



OPEN

Identification of a predominant genotype of *Mycobacterium tuberculosis* in Brazilian indigenous population

S. A. Hadi^{1,6}, I. V. Kolte^{2,6}, E. P. Brenner^{1,6}, E. A. T. Cunha³, V. Simonsen⁴, L. Ferrazoli⁴, D. A. M. Villela⁵, R. S. Santos², J. Ravi¹, S. Sreevatsan^{1✉} & P. C. Basta^{2✉}

After nearly a century of vaccination and six decades of drug therapy, tuberculosis (TB) kills more people annually than any other infectious disease. Substantial challenges to disease eradication remain among vulnerable and underserved populations. The Guarani-Kaiowá people are an indigenous population in Paraguay and the Brazilian state of *Mato Grosso do Sul*. This community, marginalized in Brazilian society, experiences severe poverty. Like other South American indigenous populations, their TB prevalence is high, but the disease has remained largely unstudied in their communities. Herein, *Mycobacterium tuberculosis* isolates from local clinics were whole genome sequenced, and a population genetic framework was generated. Phylogenetics show *M. tuberculosis* isolates in the Guarani-Kaiowá people cluster away from selected reference strains, suggesting divergence. Most cluster in a single group, further characterized as *M. tuberculosis* sublineage 4.3.3. Closer analysis of SNPs showed numerous variants across the genome, including in drug resistance-associated genes, and with many unique polymorphisms fixed in each group. We report that local *M. tuberculosis* strains have acquired unique polymorphisms in the Guarani-Kaiowá people, and drug resistance characterization is urgently needed to inform public health to ensure proper care and avoid further evolution and spread of drug-resistant TB.

The WHO EndTB strategy focuses on the reduction of incidence and mortality rates towards the elimination of tuberculosis (TB) by 2035, yet the disease remains a substantial public health threat worldwide. Ethnic minorities who face precarious living conditions and lack access to healthcare are at greater risk. TB eradication in aboriginal (indigenous) groups is further complicated by sparse data collection from these communities. The existing evidence indicates that incidence of active TB and prevalence of latent TB infection (LTBI) are significantly higher in indigenous groups compared to non-indigenous populations globally¹. This is a result of socioeconomic inequalities, including limited access to education, lack of employment opportunities and marginalization². To achieve TB eradication in these communities, high-resolution surveillance tools are needed. The use of whole-genome sequencing (WGS) in TB research provides the highest possible resolution to study disease transmission. WGS can diagnose infections directly from primary clinical samples, provide results faster, predict antimicrobial agent resistance-associated genotypes, and has applications in outbreak investigations to better define transmission clusters^{3–8}.

Despite significant advances in TB control in recent decades, Brazil remains a high-burden country with an estimated 95,000 new cases of TB diagnosed in 2018⁹. Indigenous groups in Brazil suffer from incidence rates three times than that of the general population¹⁰. Studies have suggested nearly 50% LTBI prevalence rates in some indigenous territories, along with evidence of recent and ongoing transmission^{11,12}. The Guarani-Kaiowá people in the Brazilian state of *Mato Grosso do Sul* (MS) are Brazil's second largest indigenous group at around 44,000 individuals. For the last decade, MS has borne the country's highest rate of TB among its indigenous population. In 2013, a study of the indigenous people in MS found annual incidence rates of diagnosed active

¹College of Veterinary Medicine, Michigan State University, 784 Wilson Drive, East Lansing, MI 48824, USA. ²Escola Nacional de Saúde Pública, Oswaldo Cruz Foundation, Rua Leopoldo Bulhões, 1480, Rio de Janeiro, RJ 21041-210, Brasil. ³Laboratório Central de Saúde Pública de Mato Grosso do Sul, Campo Grande, Brasil. ⁴Adolfo Lutz Institute, São Paulo, Brasil. ⁵Programa de Computação Científica, Fundação Oswaldo Cruz, Rio de Janeiro, Brasil. ⁶These authors contributed equally: S. A. Hadi, I. V. Kolte, and E. P. Brenner. ✉email: sreevats@msu.edu; pcbasta@ensp.fiocruz.br

TB to be six times the state average¹³. The southern part of the state, Amambai district, stood out with an average annual TB incidence rate of around 400/100,000 in the population of Guarani-Kaiowá ethnicity¹³.

The aim of this study was to characterize the genomic patterns of diversification of drug resistance, and explore whether hyperendemic, locally circulating *Mycobacterium tuberculosis* in the indigenous population of the Guarani-Kaiowá may be a product of regular introductions from outside the community, or unchecked transmission within.

Results

Characterization of the patient population. Selected samples reflected the characteristics of the overall TB patient population in the study area^{14,15} (Supplemental 1—Study Characteristics). Diagnosed patients were 62% male (71% of patients from whom we obtained WGS data). TB patients were relatively young, with a median age of 32.5 in the study area (33 years for our subset used WGS analysis). Nearly 50% of patients sought healthcare within a week of symptom onset, but around 20% (in total and WGS subset cases) reported waiting five weeks or longer to seek healthcare. 9/68 patients had been treated for TB one or more times in the past.

Whole-genome sequencing. After genome processing, the 414 lineage-determining single-nucleotide polymorphisms (SNPs) published in Coll et al. were utilized¹⁶. Four main groups emerged from the analysis (Supplemental 2—SNP Barcoding for Lineage identification). The dominant sublineage (n = 54/66, Supplemental 2, col. D:BE, blue) was 4.3.3. The 4.3.3-determining *Rv1248c* C2526G was always paired with an adjacent G2525T mutation (S842Y). This SNP, among others, may suggest divergence of sublineage of 4.3.3 in Brazil that warrants further exploration. In total, 97% of all *M. tuberculosis* analyzed in this study was identified as sublineage 4.3. When analyzing 9 Brazilian isolates from another study as local references, only 2 out of 7 that were classified as sublineage 4.3 (local references 3 and 5) shared some of the SNPs fixed in our dataset. Separately, 61 SNPs found at greater than 80% prevalence in our isolates were completely absent in all global and local references, including in local references 3 and 5. Remaining isolates belonged to sub-lineages 4.3.4.2 (n = 9/66, Supplemental 2, col. BG:BO, green), 4.4.1.1 (n = 2/66, col. BP:BQ, red), and ambiguous S58 (col. BF, yellow) in either 4.3.2 or 4.3.3, carrying an unusual mixed SNP genotype from multiple lineages (although *Rv1248c* G2525T/C2526G was absent).

A maximum likelihood (ML) phylogenetic tree was generated in MEGA-X to identify genotype clustering patterns. The tree with the highest log likelihood is shown in Fig. 1A. Branch lengths are drawn to scale and represent the number of substitutions. All isolates fell well-separated from global *M. tuberculosis* references [closest = H37Rv, >600SNPs avg. pairwise distance (Supplemental 3—Pairwise distance)]. Bootstrap values were well above recommended cutoffs for statistical support¹⁷. An additional maximum likelihood analysis' tree files, including a tree with bootstraps labeled, are provided in Supplemental 4. The blue clade representing 4.3.3 contains most isolates (n = 54/66, ~82%), though four of these isolates (S01/S09, and S03/S39) stand on relatively distant, robust branches and appear distinct. These isolates fell at least ~180 SNPs away from closest local reference (local reference 5 to S01). The green clade, 4.3.4.2, consisted of 9 isolates, grouped next to four local references (local references 1, 2, 6, and 7), but still showed a minimum SNP distance of ~75 SNPs. The red clade (S63/S29) grouped with local reference 9 at ~245 SNPs distance. Finally, S58 (yellow clade) grouped ~200 SNPs away from closest local reference 4.

Next, a Bayesian phylogenetic consensus tree was generated in MrBayes and is shown in Fig. 1B. The tree closely replicates the reconstruction by the ML tree, and all results from the previous tree apply. Support for the run and Bayesian tree were excellent (Supplemental 5). Finally, an additional ML tree built against H37Rv with only local strains is shown in Fig. 1C, providing an at-a-glance visualization of how barcoding aligned with phylogenetic grouping.

Both the barcoding strategy and phylogenetic trees using different methodologies produced nearly identical findings. Over 80% of isolates grouped into a dominant clade typed as 4.3.3, and some distance was observed even between recent Brazilian strains from another study and the isolates circulating in the studied population.

