







Article

Genetic Characterization and Population Structure of Drug-Resistant *Mycobacterium tuberculosis* Isolated from Brazilian Patients Using Whole-Genome Sequencing

Leonardo Souza Esteves^{1,2,3,*}, Lia Lima Gomes³, Daniela Brites^{4,5}, Fátima Cristina Onofre Fandinho⁶, Marcela Bhering⁶ , Márcia Aparecida da Silva Pereira⁶, Emilyn Costa Conceição⁷ , Richard Salvato² , Bianca Porphirio da Costa⁶, Reginalda Ferreira de Melo Medeiros⁶, Paulo Cesar de Souza Caldas⁶, Paulo Redner⁶ , Margareth Pretti Dalcolmo⁶ , Vegard Eldholm⁸, Sebastien Gagneux^{4,5}, Maria Lucia Rossetti⁹, Afrânio Lineu Kritski^{1,†}  and Philip Noel Suffys^{3,†}

¹ Programa Acadêmico de Tuberculose da Faculdade de Medicina, Centro de Ciências da Saúde, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro 21941-590, RJ, Brazil; kritskia@gmail.com

² Centro de Desenvolvimento Científico e Tecnológico (CDCT), Secretaria Estadual de Saúde (SES-RS), Porto Alegre 90450-190, RS, Brazil; richardsalvato@hotmail.com

³ Laboratório de Biologia Molecular Aplicado à Micobactérias, Fundação Oswaldo Cruz (FIOCRUZ), Instituto Oswaldo Cruz (IOC), Rio de Janeiro 21040-360, RJ, Brazil; lialima2212@gmail.com (L.L.G.); psuffys@gmail.com (P.N.S.)

⁴ Swiss Tropical and Public Health Institute (Swiss TPH), CH-4123 Allschwil, Switzerland; d.brites@swisstph.ch (D.B.); sebastien.gagneux@swisstph.ch (S.G.)

⁵ University of Basel, CH-4001 Basel, Switzerland

⁶ Centro de Referência Professor Hélio Fraga, Fundação Oswaldo Cruz (FIOCRUZ), Rio de Janeiro 22780-195, RJ, Brazil; fatima.fandinho@fiocruz.br (F.C.O.F.); marcelabhering@yahoo.com.br (M.B.); marcapas2011@gmail.com (M.A.d.S.P.); bporphirio@yahoo.com.br (B.P.d.C.); regimedeiros@hotmail.com (R.F.d.M.M.); paulocesarcaldas@ensp.fiocruz.br (P.C.d.S.C.); predner@gmail.com (P.R.); margarethdalcolmo@ensp.fiocruz.br (M.P.D.)

⁷ Department of Science and Innovation—National Research Foundation Centre of Excellence for Biomedical Tuberculosis Research, South African Medical Research Council Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town 7505, South Africa; emilyncosta@gmail.com

⁸ Norwegian Institute of Public Health, 0213 Oslo, Norway; v.eldholm@gmail.com

⁹ Laboratório de Biologia Molecular, Universidade Luterana do Brasil (ULBRA), Canoas 92425-020, RS, Brazil; mrossett@terra.com.br

* Correspondence: biomedicineleo@yahoo.com.br; Tel.: +55-(51)-994151349

† These authors contributed equally to this work.



Citation: Esteves, L.S.; Gomes, L.L.; Brites, D.; Fandinho, F.C.O.; Bhering, M.; Pereira, M.A.d.S.; Conceição, E.C.; Salvato, R.; Costa, B.P.d.; Medeiros, R.F.d.M.; et al. Genetic Characterization and Population Structure of Drug-Resistant *Mycobacterium tuberculosis* Isolated from Brazilian Patients Using Whole-Genome Sequencing. *Antibiotics* **2024**, *13*, 496. <https://doi.org/10.3390/antibiotics13060496>

Academic Editors: Maria Rosario Rodicio and Carlos Rodriguez-Lucas

Received: 12 April 2024

Revised: 21 May 2024

Accepted: 22 May 2024

Published: 28 May 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: The present study aimed to determine the genetic diversity of isolates of *Mycobacterium tuberculosis* (Mtb) from presumed drug-resistant tuberculosis patients from several states of Brazil. The isolates had been submitted to conventional drug susceptibility testing for first- and second-line drugs. Multidrug-resistant (MDR-TB) (54.8%) was the most frequent phenotypic resistance profile, in addition to an important high frequency of pre-extensive resistance (p-XDR-TB) (9.2%). Using whole-genome sequencing (WGS), we characterized 298 Mtb isolates from Brazil. Besides the analysis of genotype distribution and possible correlations between molecular and clinical data, we determined the performance of an in-house WGS pipeline with other online pipelines for Mtb lineages and drug resistance profile definitions. Sub-lineage 4.3 (52%) was the most frequent genotype, and the genomic approach revealed a p-XDR-TB level of 22.5%. We detected twenty novel mutations in three resistance genes, and six of these were observed in eight phenotypically resistant isolates. A cluster analysis of 170 isolates showed that 43.5% of the TB patients belonged to 24 genomic clusters, suggesting considerable ongoing transmission of DR-TB, including two interstate transmissions. The in-house WGS pipeline showed the best overall performance in drug resistance prediction, presenting the best accuracy values for five of the nine drugs tested. Significant associations were observed between suffering from fatal disease and genotypic p-XDR-TB ($p = 0.03$) and either phenotypic ($p = 0.006$) or genotypic ($p = 0.0007$) ethambutol resistance. The use of WGS analysis improved our understanding of the population structure of MTBC in Brazil and the genetic and clinical data correlations and

demonstrated its utility for surveillance efforts regarding the spread of DR-TB, hopefully helping to avoid the emergence of even more resistant strains and to reduce TB incidence and mortality rates.

