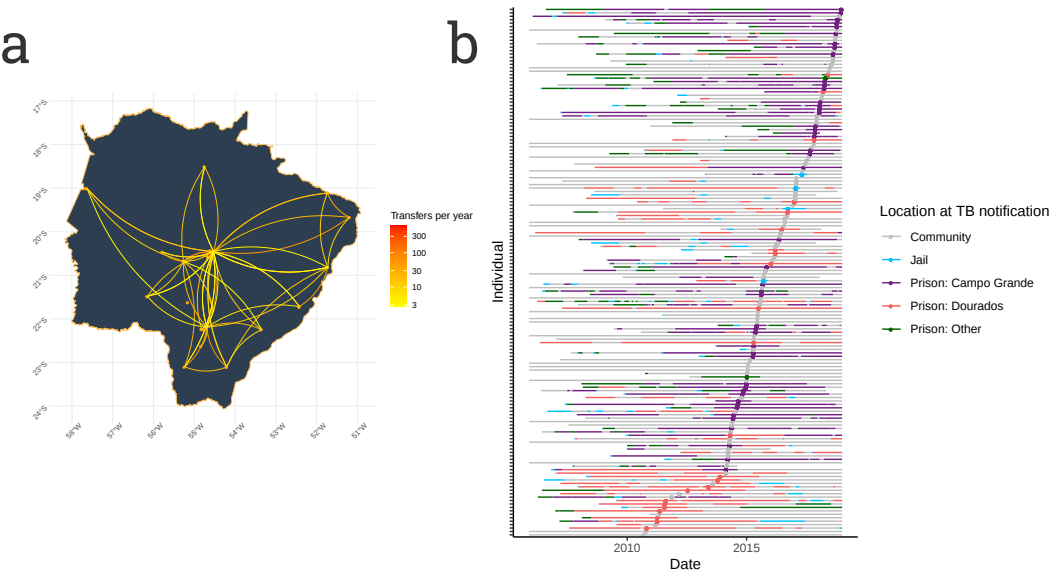
c) Genomic cluster

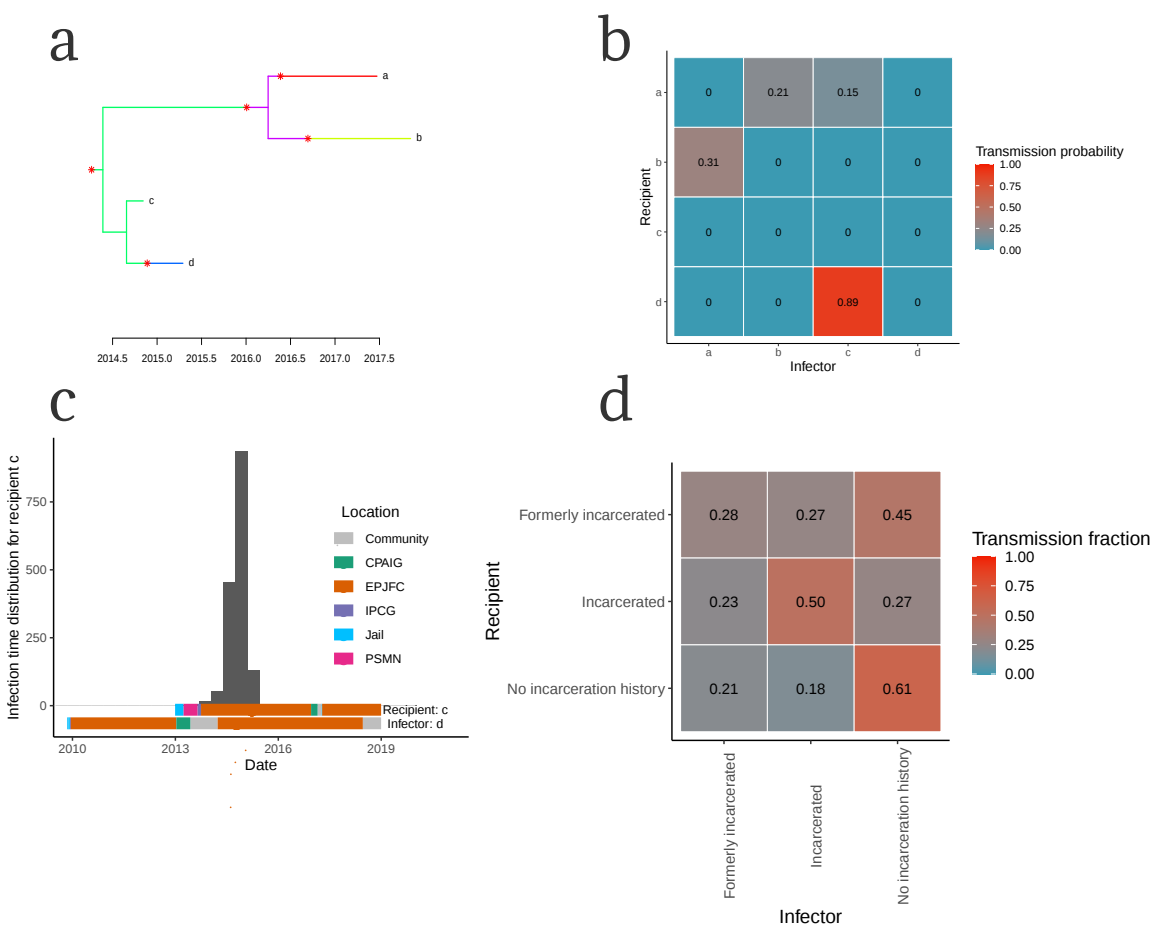
- 1
- 2
- 3
- 4
- 5
- Other
- Unique

Lineage

- 4.1
- 4.2
- 4.3
- 4.4
- 4.7
- 4.8
- 4.9


a
b
1990
2000
2010
2020
Incarceration status
Formerly incarcerated
Incarcerated
No incarceration history
City
Campo Grande
Corumba/ Ladario
Dourados
Ponta Pora
1
2
4
Formerly incarcerated
Incarcerated
No incarceration history


a
17°S
18°S
19°S
20°S
21°S
22°S
23°S
24°S
58°W
57°W
56°W
55°W
54°W
53°W
52°W
51°W
Transfers per year
300
100
30
10
3
b
Individual
2010
2015
Date
Location at TB notification
Community
Jail
Prison: Campo Grande
Prison: Dourados
Prison: Other


a
a
b
c
d
2014.5
2015.0
2015.5
2016.0
2016.5
2017.0
2017.5
b
Recipient
Infector
a
b
c
d
0
0.21
0.15
0
0.31
0
0
0
0
0
0
0
0
0
0.89
0
Transmission probability
1.00
0.75
0.50
0.25
0.00
c
Infection time distribution for recipient c
750
500
250
0
Location
Community
CPAIG
EPJFC
IPCG
Jail
PSMN
Recipient: c
Infector: d
2010
2013
2016
2019
Date
d
Recipient
Formerly incarcerated
Incarcerated
No incarceration history
0.28
0.27
0.45
0.23
0.50
0.27
0.21
0.18
0.61
Transmission fraction
1.00
0.75
0.50
0.25
0.00
Formerly incarcerated
Incarcerated
No incarceration history
Infector