Single-nucleotide polymorphism analysis and drug resistance-associated genes. Of 68 patients, 9 had been treated for tuberculosis at least once before. One of these isolates failed QC (S02), but of the remaining 8, 6 were infected by 4.3.3, and 2 by 4.3.4.2. To investigate potential drug resistance in this community, an analysis of 13 drug resistance (DR) associated genes was performed, revealing 24 single nucleotide polymorphisms (SNPs). Of these, 79.2% (19/24) led to missense mutations. Among these missense mutations, 57.9% (11/19) were unique to the Guarani-Kaiowá isolates when compared against the 9 Brazilian isolates from Brynildsrud et al.¹⁸. Details of each SNP for every drug resistance-associated gene analyzed for all isolates can be seen in Supplemental 6 (SNP Summary: Sheet 2), which were extracted from the core SNPs (Supplemental 6—SNP Summary: Sheet 1).

Pyrazinamide resistance is associated with *pncA*^{19,20}. All isolates in our dataset appeared genotypically sensitive to pyrazinamide. Additionally, streptomycin resistance-associated genes *rpsL* and *rrs* were found to be wild-type.

Isoniazid resistance is associated with *inhA*, *ahpC*, *kasA*, and *katG*^{21,22}. All four genes were analyzed for variants, including 100 bp upstream of the start codon for *inhA*. No SNPs were seen in *inhA* or *ahpC*. Two SNPs were identified in *kasA* and six in *katG*. These eight SNPs all led to missense mutations (Fig. 2). In *kasA*, G805A was present in all 54 isolates in 4.3.3. This SNP was shared only by local references 3 and 5. In *KatG*, two mutations were found in three isolates at nucleotide C944, one leading to S315T (S03, S39) and the other to S315N (S17). This SNP was also found to be commonly circulating in all local reference strains except in local reference 1.

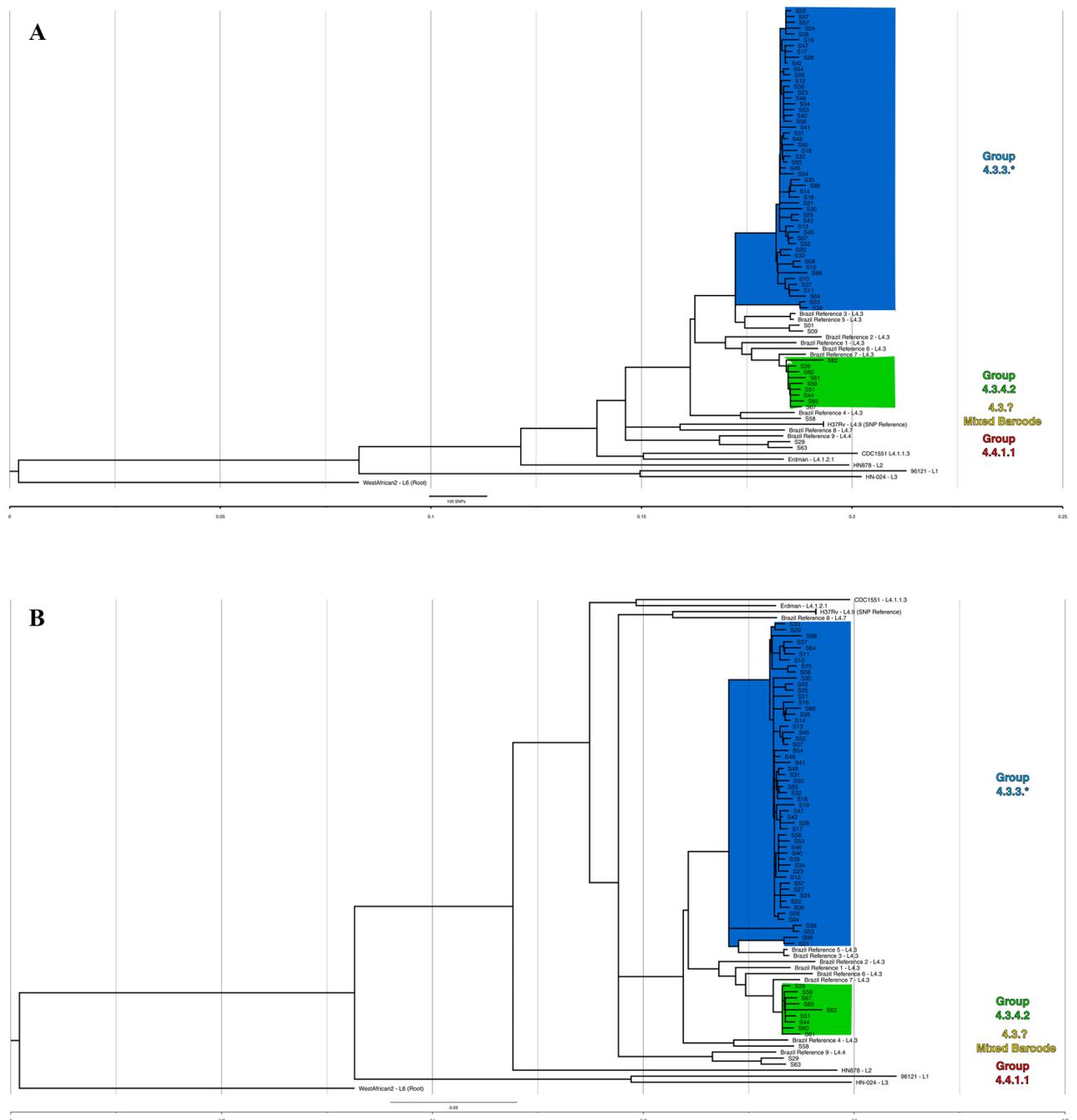


Figure 1. (A) Maximum likelihood (ML) tree showing SNP-based phylogenetic analysis of core SNPs extracted by snippy. Analysis by MEGA-X (default parameters, HKY model) with 500 bootstrap replicates. Tree with highest log likelihood is shown (– 53,357.07). Tree was visualized in FigTree, and labeling generated in Inkscape software. The tree is rooted to *M. tuberculosis* GM041182 (West African 2). The tree is drawn to scale, with branch lengths measured in average substitutions per site (7420 sites total). Bootstrap values are robust across the entire tree and are available in Supplemental 4. Pairwise distances are available in Supplemental 3. Coloring represents groups identified by Coll et al.’s TB barcoding strategy, that groups the isolates based on the presence or absence of lineage-determining SNPs (Supplemental 2). The tree cleanly groups similarly barcoded TB isolates and suggests divergence of sampled isolates even from recent local reference isolates. (B) Bayesian consensus phylogram showing SNP-based phylogenetic analysis of 7420 core SNPs extracted by snippy. Analysis performed by MrBayes (GTR model, nst = mixed, mcmc = 500,000). Tree was visualized in FigTree, and labeling generated in Inkscape software. The tree is rooted to *M. tuberculosis* GM041182 (West African 2). All tree statistics and raw files are available in Supplemental 5. Coloring represents groups identified by Coll et al.’s TB barcoding strategy available in Supplemental B. The tree demonstrates nearly identical grouping of all isolates as the maximum likelihood tree in (A) and lends even more support for barcoding-based grouping. (C) Maximum likelihood (ML) tree showing SNP-based phylogenetic analysis of core SNPs extracted by snippy without global reference lineages. Analysis by MEGA-X (default parameters, HKY model) with 1000 bootstrap replicates. Tree with highest log likelihood is shown (– 30,203.04). Tree was visualized in FigTree, and labeling generated in Inkscape software. The tree is rooted to *M. tuberculosis* H37Rv (Lineage 4.9). The tree is drawn to scale, with branch lengths measured in average substitutions per site (3863 sites total). Bootstrap values are labeled at branch points if greater than 70%, and any points where support falls under 70% is labeled in red and further values are not provided for the clade. Coloring represents groups identified by Coll et al.’s TB barcoding strategy available in Supplemental 2. Without global reference lineages, barcoded isolates still group reliably. S09/S01 and S03/S39 (blue clade, bottom four isolates) stand out by branch lengths as divergent from rest of blue clade.