Keywords: *Mycobacterium tuberculosis*; drug resistance; whole-genome sequencing; genetic diversity; Brazil; novel mutations

1. Introduction

Tuberculosis (TB) accounted for 1.4 million deaths and an estimated 10.6 million new cases in 2023, and was the leading cause of death from a single infectious agent worldwide until the COVID-19 pandemic [1]. The emergence of drug-resistant (DR) TB is a global threat that hinders successful TB treatment. Multidrug-resistant TB (MDR-TB), defined as the simultaneous resistance to rifampicin (RIF) and isoniazid (INH), results in a worse prognosis, prolonged TB treatment with second-line drugs that are more toxic, more expensive, and that could possibly evolve into pre-extensive drug-resistant TB (p-XDR-TB). The latter is defined as strains that fulfill the definition of multidrug-resistant or rifampicin-resistant (MDR/RR-TB), plus resistance to any fluoroquinolone, and subsequently can evolve into extensive drug-resistant TB (XDR-TB), defined as strains p-XDR-TB plus resistance to bedaquiline and/or linezolid.

Brazil is one of the thirty countries with the highest TB burden, and although the incidence rate decreased until 2014, during the period of 2015–2019, a notified increased incidence of 34.3 to 37.4 cases/100.000 inhabitants was observed [2]. In addition, due to the COVID-19 pandemic, it is estimated that after a decade of decline, TB mortality has increased in Brazil and globally [3]. An additional concern for TB control in Brazil is the considerable number of DR-TB patients; in 2018, around 2,500 MDR/RR-TB cases were estimated [4], including an increase in MDR-TB among patients that had not previously been treated in the Rio de Janeiro State [5].

Rapid DR-TB detection and epidemiological surveillance, as well as knowledge about the genetic diversity of isolates of the *Mycobacterium tuberculosis* complex (MTBC) in different settings, are factors that may contribute to DR-TB elimination. In this scenario, molecular tools have become important, and quite recently, next-generation sequencing (NGS) has made it possible to quickly characterize the whole genome of MTBC strains, enabling the identification of both resistance-related genetic variants and lineages involved in ongoing transmission [6,7]. The culture-based phenotypic drug susceptibility test (DST), although still the current gold standard, has limitations due to the slow growth rate of MTBC organisms. Thus, molecular methods for drug resistance prediction are being steadily introduced as a routine in low-TB-incidence countries [8].

WGS is a promising tool and an approach to DR/MDR-TB detection, since it provides detailed sequence information from different genomic regions, thus enabling drug resistance prediction [9]. However, the high amount of sequencing data generated by WGS has created the challenge to develop bioinformatics tools to translate the data into information of clinical and laboratory interest [8]. To permit the use of WGS by professionals with little or no bioinformatics skills, user-friendly tools for the analysis and interpretation of WGS data have been developed and implemented, permitting the accessibility and broad implementation of NGS-based approaches [10–13]. Nonetheless, due to the complexity of large-scale data analyses, some bioinformatics command-line skills are still required and sometimes, a user-friendly graphical interface is not available [14,15].

Due to the global variability in the prevalence of MTBC lineages and the evidence of a differential association with drug resistance, it is evident that the combined detection of both characteristics in different countries and regions may interfere with DR-TB management [16]. Therefore, the genomic information of MTBC strains and a conventional phenotypic DST regarding clinical outcomes are urgently needed [17–19]. In addition, the evaluation of the different available tools for the extraction of a DST profile from WGS

has become important to evaluating the regional differences in DR-TB surveillance and to understanding local TB transmission [20–23].

Here we conducted a genetic diversity study on genomes from a large collection of Mtb isolates from several states of Brazil that mostly had a phenotypic DST for primary and secondary drugs. WGS data were evaluated by an in-house WGS pipeline and different online available pipelines, including Mykrobe Predictor, KvarQ, TB Profiler version 0.3.4, and the more recent TB Profiler 5.0 to predict drug susceptibility, and the in-house WGS pipeline, KvarQ, TB Profiler 5.0, and RD-Analyzer to predict the genotype (TB lineages).

2. Results

2.1. Study Samples, Patient Data, and Phenotypic Drug Susceptibility Testing

Among the 298 Mtb isolates from presumed DR-TB patients with high-quality WGS data, 294 had conventional phenotypic DST data for at least one drug. The phenotypic DST resulted in 64 (21.8%) pan-susceptible, 22 (7.5%) isoniazid-monoresistant (IMR), 9 (3.1%) rifampicin-monoresistant (RMR); 1 (0.3%) ethambutol-monoresistant (EMR), 10 (3.4%) poly-resistant (Poly-R), 161 (54.8%) MDR, and 27 (9.2%) p-XDR (Supplementary Table S1).

Among the 170 patients for which we had identification, 112 had sociodemographic, clinical, radiographic, and treatment outcome data available (Supplementary Table S2). Among them, 65 (57.5%) were male, 7 (6.3%) were smokers, 8 (7.1%) were illicit drug users, 16 (14.2%) had diabetes, 12 (10.7%) were alcohol users, and 7 (6.3%) were HIV-infected. In relation to chest radiographic images, 72 patients (64.3%) had bilateral cavitory disease, 15 (13.4%) had unilateral, and 16 (14.3%) had no image available. Regarding treatment outcomes, 50 (44.6%) were cured or completed treatment, 32 (28.6%) died due to TB, 16 (14.3%) were lost to follow-up, 5 (4.5%) were still under treatment, 5 (4.5%) died from another cause, and 4 (3.6%) patients had no clinical data available.

2.2. Novel Mutations, In Silico Drug Susceptibility, and Resistance Prediction Using Different Pipelines

Considering the in-house WGS pipeline, the 298 sequenced genomes were classified as follows: 80 (26.8%) pan-susceptible, 11 (3.7%) IMR, 4 (1.3%) RMR, 1 (0.3%) EMR, 1 (0.3%) SMR, 10 (3.4%) Poly-R, 124 (41.6%) MDR, and 67 (22.5%) p-XDR. The nature and frequencies of the mutations detected with this pipeline are presented in Figure 1 and in Supplementary Table S3.

Using the 294 genomes that had a DST available for at least one drug, we compared the performance in drug resistance using five predictive drug resistance pipelines against the phenotypic DST (three different pipelines and two versions of TB Profiler). The in-house WGS pipeline showed slightly better overall results when considering sensitivity, specificity, positive and negative predictive values, and accuracy compared to Mykrobe Predictor, both versions of TB Profiler, and KvarQ (Table 1). For the prediction of resistance to RIF and INH, in all tools, the sensitivity was higher than 80% but the specificity was lower than 95%. Regarding the prediction of other first-line drugs, the in-house WGS pipeline showed sensitivity higher than 80% to EMB, while high specificity to STR was observed using the in-house WGS pipeline and Mykrobe Predictor. With respect to second-line drugs, the highest sensitivity to FQ ofloxacin (OFL) was obtained using TB Profiler, and over 90% specificity to AMK, KAN, and CAP was observed in all pipelines. The highest sensitivity to PZA was obtained in the most updated version of TB Profiler (68%).

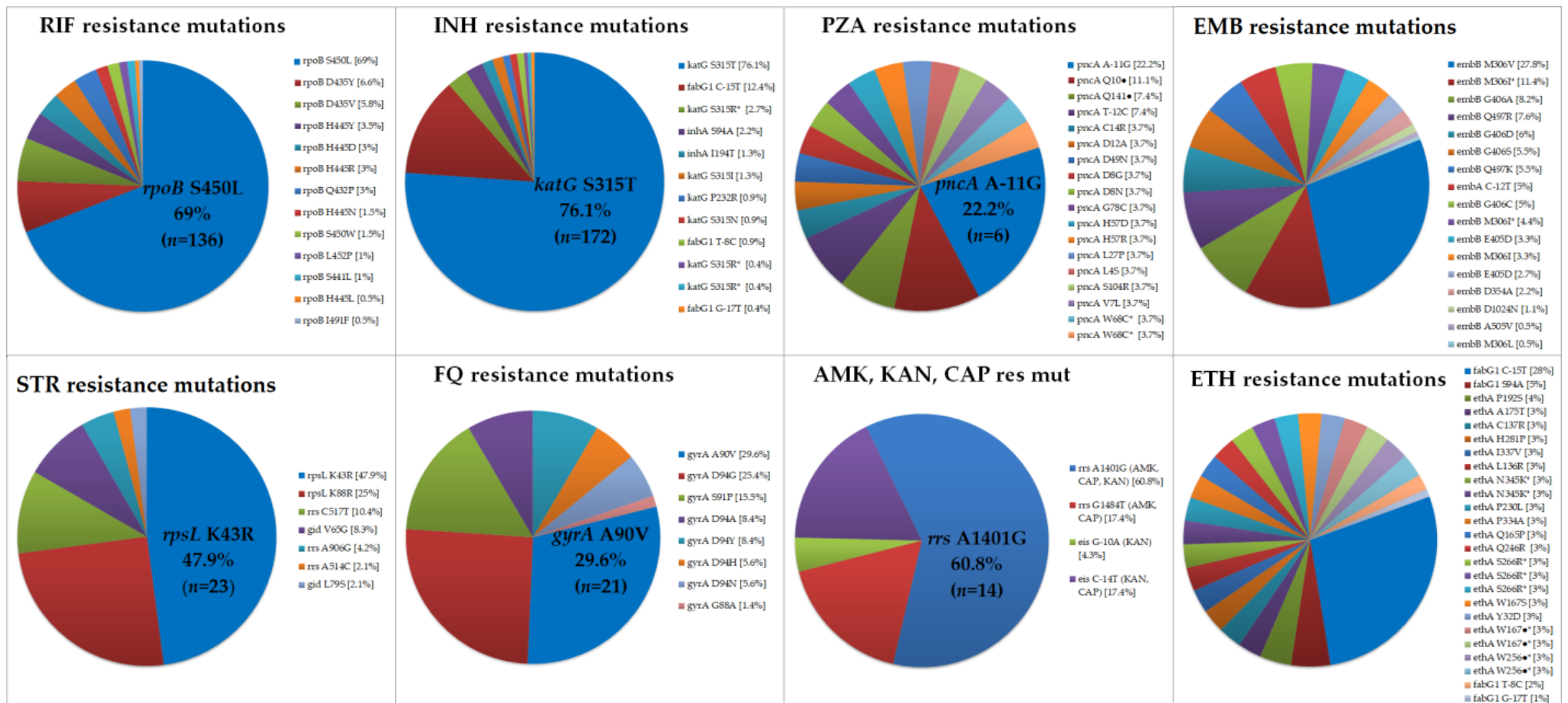


Figure 1. Proportion of resistance-related mutations of 298 isolates included in the study based on in-house WGS pipeline analysis. The most frequent mutation found is highlighted. RIF = rifampicin; INH = isoniazid; PZA = pyrazinamide; EMB = ethambutol; STR = streptomycin; FQ = fluoroquinolone; AMK, KAN, CAP res mut = amikacin, kanamycin, capreomycin resistance mutation; ETH = ethionamide. * = same amino acid changed but different mutation involved. • = stop codon.

Table 1. Evaluation of WGS-based drug susceptibility prediction based on the in-house WGS pipeline, KvarQ, Mykrobe Predictor, TB Profiler, and TB Profiler 5.0 presenting sensitivity, specificity, positive and negative predictive values, and accuracy.