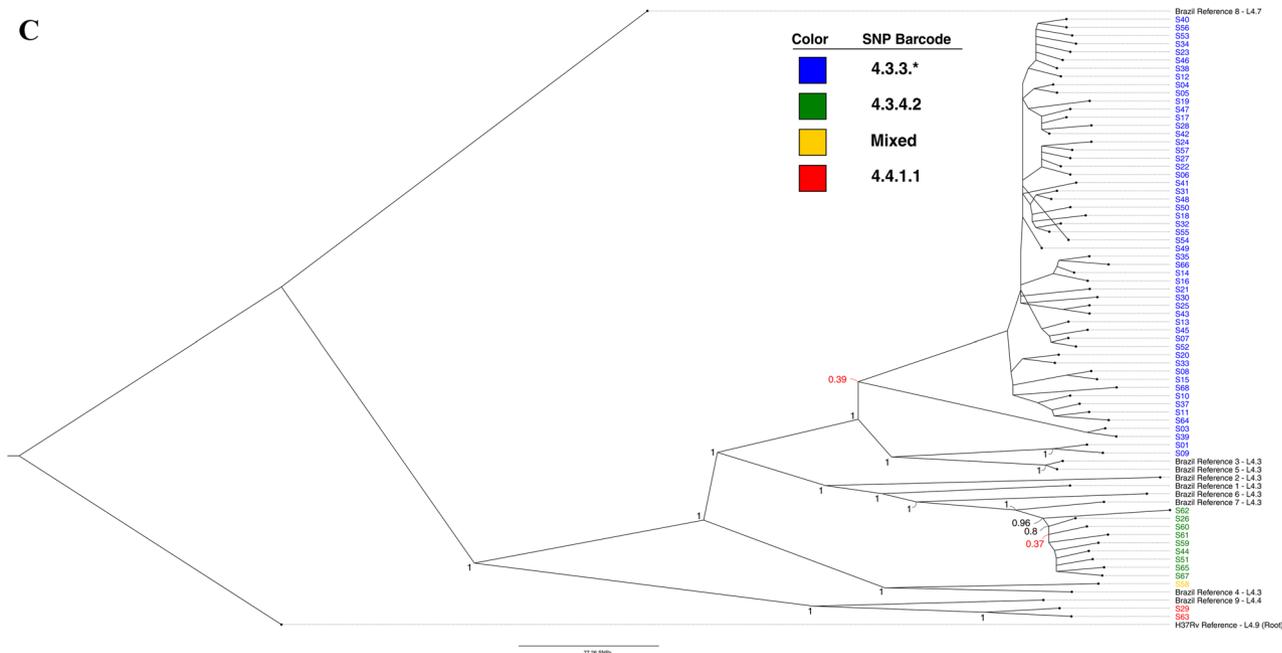


Figure 1. (continued)

Only two samples in our dataset were known to be phenotypically resistant to isoniazid—S03 and S17—and both had SNPs in *kasA* and *katG*.

The 80 bp region in *rpoB*—the rifampicin resistance-determining region (RRDR)—did not carry any mutations. However, two missense mutations were seen outside the RRDR in three samples—S03, S68 and S12 (Fig. 3). S03 was the only sample that was phenotypically rifampin resistant, confirming S03 as multidrug resistant both phenotypically and genotypically due to presence of mutations in *kasA*, *katG*, and *rpoB*, although the *rpoB* SNP fell outside the RRDR. No mutations were observed in *rpoA* in any of the isolates, while S68 showed a missense mutation C3119G (P1040R) in *rpoC*. It is of note that S50 was phenotypically resistant to rifampin, but no SNPs were identified in *rpoB*.

Ethambutol resistance-associated mutations occur in three genes identified as *embABC*. No SNPs were detected in *embA*. One silent mutation in *embB*, and 3 silent plus 2 missense mutations in *embC* were identified (Fig. 4). *embC* missense mutations were seen in S01, S09 and S08, but drug susceptibility data was not available for any of these isolates. The silent mutation at C2781T in *embC* was found in all isolates and references except H37Rv.

For the second line anti-tuberculosis drug fluoroquinolone, two genes—*gyrA* and *gyrB*—were analyzed. All SNPs identified were missense—5 in *gyrA* and 2 in *gyrB*. Out of the 5 SNPs in *gyrA*, G739A (G247S) was fixed in all 54 isolates that fell in sublineage 4.3.3. Three SNPs were common to all 66 isolates as well as our global references except for H37Rv: *gyrA* E21Q, S95T, and G668D. These 3 SNPs were present in almost all Brazilian reference strains, with the exception of S95T & G668D that were absent in local reference 8. The 2 missense SNPs in *gyrB* were found only in two isolates and were absent in Brazilian references (Fig. 5).

Discussion

Although TB is hyperendemic among indigenous populations like the Guarani-Kaiowá in Brazil, there has been a lack of investigation into TB transmission in these settings, a red flag for public health policy. This is, to our knowledge, the first study applying WGS to TB samples from an indigenous population in a high-burden, low-and-middle-income country. Our phylogenomic study indicates that, although cases were diagnosed across six different territories over a study period of six years, a majority of the sequenced cases are caused by *M. tuberculosis* clustered into sublineage 4.3.3 (Fig. 1, blue clades). The three barcoded sublineages observed showed divergence from global references (~600 + SNP average distance), necessitating the inclusion of other recently published Brazilian isolates for comparison (Supplemental 7—Isolates & References: Sheet 3). These nine additional isolates (referred to as “local references”) and an analysis methodology were graciously shared by Drs. Brynildsrud and Eldholm, and this allowed a more refined comparison between Brazilian TB strains¹⁸. Isolates remain divergent by phylogeny and SNP analysis even when including these local references from 2009 to 2012. While local reference strains reliably cluster with our isolates, the average SNP distance within 4.3.3 isolates was ~72 SNPs (SD ± 45), and the distance from two closest 4.3 local references was ~210 SNPs (SD ± 10). The strains observed in this study are Latin American-Mediterranean (LAM) type TB, which is itself not rare in the region. However, the unusually high prevalence of *M. tuberculosis* sublineage 4.3 (64/66, ~97%) is noteworthy. It is far greater than any LAM incidence observed in any prior report the authors could identify in South America or globally, with frequencies typically up to 50%^{23–26} and peaking in South America at only ~65% in one study²⁴. We speculate that the robust grouping of isolates distant from global references, the divergence from local

references, and the prevalence of a single type of TB (LAM) might represent that disease in this community is not the result of modern, recurring introductions. Rather, mutations absent in global lineages yet fixed in this study potentially suggest earlier, historical introductions of *M. tuberculosis* and a subsequent circulation and diversification of local sublineage 4.3 (LAM) within the Guarani-Kaiowá population. This hypothesis would be expected for an isolated community, but the Guarani-Kaiowá cannot live off their overcrowded territories and are therefore not living in isolation. Our findings then would rather support the extreme level of marginalization suffered by this indigenous group—to the point where the fingerprints of interaction appear largely undetected by high-resolution genomic data.

In both phylogenetic trees, we observed universal agreement for all barcoded groups. Isolates grouped into the blue clades (Fig. 1A,B) are dominant and responsible for most sampled infections. Both trees and SNP-based classifications support additional uncommon subtypes circulating in this population, like the distant sublineage 4.4.1.1 isolates (red clade) of S29/S63, or the unusual genotype observed in S58 (yellow clade). We find 111 SNPs that are absent in global references but common at 80% + prevalence in local references and isolates, an additional 63 SNPs present only in 4.3.3 and the closest related local references 3 and 5, and a further 61 SNPs found at 80% + prevalence solely in our 4.3.3 isolates alone (Supplemental 6—SNP Summary: Sheet 1). Isolates show variability in drug resistance-associated genes, a possible portent of developing resistance. Two variants—*kasA* G805A and *gyrA* G739A—have become dominant and fixed in most sampled infections among the Guarani-Kaiowá population. These fixed variants were specifically identified in isolates of sublineage 4.3.3. Additionally, 4 SNPs were identified in all 66 isolates in the fluoroquinolone (*gyrA*) and ethambutol (*embC*) resistance-associated genes. Without careful genome assembly and associated error-checking of the local reference paired-end reads we used, as well as still more local references to compare against, we cannot say for certain whether all these SNPs are truly unique, nor can we speculate what effects they may have, but they do continue to support diversification within the studied population.