ATB	DST		WGS Pipeline					KvarQ					Mykrobe Predictor					TB Profiler					TB Profiler 5.0				
	RES	SUS	SEN	SPE	PPV	NPV	ACC	SEN	SPE	PPV	NPV	ACC	SEN	SPE	PPV	NPV	ACC	SEN	SPE	PPV	NPV	ACC	SEN	SPE	PPV	NPV	ACC
RIF ¹	199	95	89.95%	84.21%	92.27%	80%	88.1%	88.94%	84.21%	92.19%	78.43%	87.41%	89.95%	84.21%	92.27%	80%	88.1%	87.94%	82.11%	91.15%	76.47%	86.01%	89.95%	83.15%	91.79%	79.79%	87.75%
INH ¹	217	75	85.25%	88%	95.36%	67.35%	85.96%	83.87%	88%	95.29%	65.35%	84.93%	84.79%	88%	95.34%	66.67%	85.62%	88.94%	82.67%	93.69%	72.09%	87.33%	89.4%	85.33%	94.63%	73.56%	88.35%
PZA ¹	25	138	0%	94.2%	0%	83.87%	79.75%	40%	71.74%	20.41%	86.84%	66.87%	Na	Na	Na	Na	Na	24%	61.59%	10.17%	81.73%	55.83%	68%	50.72%	20%	89.74%	53.37%
EMB ¹	67	213	82.09%	65.73%	42.97%	92.11%	69.64%	68.66%	75.12%	46.46%	88.4%	73.57%	64.18%	78.4%	48.31%	87.43%	75%	76.12%	67.61%	42.5%	90%	69.64%	80.59%	65.73%	42.52%	91.5%	69.28%
STR ²	39	75	56.41%	97.33%	91.67%	81.11%	83.33%	48.72%	88%	67.86%	76.74%	74.56%	43.59%	97.33%	89.47%	76.84%	78.95%	53.85%	88%	70%	78.57%	76.32%	69.23%	94.66%	87.09%	85.54%	85.96%
OFL ²	29	45	75.86%	91.11%	84.62%	85.42%	85.14%	55.17%	93.33%	84.21%	76.36%	78.38%	72.41%	88.89%	80.77%	83.33%	82.43%	86.21%	80%	73.53%	90%	82.43%	79.31%	82.22%	74.19%	86.04%	81.08%
AMK ²	12	68	75%	97.06%	81.82%	95.65%	93.75%	41.67%	98.53%	83.33%	90.54%	90%	75%	98.53%	90%	95.71%	95%	75%	91.18%	60%	95.38%	88.75%	75%	92.64%	64.28%	95.45%	90%
KAN ²	7	29	42.86%	100%	100%	87.88%	88.89%	28.57%	100%	100%	85.29%	86.11%	42.86%	96.55%	75%	87.5%	86.11%	57.14%	96.55%	80%	90.32%	88.89%	42.86%	96.55%	75%	87.5%	86.11%
CAP ²	7	29	42.86%	96.55%	75%	87.5%	86.11%	0%	80.56%	0%	100%	80.56%	42.86%	96.55%	75%	87.5%	86.11%	42.86%	96.55%	75%	87.5%	86.11%	42.86%	96.55%	75%	87.5%	86.11%

ATB = antibiotic; RES = resistant; SUS = susceptible; SEN = sensitivity; SPE = specificity; PPV = positive predictive value; NPV = negative predictive value; ACC = accuracy. ¹ = first-line drugs; ² = second-line drugs.

The agreement among phenotypic DST and the in-house WGS pipeline, KvarQ, Mykrobe Predictor, TB Profiler 0.3.4, and TB Profiler 5.0 was assessed for all drugs considering resistance and susceptibility prediction (regardless of mutation type in pipelines). The highest agreement level among the six methods was found in RIF ($k = 0.89$), and the worst agreement level was in PZA ($k = 0.32$). An excellent level of agreement was also detected in INH, AMK, KAN, and CAP (Supplementary Table S4).

Using TB Profiler 5.0, which is able to detect and report unique mutations, i.e., those not yet reported, twenty novel mutations in three resistance genes (*katG*, *pncA*, and *ethA*) were identified, of which six were present in eight phenotypically resistant isolates (gene-drug-related), as shown in Table 2. Unfortunately, ethionamide susceptibility was not phenotypically characterized in the present study. Mostly isolates with novel mutations were phenotypically MDR (68.7%) and genotypically p-XDR (46.8%). All mutations observed using TB Profiler 5.0 are shown in Supplementary Table S5.

Table 2. Novel mutations observed in this study using TB Profiler 5.0.

n Isolates	Drug	DST	Novel Mutation	Phenotypic Profile	Genotypic Profile ‡
1	INH	R	<i>katG</i> _c.2070delC ¹	MDR	p-XDR
1	INH	S	<i>katG</i> _c.1141dupG ²	RMR	IMR
1	ETH	NA	<i>ethA</i> _c.306_307delCA	MDR	p-XDR
1	ETH	NA	<i>ethA</i> _c.-382_*857del	p-XDR	p-XDR
1	ETH	NA	<i>ethA</i> _c.40dupA	MDR	MDR
4	ETH	NA	<i>ethA</i> _c.851dupC	MDR	2 MDR/2 p-XDR
4	PZA	2R/2NA	<i>pncA</i> _c.193_200dupTCCTCGTC	MDR	MDR
1	PZA	NA	<i>pncA</i> _c.289_293dupGGTGC	MDR	MDR
1	PZA	NA	<i>pncA</i> _c.502delA	Poly-R	Poly-R
1	PZA	R	<i>pncA</i> _c.452dupT	MDR	p-XDR
2	PZA	R	<i>pncA</i> _c.75_79delCGCGC	p-XDR	p-XDR
4	PZA	R/2S/NA	<i>pncA</i> _c.443_444dupGC ³	MDR	2 MDR/2 p-XDR
1	PZA	S	<i>pncA</i> _c.117_124delGGACTACC	MDR	p-XDR
1	PZA	S	<i>pncA</i> _c.300delC	MDR	Poly-R
2	PZA	R/NA	<i>pncA</i> _c.305dupC	p-XDR/Poly-R	p-XDR
1	PZA	S	<i>pncA</i> _c.329_338delACGAGAACGG	p-XDR	Poly-R
1	PZA	S	<i>pncA</i> _c.-3449_*7353del	p-XDR	p-XDR
1	PZA	S	<i>pncA</i> _c.423_424delGA	p-XDR	p-XDR
2	PZA	S	<i>pncA</i> _c.454_455insT	MDR	MDR
1	PZA	S	<i>pncA</i> _c.527dupG	MDR	MDR

‡ = genotypic resistance profile classified according to TB Profiler 5.0. ¹ Genotypic isoniazid resistance profile = *ahpC*_c.-81C>T, *katG*_c.2070delC. ² Genotypic isoniazid resistance profile = *ahpC*_c.-48G>A, *katG*_c.1141dupG. ³ Genotypic pyrazinamide resistance profile = *pncA*_c.443_444dupGC, *pncA*_p.Phe81Val. Isolate PZA phenotypically susceptible. * = deletion beyond the boundaries of coding sequence.

2.3. Genomic Diversity, Phylogenetic Analysis, and Lineage Classification Using Different Pipelines

Among the 298 genomes analyzed by the in-house WGS pipeline, 99% belonged to lineage 4, of which 155 (52%) belonged to sub-lineage 4.3 [LAM], 42 (14.1%) to sub-lineage 4.10 [PGG3], 40 (13.4%) to sub-lineage 4.1.2 [Haarlem], 35 (11.7%) to sub-lineage 4.1.1 [X], 17 (5.7%) to sub-lineage 4.4 [Vietnam], 2 (0.6%) had not been sub-lineage-detected and 4 (1.3%) demonstrated a mixed classification of L4.3/L4.1.1 ($n = 1$), L4.3/L4.1.2 ($n = 1$), L4.3/L4.10 ($n = 1$), and L4.1.1/L4.4 ($n = 1$), indicating mixed infections or contaminations; three (1%) were classified as lineage 1.

A phylogenetic tree (Figure 2) based on 38,563 genome-wide SNPs was constructed using 293 genomes. Five genomes with a high count of mixed SNP calls were excluded (four with mixed sub-lineage classification and one L4.1.2 [Haarlem]). The topology was congruent with different lineage and sub-lineage classifications assigned by all SNP-based pipelines as previously described [24,25].