Taken together, this is alarming for the Guarani-Kaiowá people from a public health policy perspective, as patients are currently treated as drug-sensitive unless treatment fails, due to insufficient provision of susceptibility kits to the laboratory environment systems (LACENs). In addition to needless suffering, this may have allowed drug-resistant *M. tuberculosis* to arise unchecked in Guarani-Kaiowá communities. Drug susceptibility testing should be provided by the Brazilian Ministry of Health as part of the national TB control program, as it is urgently needed in these vulnerable populations; both to lessen the existing high disease burden as well as to prevent the emergence of MDR-TB that can cause further harm and confound control efforts.

A recent study of *M. tuberculosis* isolates from the Inuit population in Greenland detected several micro-epidemics that could be confirmed with epidemiological data and noted remarkable geographical confinement of lineages within individual villages²⁷. Similar observations have been made among Inuit populations in Canada²⁸. The WGS studies among the Inuit have described a pattern of a limited number of endemic lineages circulating since their introduction by white colonizers in the early twentieth century. The occasional outbreaks or micro-epidemics among the Inuit were not driven by an adaptation of *M. tuberculosis* but rather a change in circumstances in the communities²⁸ or down-scaling of TB programs following a period of low incidence²⁷. A study from New Zealand of *M. tuberculosis* isolates from the Maori population identified L4.4.1.1 sublineage clade of European origin (a sublineage also detected in the current study, albeit at low incidence). Molecular dating showed that this clade was introduced at the time of arrival of European traders. The clade is also found in *M. tuberculosis* isolates from indigenous populations in Canada²⁹. While WGS studies of *M. tuberculosis* complex have not yet been carried out among indigenous populations in other regions, high levels of Mycobacterial interspersed repeat units—variable number tandem repeat genotype similarity have also been reported from the Warao in Venezuela³⁰. A restriction fragment length polymorphism (RFLP) study from MS using data from 1991 to 2001 showed high genotype similarity¹², while a later RFLP study from the same area using data from 2009 to 2015 revealed a pattern of more fragmented clustering³¹. This was attributed to a more aggressive policy of LTBI treatment of contacts, thus limiting the reactivation of previously dominant endemic strains³¹.

Our study analyzed genotypic drug resistance patterns in *M. tuberculosis* strains circulating in this small indigenous community. A high prevalence of multiple frontline anti-TB drug-specific mutations identified in this study suggests an urgent need for a change in public policy tailored to this population.

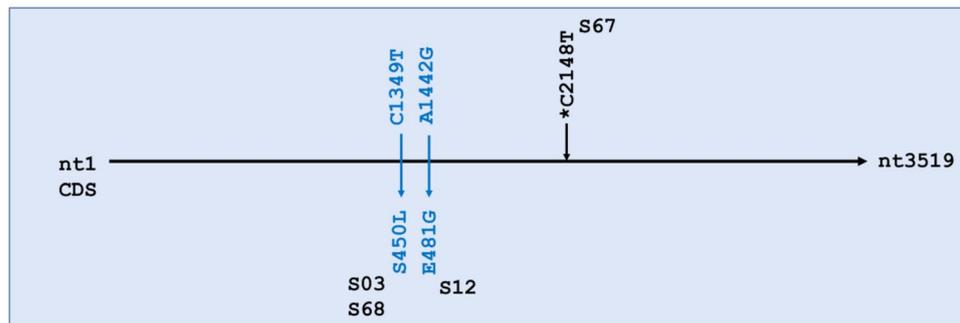
Some isoniazid resistance-associated genes, such as *inhA* or *ahpC*, showed no changes, while *katG* and *kasA* had multiple mutations. KatG normally activates the isoniazid prodrug (INH), while KasA contributes to mycolic acid synthesis that isoniazid impairs. Changes to *katG* Ser315 are strongly associated with isoniazid resistance³². Globally, a single change in *katG* Ser315 is responsible for 64.2% of all phenotypic resistance to isoniazid, the majority of which (95.3%) cause S315T and a few (3.6%) are associated with S315N²¹. These substitutions were also observed in a study in China, where single *katG* mutations occurred in majority of the cases, most of which had S315T and two bore S315N³³. In our analysis, S315T was observed in two isolates S03 and S39, while S315N was identified in S17. Though we lack phenotypic data on S39, both S03 and S17 were phenotypically resistant to isoniazid. Jagielski et al. reported on SNPs in *kasA* and found that G269S was present in 22% of MDR clinical cases, and in 10% of isoniazid mono-resistant³⁴. Older studies have also reported this^{32,35–37}. In our work, 81.8% of isolates (54/66) showed G269S (G805A), suggesting a much higher prevalence than found elsewhere. Unfortunately, highly limited phenotypic data (n = 2/54) stymies further interrogation beyond S03 and S17, which were confirmed isoniazid resistant. These findings merit in-depth study of the population, as literature suggests some of the observed and commonplace SNPs are associated with drug resistance and prompting concern about drug-resistant tuberculosis becoming ubiquitous in the Guarani-Kaiowá people.

In *rpoB*, drug resistance-associated mutations are mostly confined to the RRDR (codon 507 to 533), coding for the core of the beta-subunit of RNA polymerase. Almost 95% of all resistance-associated mutations exist in this hotspot. In our study, all 3 SNPs were found to be outside the RRDR. Of 66 isolates, S03 was the only one known to be multidrug-resistant, yet both S68 and S03 had the same amino acid substitution S450L that is commonly



Figure 2. Isoniazid resistance-associated genes *katG* and *kasA* had multiple single nucleotide polymorphisms, whereas *ahpC* and *inhA* were wildtype. Missense mutations are colored blue.

Rifampin - *rpoB*



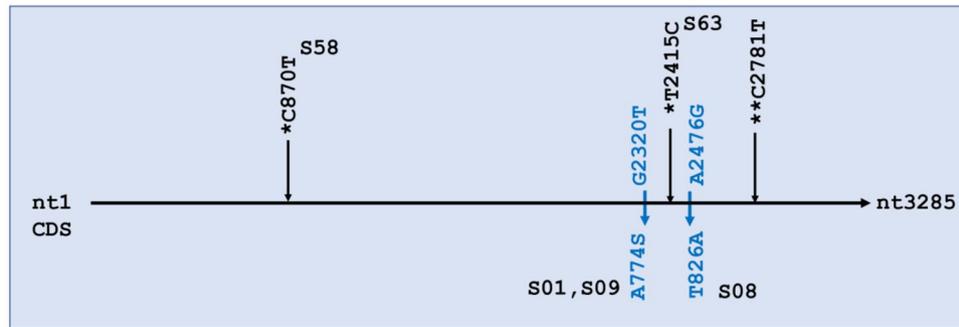
*Black arrow represents silent mutation

Figure 3. Rifampin resistance-associated gene *rpoB* had only three single nucleotide polymorphisms (SNPs). Missense mutations are colored blue. S67 had a silent mutation represented by the *black arrow.

associated with resistance to rifampicin^{38,39} This mutation falls inside the polymerase's highly conserved catalytic pocket (codon 443–451)⁴⁰, suggesting that S68 might also have been multi drug-resistant like S03. Furthermore, S450L has been reported to accumulate significantly more compensatory mutations than other rifampicin resistance-associated SNPs⁴¹. This was observed in our study uniquely in S68, with a missense mutation in *rpoC* (P1040R) lending further support to possible rifampin-resistance in S68. We lack the phenotypic data to confirm our hypothesis, again calling for intensive drug testing in the community before initiation of treatment^{42–44}.

Fluoroquinolone, the second-line TB drug, targets a DNA gyrase encoded by *gyrA* and *gyrB*. Like rifampin's *rpoB* hotspot, drug resistance to fluoroquinolone is also associated with specific mutations in the quinolone resistance-determining region (QRDR) of *gyrA* & *gyrB*⁴⁵. In *gyrA*, T378S was a substitution found to be unique to our study (Fig. 5). Another substitution, G247S, was found 100% of all our 4.3.3 isolates and was not seen in any other isolates or global references. Only the two Brazilian reference strains (reference 3 and 5) that grouped on the periphery of our 4.3.3 isolates showed this SNP, suggesting these references may be tightly related to our predominant sublineage. This mutation falls outside the QRDR and has not been associated with fluoroquinolone resistance⁴⁶. Whether it led to drug resistance in patients is unknown, but it is important to note that while indigenous patients are not treated with fluoroquinolones for TB in Brazil, drugs in this major group are regularly administered for other ailments such as pneumonia, foodborne diseases, and urinary tract infections caused by a wide variety of bacteria such as methicillin-resistant *Staphylococcus aureus*⁴⁷, *Enterobacteriaceae*⁴⁸, *Pseudomonas aeruginosa*⁴⁹, and *Salmonella*⁵⁰. It is possible that the use of fluoroquinolones for other conditions in parallel with endemic TB has led to the development of unique circulating TB strains. Fixing of SNPs in the

Ethambutol - *embC*

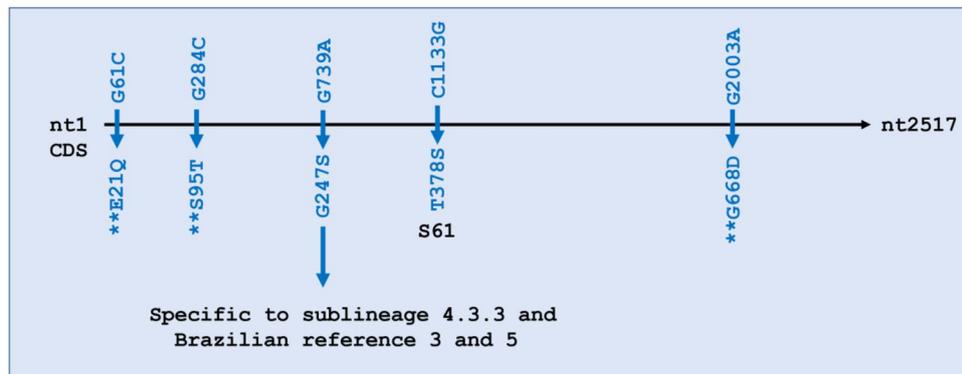


*Black arrow represents silent mutation

**Found in all isolates and global reference lineages except MTB H37Rv

Figure 4. Ethambutol resistance-associated gene *embC* had 2 missense single nucleotide polymorphisms (SNPs) and one silent SNP (**C2781T) found in all 66 isolates in the data set as well all reference strains except H37Rv and is not considered particularly informative. All samples were wildtype for *embA*. Missense mutations are colored blue. Silent mutations are represented in *black color.

Fluoroquinolone - *gyrA*



**Found in all isolates and global reference lineages except MTB H37Rv

Fluoroquinolone - *gyrB*

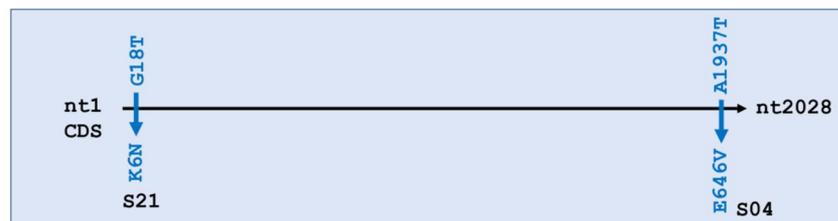


Figure 5. Fluoroquinolone resistance-associated genes *gyrA* and *gyrB*, had a few yet significant single nucleotide polymorphisms (SNPs). All SNPs were missense (blue). Three SNPs (**) in *gyrA* were fixed in the entire data set and all references except H37Rv, whereas one SNP in *gyrA* was found in 54/66 isolates. The former are not considered particularly informative, while the latter does appear to be. Two SNPs were observed in *gyrB*.

entire data set and the establishment of unique SNPs beyond what are observed in existing sub-lineages support diversification in the indigenous population.

Multiple mutations were identified in nearly all drug resistance-associated genes that were analyzed, with many in isoniazid, ethambutol, and fluoroquinolone drug resistance-associated genes. A 2000–2006 study analyzing 783 TB isolates (82.4% new cases and 17.6% previously treated patients) from the general population in *Mato Grosso do Sul* found evidence of drug resistance in 18.3% (all combinations) of cases. Looking at single-drug primary resistance, streptomycin showed the highest level of resistance (3.4%), followed by isoniazid (2.9%),

ethambutol (1.7%), and rifampin where no resistance was detected. The resistance to streptomycin was believed to be due to endogenous reactivation of TB strains from before the 1980s, as streptomycin was not given to new cases. Acquired resistance was found to isoniazid (7.2%), streptomycin (3.6%), rifampin (2.2%), and ethambutol (1.4%). Although levels of primary resistance were lower than observed in contemporary data from other regions, the levels of acquired resistance were already high in 2000–2006⁵¹. As the relative risk of TB is six times higher for the indigenous versus the non-indigenous population in MS¹⁰ we could speculate that a significant proportion of the samples from the 2000–2006 study could have been from indigenous patients.

Though our study lacked lab-confirmed drug resistance data for all samples, ample literature exists that support that the genotypic mutations seen in our work has the potential for resistance to the standard treatment protocol. We recommend that regular surveillance and drug-susceptibility information be made compulsory before the initiation of treatment.

In summary, these data provide the first detailed genomic view of hyperendemic TB circulating within the Guarani-Kaiowá people. The vast majority of *M. tuberculosis* isolates fell into a dominant clade of the Latin American sub-lineage 4.3.3 by Coll et al.'s SNP barcoding strategy¹⁶. This grouping was fully supported by two phylogenetic methods using a core SNP set of 7420 total positions taken from our 66 isolates, 7 global references, and 9 local references. Analysis of SNPs both for drug resistance-associated genes—specifically *gyrA* (fluoroquinolone) and *kasA* (isoniazid)—and for lineage-determination revealed diversification, with unique SNPs fixed in the population. Among 4.3.3-classified isolates, there was the universal presence of a two nucleotide *Rv1248c* G2525T/C2526G change. Unlike 4.3.3 sub-lineage determining *Rv1248c* C2526G, these two changes lead to an amino acid substitution (S842Y). When looking at references, this substitution was only observed in 2/7 of Brazilian 4.3 references. This and the presence of other unique SNPs fixed in the population (*gyrA* and *kasA*) leads us to speculate that most TB in the Guarani-Kaiowá people is the result of a historical introduction of the commonly circulating 4.3.3 lineage, with subsequent diversification in the community. The incidence of sublineage 4.3 is far higher than observed anywhere else in South America, suggesting it is hyperendemic in the population. Molecular and phenotypic characterization of this microevolution is important to discern whether drug resistance has become fixed in this community. The widespread SNPs observed in this study suggest that a more targeted public health strategy should be implemented in the region, where TB cases are not screened for drug resistance unless all prior treatments have failed.

Methods

Study area, population, and design. The population of the Brazilian state of MS is around 2.7 million, including approximately 74,000 indigenous people from nine different groups, of which the Guarani-Kaiowá and the Terena ethnicities represent the majority. Despite international condemnation of human rights violations against the Guarani-Kaiowá⁵², the Guarani-Kaiowá remain confined to overcrowded reservations⁵³. Most income for the Guarani-Kaiowá in southern MS comes from seasonal work on plantations, where overcrowded workers' barracks have been identified as TB transmission hotspots⁵⁴. Poor living conditions of the Guarani-Kaiowá, with high rates of violence⁵³, low education and income levels⁵⁵, and food insecurity⁵⁶ favor transmission of TB.

Control of TB in the indigenous populations of Brazil is the responsibility of the Special Secretariat for Indigenous Health of the Ministry of Health (Portuguese: SESAI). SESAI is composed of 34 Special Indigenous Health Districts which organize healthcare centers located in areas with significant indigenous populations. Each center manages smaller healthcare stations inside each indigenous territory as primary points of care. We examined TB cases from the *Amambai* and *Caarapó* healthcare centers during the study period of 2011 until 2016 (Supplemental 1—Study characteristics).