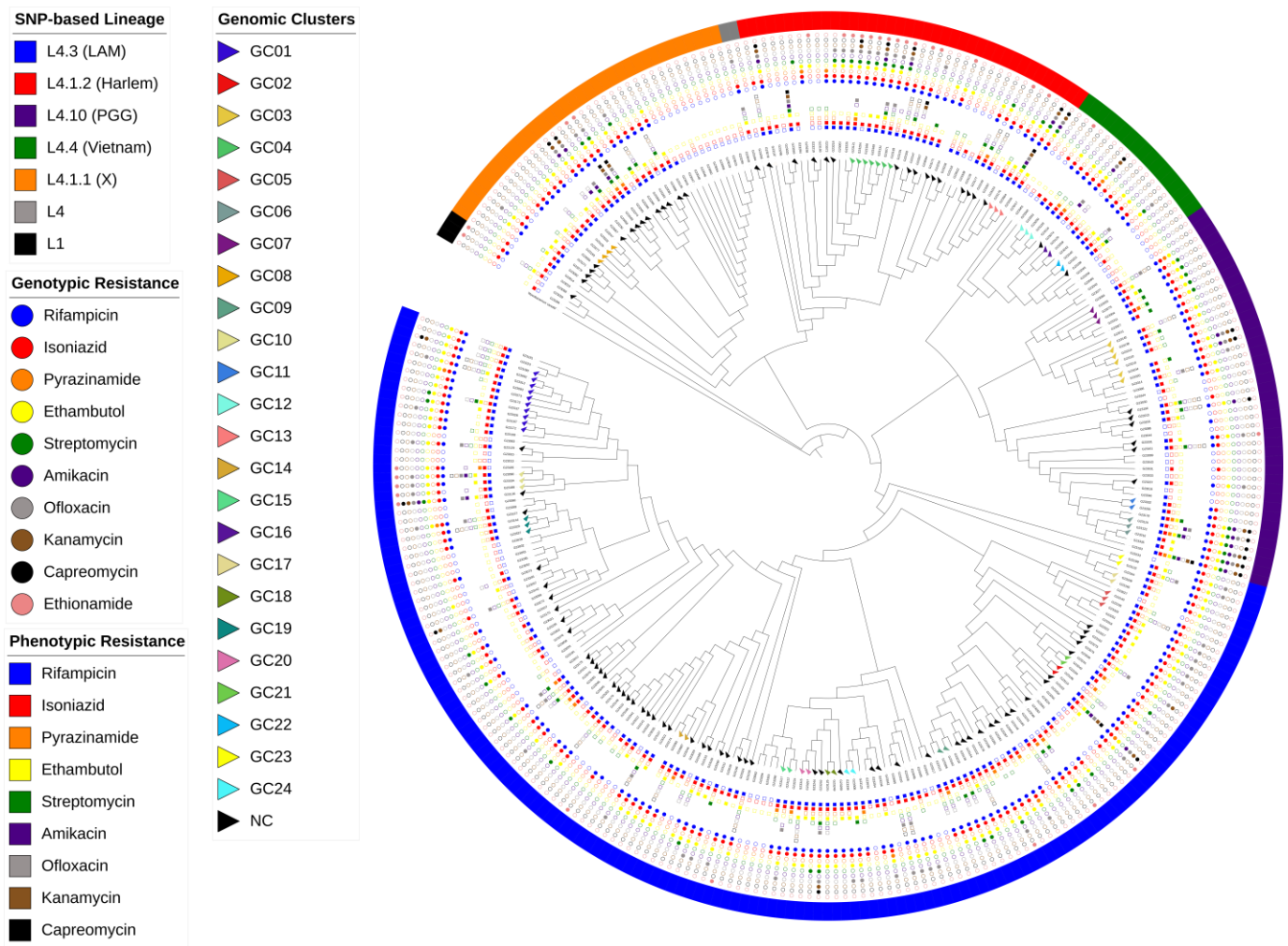


Figure 2. Maximum likelihood phylogenetic tree for the 293 *M. tuberculosis* isolates constructed based on 38,563 SNPs rooted with *M. canettii*. Presented from the outside to the inside: SNP-based lineage classification by the in-house WGS pipeline; genotypic resistance profile (circle); phenotypic resistance profile (square). Tips (triangle) are shown colored according to the genomic cluster from 170 patients with identification (see legend). GC = genomic cluster; NC = non-clustered. Tree branch scale represents the number of nucleotide substitutions per site.

Upon comparing the SNP-based lineage classification pipelines (in-house WGS pipeline, KvarQ-barcode-Coll14, and TB Profiler) and an RD-based analysis pipeline (RD-Analyzer), the L4.3 [LAM] (and higher resolutions classifications when available, e.g., L4.3.2, L4.3.3, L4.3.4, L4.3.4.1, L4.3.4.2, L4.3.4.2.1) was predominant with all tools. The classification results and lineage proportions are presented in Supplementary Figure S1.

Some discordant genotypic classifications were encountered and described in Supplementary Figure S2, mainly among RD-based pipeline analyses and SNP-based pipelines. Nevertheless, the free marginal kappa coefficient obtained presented an excellent agreement level ($k = 0.89$ [95%CI: 0.86–0.93]), and the overall agreement was 92.06%.

With respect to the sample origin (Figure 3), 212 (71.2%) were isolates from patients residing in the southeast region, including 173 (58.1%) isolates from Rio de Janeiro and 39 (13.1%) from São Paulo; 46 (15.4%) were from the Midwest region, including 42 (14.1%) from the Distrito Federal, 2 (0.7%) from Goiás, and 2 (0.7%) from Mato Grosso; 15 (4.9%) were from the northeast, including 6 (2%) from Pernambuco, 4 (1.3%) from Ceará, 3 (1%) from Maranhão, 1 (0.3%) from Piauí, and 1 (0.3%) from Sergipe; 9 (3%) were from the south region, including 6 (2%) from Santa Catarina and 3 (1%) from Paraná; finally, 6 (2%) were

from the north region, including 3 (1%) from Amazonas, 2 (0.7%) from Tocantins, and one (0.3%) from Acre. For eight (2.7%) isolates, we had no data regarding state origin.

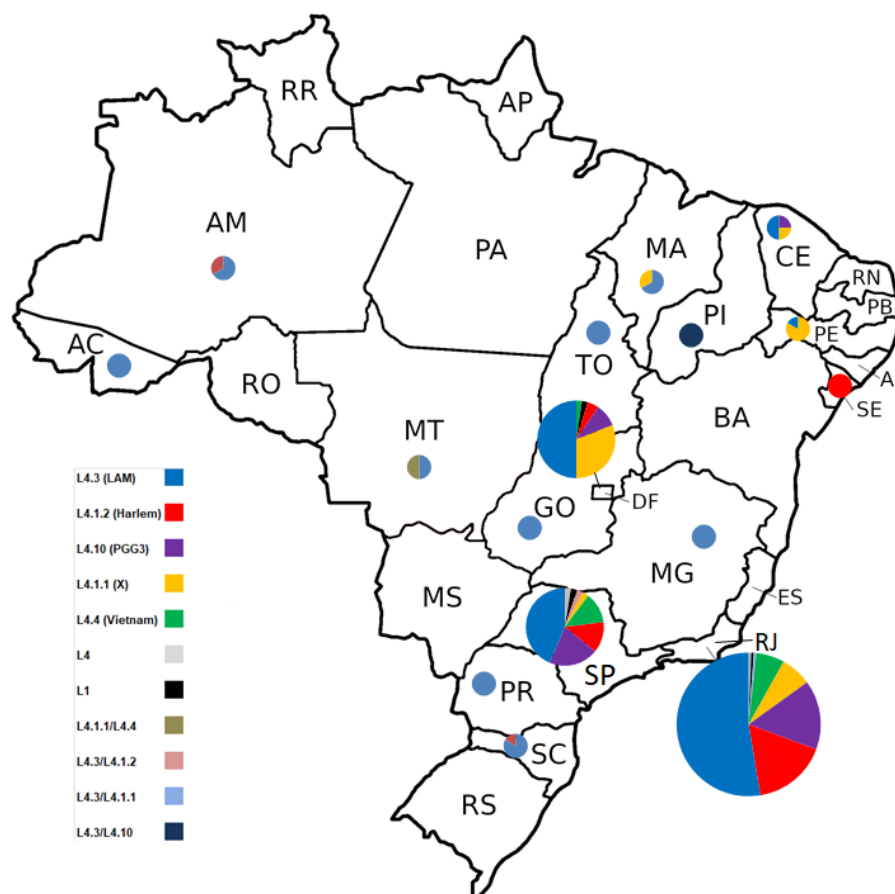


Figure 3. Geographic proportion distribution of *Mycobacterium tuberculosis* lineages classified by in-house WGS pipeline of 290 isolates. Eight isolates (three L4.3 [LAM], two L4.10 [PGG3], two L4.1.1 [X], and one L4.1.2 [Haarlem]) have no data regarding geographic origin.