TB diagnosis in the study area was carried out at local healthcare centers. Ziehl–Neelsen bacilloscopy was performed until 2014 when GeneXpert took its place. Ministry of Health's TB protocol⁵⁷, positive bacilloscopy/ GeneXpert results should be confirmed by the culture at the General Laboratory of MS; this confirmation is not always possible due to logistical challenges in getting samples to the laboratory⁵⁸ on top of well-documented difficulties with successfully culturing *M. tuberculosis*⁵⁹.

Per the Ministry of Health⁵⁷, the standard treatment regimen for TB during the study period was two months of rifampin, isoniazid, pyrazinamide, and ethambutol, followed by four months rifampin and isoniazid. Streptomycin and fluoroquinolone are second-line TB drugs and are used only in special regimes of treatment, with the patient under close medical supervision. Testing cultures for susceptibility to first-line drugs was done at the Central Public Health Laboratories of *Mato Grosso do Sul* (*Laboratório Central de Saúde Pública de Mato Grosso do Sul*, Portuguese acronym LACEN-MS) only with clinical indication.

The indigenous territories included in this study (*Amambai*, *Guassuty*, *Jaguari*, *Kurussu Amba*, *Lima Verde*, *Taquaperi*, and *Caarapó*) are assisted by *Amambai* and *Caarapó* health units and are located on the border of Paraguay. Together, these Guarani-Kaiowá communities encompass approximately 18,000 people (Supplemental 1—Study characteristics).

Tuberculosis isolate identification, regrowth and DNA extraction. Patients diagnosed with TB during the study period were identified in the records of the SESAI. *M. tuberculosis* isolates from these patients were identified in the MS Laboratory Environment Management System. All cultures still in storage at LACEN-MS were regrown in the Ogawa-Kudoh medium. DNA from the regrown cultures was extracted and purified at São Paulo's Instituto Adolfo Lutz using the cetyltrimethylammonium bromide (CTAB)-lysozyme method as described by van Embden et al.⁶⁰. Purified DNA concentrations were determined by Gene-Quant II (Pharmacia Biotech).

SESAI registered 277 incident cases during the study period. We set a cutoff for sample metadata completeness, which limited our set. Furthermore, not all samples arrived with sufficient quality for sequencing in Michigan, and WGS was ultimately performed on 68/277 samples.

Whole-genome sequencing and barcoding. Sixty-eight samples were submitted for WGS on the MiSeq (Illumina) platform at Michigan State University. Samples were numbered S1–S68. Samples S02 and S36 failed quality checks and were excluded. The remaining 66 isolates were trimmed by Trimmomatic⁶¹ (trimmomatic PE, trimming TruSeq2 adapters, with parameters 2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) and reference-based assembly was performed using *M. tuberculosis* H37Rv in CLC Genomics Workbench v12.0. Representative global reference lineages included were *M. tuberculosis* 96121 (Manila-Lineage 1), HN878 (Lineage 2), HN-024 (Lineage 3), CDC1551 (Lineage 4), H37Rv (Lineage 4), Erdman (Lineage 4), and GM041182 (West African 2-Lineage 6). Multiple representatives of lineage 4 were included due to its historical and modern clinical importance. A whole genome multiple sequence alignment was constructed with default parameters in CLC (Supplemental 8—Alignment). To identify the sublineages of *M. tuberculosis* circulating in the community, we utilized specific SNPs for “barcoding” of TB, as reported by Coll et al.¹⁶. The whole-genome alignment of Brazilian isolates was searched manually for each of the 414 reported SNP barcodes. If all diagnostic SNPs for a particular sub-lineage were found in an isolate, it was considered part of that group. A summary is shown in Supplemental 2—SNP Barcoding for lineage identification.

Core SNP extraction, phylogeny generation, and analysis. After initial analysis, we sought local references to better understand observed diversity. Unfortunately, assembled Brazilian whole genome sequences are vanishingly rare, and we reached out to Drs. Brynildsrud and Eldholm, who in 2018 were first and corresponding authors, respectively, on an expansive investigation into *Mycobacterium tuberculosis* lineage 4 isolates. Their dataset includes hundreds of paired-end FASTQ read sets for South American tuberculosis genomes. Reads for 9 Brazilian isolates unrelated to our current study were downloaded from NCBI (BioProject PRJEB27366) and processed utilizing the snippy pipeline⁶². Briefly, we included CLC-constructed contigs for all 66 of our isolates that passed QC, assembled whole genomes for global references, and raw paired-end FASTQ reads for 9 isolates (referred to in this manuscript as “local references”) from the Brynildsrud et al., dataset (Supplemental 7—Isolates & References). These local references represent sublineages 4.3 (LAM, n=7), 4.4 (S-type, n=1), and 4.7 (Congo, n=1)⁶³. They were first processed by ABySS⁶⁴ (abyss-pe, default parameters, k=96) to construct contigs, and they, along with all contigs from our isolates and assembled genomes for global references, were batch-processed in snippy with the *M. tuberculosis* H37Rv Genbank (.gb) file as reference. The snippy output folders for each isolate and local references were processed by snippy-core using H37Rv.gb as reference. This output core alignment, comprised solely of variant sites and excluding complex changes like indels, was analyzed in 16 drug resistance-associated genes. To avoid the effects of uninformative, false positive hits in repetitive regions of the genome, snippy-core was run again with the mask parameter and the *M. tuberculosis* H37Rv-based .bed mask file included in the snippy package to filter out loci like PE/PPE family proteins. The unmasked and masked core SNP sets were compared, and ~3% of SNPs in both isolates and local references were considered uninformative by this approach. Masked core SNPs were passed into MEGA-X v10.1.8⁶⁵ and MrBayes v3.2.7a⁶⁶ separately. Maximum-likelihood trees were produced with 500 bootstrap replicates and the HKY model with other parameters default, and the tree with highest log-likelihood is shown (Fig. 1A). In MrBayes, the GTR model was used, with the parameter nst=mixed used to sample across the GTR model space. After a MCMC run length of 500,000, the standard deviation of split frequencies fell well below 0.01 as recommended, minimum ESS values were 300 or higher, above the 100 recommended, and PSRF values were ~1.000 (Supplemental 5—Bayesian Analysis). A quick comparison of ML phylogeny (HKY, default parameters, bs=100) between masked and unmasked core SNPs showed greater bootstrap values for nodes derived from the masked set, suggesting removing uninformative hits yielded a more robust phylogeny (Supplemental 9—Masked vs Unmasked Phylogeny). As such, the masked SNP set was used for final analysis. Trees were visualized in FigTree and MEGA-X. Final labeling of trees (Fig. 1) was performed with Inkscape v.1.0⁶⁷, but unmodified Newick tree files are also included for transparency (Supplemental 4 and 5).

Ethical aspects. This study was performed in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the National School of Public Health (Report 354.060/2013) and the National Committee for Ethics in Research (Report 650.820/2014), which are both a part of the Brazilian Ministry of Health. Written informed consent was obtained from all study subjects.

Data availability

Raw FASTQ sequences for all 68 sequenced isolates (including the two that did not pass our QC) are freely available through NCBI (BioProjectID: PRJNA669332). This work would not have been possible without both global reference and Brazilian sequences also made available through NCBI. Accession IDs are available for all references in Supplemental 7 (Isolates & References).

Received: 19 August 2020; Accepted: 7 December 2020

Published online: 13 January 2021

References

1. Tollefson, D. et al. Burden of tuberculosis in indigenous peoples globally: A systematic review. *Int. J. Tuberc. Lung Dis.* <https://doi.org/10.5588/ijtld.12.0385> (2013).

2. Cormier, M. *et al.* Proximate determinants of tuberculosis in Indigenous peoples worldwide: A systematic review. *Lancet Glob. Heal.* [https://doi.org/10.1016/S2214-109X\(18\)30435-2](https://doi.org/10.1016/S2214-109X(18)30435-2) (2019).
3. Crisan, A., McKee, G., Munzner, T. & Gardy, J. L. Evidence-based design and evaluation of a whole genome sequencing clinical report for the reference microbiology laboratory. *PeerJ* <https://doi.org/10.7717/peerj.4218> (2018).
4. Lee, R. S. & Behr, M. A. The implications of whole-genome sequencing in the control of tuberculosis. *Ther. Adv. Infect. Dis.* <https://doi.org/10.1177/2049936115624630> (2016).
5. Walker, T. M. *et al.* Whole-genome sequencing for prediction of *Mycobacterium tuberculosis* drug susceptibility and resistance: A retrospective cohort study. *Lancet Infect. Dis.* [https://doi.org/10.1016/S1473-3099\(15\)00062-6](https://doi.org/10.1016/S1473-3099(15)00062-6) (2015).
6. Pankhurst, L. J. *et al.* Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: A prospective study. *Lancet Respir. Med.* [https://doi.org/10.1016/S2213-2600\(15\)00466-X](https://doi.org/10.1016/S2213-2600(15)00466-X) (2016).
7. van der Werf, M. J. & Ködmön, C. Whole-genome sequencing as tool for investigating international tuberculosis outbreaks: A systematic review. *Front. Public Health* <https://doi.org/10.3389/fpubh.2019.00087> (2019).
8. Lalor, M. K. *et al.* The use of whole-genome sequencing in cluster investigation of a multidrug-resistant tuberculosis outbreak. *Eur. Respir. J.* <https://doi.org/10.1183/13993003.02313-2017> (2018).
9. WHO. *Global Tuberculosis Report*. Geneva: World Health Organization (2018).
10. De Viana, P. V. S., Gonçalves, M. J. F. & Basta, P. C. Ethnic and racial inequalities in notified cases of tuberculosis in Brazil. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0154658> (2016).
11. Malacarne, J. *et al.* Prevalence and factors associated with latent tuberculosis infection in an indigenous population in the Brazilian Amazon. *Rev. Soc. Bras. Med. Trop.* <https://doi.org/10.1590/0037-8682-0220-2016> (2016).
12. Cunha, E. A. *et al.* Incidence and transmission patterns of tuberculosis among indigenous populations in Brazil. *Mem. Inst. Oswaldo Cruz* <https://doi.org/10.1590/0074-0276130082> (2014).
13. Basta, P. C. *et al.* Social inequalities and tuberculosis: An analysis by race/color in mato grosso do sul, Brazil. *Rev. Saude Publica* <https://doi.org/10.1590/S0034-8910.2013047004628> (2013).
14. Malacarne, J. *et al.* Factors associated with TB in an indigenous population in Brazil: The effect of a cash transfer program. *Rev. Inst. Med. Trop. Sao Paulo* <https://doi.org/10.1590/S1678-9946201860063> (2018).
15. Freitas, L. P., Souza-Santos, R., Kolte, I. V., Malacarne, J. & Basta, P. C. Socioeconomic status of indigenous peoples with active tuberculosis in Brazil: A principal components analysis. *bioRxiv* <https://doi.org/10.1101/290668> (2018).
16. Coll, F. *et al.* A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat. Commun.* <https://doi.org/10.1038/ncomms5812> (2014).
17. Hillis, D. M. & Bull, J. J. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* **42**, 182–192 (1993).
18. Brynildsrud, O. B. *et al.* Global expansion of *Mycobacterium tuberculosis* lineage 4 shaped by colonial migration and local adaptation. *Sci. Adv.* <https://doi.org/10.1126/sciadv.aat5869> (2018).
19. Sreevatsan, S., Pan, X., Zhang, Y., Kreiswirth, B. N. & Musser, J. M. Mutations associated with pyrazinamide resistance in *pncA* of *Mycobacterium tuberculosis* complex organisms. *Antimicrob. Agents Chemother.* **41**, 636–640 (1997).
20. Sreevatsan, S. *et al.* Identification of a polymorphic nucleotide in *oxyR* specific for *Mycobacterium bovis*. *J. Clin. Microbiol.* **34**, 2007–2010 (1996).
21. Seifert, M., Catanzaro, D., Catanzaro, A. & Rodwell, T. C. Genetic mutations associated with isoniazid resistance in *Mycobacterium tuberculosis*: A systematic review. *PLoS ONE* **10**, e0119628 (2015).
22. Dantie, D., Weldeyohannes, D. & Mathewos, B. Review on molecular mechanism of first line antibiotic resistance in *Mycobacterium tuberculosis*. *Mycobact. Dis.* **04**, 4–7 (2014).
23. Vasconcellos, S. E. G. *et al.* Strain classification of *Mycobacterium tuberculosis* isolates in Brazil based on genotypes obtained by spoligotyping, mycobacterial interspersed repetitive unit typing and the presence of large sequence and single nucleotide polymorphism. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0107747> (2014).
24. Mendes, N. H. *et al.* Characterization of the genetic diversity of *Mycobacterium tuberculosis* in São Paulo city, Brazil. *BMC Res. Notes* <https://doi.org/10.1186/1756-0500-4-269> (2011).
25. Dantas, N. G. T. *et al.* Genetic diversity and molecular epidemiology of multidrug-resistant *Mycobacterium tuberculosis* in Minas Gerais State, Brazil. *BMC Infect. Dis.* <https://doi.org/10.1186/s12879-015-1057-y> (2015).
26. Verza, M. *et al.* Genomic epidemiology of *Mycobacterium tuberculosis* in Santa Catarina, Southern Brazil. *Sci. Rep.* <https://doi.org/10.1038/s41598-020-69755-9> (2020).
27. Bjorn-Mortensen, K. *et al.* Tracing *Mycobacterium tuberculosis* transmission by whole genome sequencing in a high incidence setting: A retrospective population-based study in East Greenland. *Sci. Rep.* <https://doi.org/10.1038/srep33180> (2016).
28. Leea, R. S. *et al.* Population genomics of *Mycobacterium tuberculosis* in the Inuit. *Proc. Natl. Acad. Sci. USA.* <https://doi.org/10.1073/pnas.1507071112> (2015).
29. Mulholland, C. V. *et al.* Dispersal of *Mycobacterium tuberculosis* driven by historical European trade in the South Pacific. *Front. Microbiol.* <https://doi.org/10.3389/fmicb.2019.02778> (2019).
30. Maes, M., Kremer, K., van Soolingen, D., Takiff, H. & de Waard, J. H. 24-Locus MIRU-VNTR genotyping is a useful tool to study the molecular epidemiology of tuberculosis among Warao Amerindians in Venezuela. *Tuberculosis* <https://doi.org/10.1016/j.tube.2008.04.003> (2008).
31. Sacchi, F. P. C. *et al.* Genetic clustering of tuberculosis in an indigenous community of Brazil. *Am. J. Trop. Med. Hyg.* <https://doi.org/10.4269/ajtmh.17-0480> (2018).
32. Hazbon, M. H. *et al.* Population genetics study of isoniazid resistance mutations and evolution of multidrug-resistant *Mycobacterium tuberculosis*. *Am. J. Trop. Med. Hyg.* **50**, 2640–2649 (2006).
33. Luo, D. *et al.* Prevalence and molecular characterization of multidrug-resistant *M. tuberculosis* in Jiangxi province, China. *Sci. Rep.* **9**, 1–8 (2019).
34. Jagielski, T. *et al.* Mutation profiling for detection of isoniazid resistance in *Mycobacterium tuberculosis* clinical isolates. *J. Antimicrob. Chemother.* **70**, 3214–3221 (2015).
35. Ramaswamy, S. V. *et al.* Single nucleotide polymorphisms in genes associated with isoniazid resistance in *Mycobacterium tuberculosis*. *J. Antimicrob. Chemother.* **47**, 1241–1250 (2003).
36. Sun, Y., Lee, A. S. G., Wong, S. & Paton, N. I. Analysis of the role of *Mycobacterium tuberculosis* *kasA* gene mutations in isoniazid resistance. *Clin. Microbiol. Infect.* **13**, 833–835 (2007).
37. Nusrath Unissa, A., Subbian, S., Elizabeth, L. & Selvakumar, N. Infection, genetics and evolution overview on mechanisms of isoniazid action and resistance in *Mycobacterium tuberculosis*. *Microbial. Infect.* **45**, 474–492 (2016).
38. Datta, G. *et al.* Longitudinal whole genome analysis of pre and post drug treatment *Mycobacterium tuberculosis* isolates reveals progressive steps to drug resistance. *Tuberculosis* **98**, 50–55 (2016).
39. Comas, I. *et al.* Whole-genome sequencing of rifampicin-resistant *Mycobacterium tuberculosis* strains identifies compensatory mutations in RNA polymerase genes. *Nat. Genet.* <https://doi.org/10.1038/ng.1038> (2012).
40. Campbell, E. A. *et al.* Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Nat. Genet.* **104**, 901–912 (2001).
41. Zuo, T., Liu, Q., Gan, M. & Gao, Q. Comprehensive identification of compensatory mutations in rifampicin-resistant *Mycobacterium tuberculosis* strains. *Chinese J. Tuberc. Respir. Med.* **41**, 207–212 (2018).