2.4. Genomic Clusters Analysis

Among the 170 isolates with patient identification available, 74 (43.5%) belonged to 24 genomic clusters with sizes ranging from 2 to 10 isolates, therefore characterized by a clustering rate of 0.294. The most frequent genotypic resistance profile in this sampling was MDR-TB (53.5%), followed by p-XDR-TB (32.9%), and a similar distribution was observed among the clustered population, composed of 63.5% MDR-TB and 33.7% p-XDR-TB isolates (Figure 4). Importantly, a significant association between being part of a cluster and having an MDR-TB genotype was observed (Chi-squared = 5.2512; $p = 0.02$).

Upon analysis of the genotypes of these 170 isolates using the in-house WGS pipeline, 95 (55.8%) belonged to sub-lineage 4.3 [LAM], 26 (15.2%) were 4.1.2 [Haarlem], 20 (11.7%) were 4.10 [PGG3], 18 (10.5%) were 4.1.1 [X], 9 (5.3%) were 4.4 [Vietnam], 1 (0.6%) was classified as lineage 1, and 1 (0.6%) as lineage 4 without sub-lineage detection. Among the isolates in clusters, L4.3 LAM was also observed most frequently ($n = 39$; 52.7%).

When concentrating on the origins of these samples, 138 were from Rio de Janeiro (81.1%), 6 each from Pernambuco (3.5%) and Santa Catarina (3.5%), 4 from Ceará (2.3%), 3 each from Amazonas (1.7%), Maranhão (1.7%), and Paraná (1.7%), 2 each from Minas Gerais (1.1%) and Goiás (1.1%), and 1 each from Tocantins (0.6%), Acre (0.6%), and Mato Grosso (0.6%). An even more pronounced number of the clustered isolates (95.9%) were from Rio de Janeiro, while only three isolates were from other states (Ceará [GC8], Goiás, and Tocantins [GC15]). The latter explains why we observed two genomic clusters with

isolates belonging to patients from different states: GC8 with one from Ceará and three from Rio de Janeiro, and GC15 with one from Goiás and one from Tocantins.

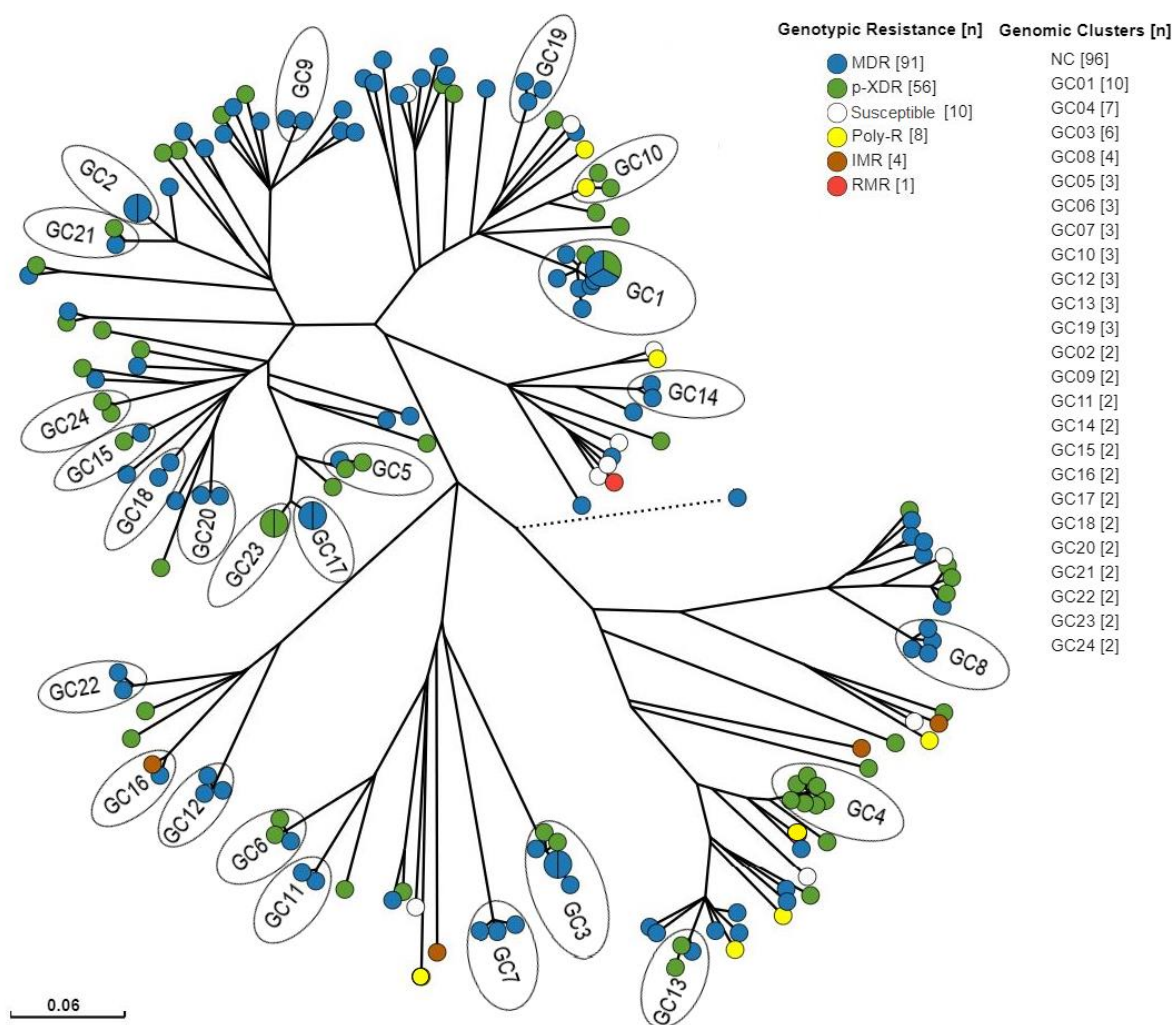


Figure 4. Radial tree of the 170 isolates used for clustering analysis. Tips were colored according to genotypic resistance profile. Pie charts in the tips represent two or more isolates with no SNP differences among genomes. NC = non-clustered. The dashed line represents the L1 isolate with greater distance in comparison to the rest of the isolates. Tree branch scale represents the number of nucleotide substitutions per site. [n] = number of isolates by category.