42. Zaw, M. T., Emran, N. A. & Lin, Z. Mutations inside rifampicin-resistance determining region of *rpoB* gene associated with rifampicin-resistance in *Mycobacterium tuberculosis*. *J. Infect. Public Health*. <https://doi.org/10.1016/j.jiph.2018.04.005> (2018).
43. Prim, R. I. *et al.* Molecular profiling of drug resistant isolates of *Mycobacterium tuberculosis* in the state of Santa Catarina, southern Brazil. *Mem. Inst. Oswaldo Cruz* **110**, 618–623 (2015).
44. Nebenzahl-guimaraes, H., Jacobson, K. R., Farhat, M. R. & Murray, M. B. Systematic review of allelic exchange experiments aimed at identifying mutations that confer drug resistance in *Mycobacterium tuberculosis*. *J. Antimicrob. Chem.* **69**, 331–342. <https://doi.org/10.1093/jac/dkt358> (2014).
45. Johnson, R. *et al.* Drug resistance in *Mycobacterium tuberculosis*. *Curr. Issues Mol. Biol* **8**, 97–112 (2006).
46. Malik, S., Willby, M., Sikes, D., Tsodikov, O. V. & Posey, J. E. New insights into fluoroquinolone resistance in *Mycobacterium tuberculosis*: Functional genetic analysis of *gyrA* and *gyrB* mutations. *PLoS ONE* **7**, e3956 (2012).
47. de Oliveira, T. L. R., Cavalcante, F. S., Chamon, R. C., Ferreira, R. B. R. & dos Santos, K. R. N. Genetic mutations in the quinolone resistance-determining region are related to changes in the epidemiological profile of methicillin-resistant *Staphylococcus aureus* isolates. *J. Glob. Antimicrob. Resist.* **19**, 236–240 (2019).
48. Minarini, L. A. R. & Darini, A. L. C. Mutations in the quinolone resistance-determining regions of *gyrA* and *parC* in enterobacteriaceae isolates from Brazil. *Braz. J. Microbiol.* <https://doi.org/10.1590/S1517-83822012000400010> (2012).
49. de Oliveira Santos, I. C. *et al.* Epidemiology and antibiotic resistance trends in clinical isolates of *Pseudomonas aeruginosa* from Rio de Janeiro, Brazil: Importance of mutational mechanisms over the years (1995–2015). *Infect. Genet. Evol.* <https://doi.org/10.1016/j.meegid.2019.05.015> (2019).
50. Almeida, F. *et al.* Phylogenetic and antimicrobial resistance gene analysis of *Salmonella* Typhimurium strains isolated in Brazil by whole genome sequencing. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0201882> (2018).
51. Marques, M., Cunha, E. A. T., Ruffino-Netto, A. & de Oliveira Andrade, S. M. Drug resistance profile of *Mycobacterium tuberculosis* in the state of Mato Grosso do Sul, Brazil, 2000–2006. *J. Bras. Pneumol.* **36**, 224–231 (2010).
52. Tauli-Corpus, V., Peoples, U. H. R. C. S. R. on R. of I. & Secretariat, U. H. R. C. *Report of the Special Rapporteur on the rights of indigenous peoples on her mission to Brazil*. (United Nations, Geneva, 2016). <https://doi.org/10.1017/s0020818300024334>
53. CIMI. *Violence Against Indigenous Peoples in Brazil: Data for 2018*. (2018).
54. Sacchi, F. P. C., Croda, M. G., Estevan, A. O., Ko, A. I. & Croda, J. Sugar cane manufacturing is associated with tuberculosis in an indigenous population in Brazil. *Trans. R. Soc. Trop. Med. Hyg.* **107**, 152–157 (2013).
55. Lemos, E. F. *et al.* Health-service performance of TB treatment for indigenous and non-indigenous populations in Brazil: A cross-sectional study. *BMC Health Serv. Res.* <https://doi.org/10.1186/1472-6963-14-237> (2014).
56. Brasil, F. *O Direito Humano à Alimentação Adequada e à Nutrição do povo Guarani e Kaiowá: um enfoque holístico: Resumo Executivo*. (2016).
57. *Manual de para o Controle da Tuberculose. Ministério da Saúde* (Ministério da Saúde, 2019).
58. Malacarne, J. *et al.* Performance of diagnostic tests for pulmonary tuberculosis in indigenous populations in Brazil: The contribution of Rapid Molecular Testing. *J. Bras. Pneumol.* <https://doi.org/10.1590/1806-3713/e20180185> (2019).
59. Nema, V. Tuberculosis diagnostics: Challenges and opportunities. *Lung India* **29**, 259–266 (2012).
60. Van Embden, J. D. A. *et al.* Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: Recommendations for a standardized methodology. *J. Clin. Microbiol.* **31**, 406–409 (1993).
61. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btu170> (2014).
62. Seemann, T. *Reads, Snippy: Fast Bacterial Variant Calling from NGS*. <https://github.com/tseemann/snippy%0A>.
63. Malm, S. *et al.* New *Mycobacterium tuberculosis* complex sublineage, Brazzaville, Congo. *Emerg. Infect. Dis.* <https://doi.org/10.3201/eid2303.160679> (2017).
64. Jackman, S. D. *et al.* ABySS 2.0: Resource-efficient assembly of large genomes using a Bloom filter. *Genome Res.* <https://doi.org/10.1101/gr.214346.116> (2017).
65. Kumar, S., Stecher, G., Li, M., Nknyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* <https://doi.org/10.1093/molbev/msy096> (2018).
66. Ronquist, F. *et al.* MrBayes 3.2: Efficient bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* <https://doi.org/10.1093/sysbio/sys029> (2012).
67. Inkscape v1.0. (2020). <https://inkscape.org/>.

Acknowledgements

Research in the Sreevatsan lab is funded by USDA (2018-67015-28288) and start-up funds provided by College of Veterinary Medicine, Michigan State University. The research project “Social Inequalities and Tuberculosis: Transmission Dynamics, Living Conditions and Interfaces between Biomedicine and Traditional Indigenous Medicine” received support from the INOVA program at the National School of Public Health at FIOCRUZ, Rio de Janeiro. Ida V Kolte received support from the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) under the Science without Borders program (#88887.091631/2014-00) to conduct this study. Daniel AM Villela is a scholar from National Council for Scientific and Technological Development. Syeda Anum Hadi is supported by Fulbright fellowship program. We greatly appreciate the prompt and helpful guidance from Ola B. Brynildsrud and Vegard Eldholm in obtaining and processing the Brazilian local references for further comparison.

Author contributions

A.H. and E.B. performed genomic analyses, extracted polymorphisms from genomes, and developed the phylogenomics models; Ida Kolte performed most of the shoe-leather epidemiology, collection of isolates, and selecting specific isolates for genome sequencing. All three helped design the study and wrote the manuscript. E.A.T.C., V.S., L.F., D.A.M.V., and R.S.S. helped collect field data, worked on microbial isolation and DNA extraction for genomics. J.R. helped in genome assembly and alignments and helped in editing the manuscript. P.B. is a physician who was in direct contact with indigenous patients and collected their samples for bacterial isolation. He also designed the prospective analysis with other coauthors in his institute. S.S. designed the genomics experiments, developed a plan for analysis and presentation, obtained funding, and oversaw the project along with Dr. P.C.B. All authors contributed to the writing and editing of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-020-79621-3>.

Correspondence and requests for materials should be addressed to S.S. or P.C.B.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2021