2.5. Evolution of Drug Susceptibility Patterns in Patients with Multiple Isolates of *Mycobacterium tuberculosis*

Among the 112 patients for whom clinical and epidemiologic data were available, 15 had at least 2 isolates with genomes sequenced, resulting in a total of 36 genomes: 2 patients with 4 isolates, 2 patients with 3 isolates, and 11 patients with 2 isolates each, who could have been included in follow-up (Table 3 and Supplementary Table S6). Only two patients presented a change in phenotypical drug resistance: One was susceptible to AMK and changed to being resistant without a change in mutational profile (patient P39). Another was phenotypically susceptible to KAN and CAP and changed to being resistant (patient P64), again without a change in mutational profile.

Table 3. Chronic TB patients with changes in genotypic profile mutations and phenotypic resistance.

Patient ID	Sample ID	Isolate Year	Resistance Mutation Genotypic Profile										Resistance Phenotypic Profile										Sex	Smoker	Diabetes	HIV	Lung lesion (First Available)	Lung lesion (Last Available)	Clinical Outcome	SNP Dist						
			INH	RIF	EMB	FQ	STR	AMK	CAP	KAN	PZA	ETH	Gen-DR	RIF	INH	PZA	EMB	STR	AMK	OFL	KAN	CAP									Sub-Lineage	Homo SNPs	Het SNPs	Spoligofamily /SIT		
P28	G23103	2010	S315T	S450L	M306V			K43R	A1401G	A1401G	A1401G			MDR	R	R	R	R	R	S	R	R	L4.10/PGG3	833	39	T1/167										
P28	G23335	2011	S315T	S450L	M306V	S91P		K43R	A1401G	A1401G	A1401G			p-XDR	R	R	-	R	-	-	-	-	L4.10/PGG3	839	6	T1/167										
P28	G23216	2011	S315T	S450L	M306V	S91P		K43R	A1401G	A1401G	A1401G			p-XDR	R	R	-	R	-	-	-	-	L4.10/PGG3	834	51	T1/167	M	No	No	Yes	UC	UC	Death	0–1		
P27	G23178	2008	S315T	S450L	M306V ¹			K43R	G1484T	G1484T				MDR	R	R	R	R	R	-	-	-	L4.10/PGG3	751	62	T1/167										
P27	G23124	2009	S315T	S450L	M306V	A90V		K43R	G1484T	G1484T				p-XDR	R	R	S	S	-	R	R	-	L4.10/PGG3	837	20	T1/167	M	No	No	No	UC	BC	Death	0		
P54	G23259	2012	S315T	S450L										MDR	R	R	S	S	-	-	-	-	L4.4/Vietnam	830	12	S/34										
P54	G23242	2012	S315T	S450L	G406D ²									MDR	R	R	S	S	-	-	-	-	L4.4/Vietnam	815	38	S/34	F	No	No	No	BC	BC	Death	0		
P63	G23094	2010		D435V										RMR	R	R	-	S	R	-	-	-	L4.1.2/Harlem	800	33	ND/Orphan										
P63	G23327	2011		D435V		D94H								p-XDR	R	R	S	S	-	-	-	-	L4.1.2/Harlem	805	10	ND/Orphan	M	No	Yes	No	BC	BC	Cured	3		
P35	G23277	2012	S315T ²	Q432P										MDR	R	R	S	S	-	-	-	-	L4.1.1/X	608	177	T1/53										
P35	G23280	2012	S315T	S450L	Q497R									MDR	R	R	S	S	-	-	-	-	L4.3/LAM	829	44	ND/Orphan	F	No	No	No	BC	BC	Abandonment	484		
P10	G23108	2008	S315T	S450L	M306V									MDR	R	R	-	S	S	-	-	-	L4.3/LAM	833	20	LAM6/64										
P10	G23167	2012	S315T	S450L	M306V	D94A								p-XDR	R	R	-	S	-	-	-	-	L4.3/LAM	826	52	LAM6/64	M	No	No	Yes	BC	BC	Death	2		
P39	G23185	2008	C-15T/S94A	S450L	G406C									C-15T/S94A	MDR	S	R	R	-	-	S	-	L4.3/LAM	849	25	LAM6/Orphan										
P39	G23090	2010	C-15T/S94A	S450L	G406C	D94G/A90V								C-15T/S94A	p-XDR	R	R	-	R	S	R	R	S	L4.3/LAM	837	18	LAM6/Orphan	F	No	Yes	No	BC	BC	Death	1	

¹ In addition to the mutation M306V, there was a gap in codons 405 and 406, which resulted in flag G→T E405D, G→C E405D, G→T G406C, G→A G406S, G→A G406D, and G→C G406A. ² Mixed SNP call. ND = not determined. Homo SNPs = all bases detected in the same genome position are the same. Het SNP = bases detected in the same genome position are different. UC = unilateral cavity. BC = bilateral cavity.

Regarding the mutational profile changes of those 15 patients, 7 patients (2 with 3 isolates each and 5 with 2 isolates each) had their genotypic drug resistance profile changed (Table 3). The other eight patients presented no change in the genomes of their respective isolates (genetic distance = 0 SNPs) (Supplementary Table S6).

The emergence of resistance mutations for FQ was observed in five patients in the *gyrA* gene (including P28 = S91P [T→C], P27 = A90V [C→T], P63 = D94H [G→C], P10 = D94A [A→C], and P39 = D94G [A→G]/A90V [C→T]); one patient was cured from the disease while the other four died of tuberculosis. Two patients had an EMB mutation resistance emergence in the *embB* gene (P54 = G406D [G→A] and P35 = Q497R [A→G]); the first was cured while the second patient abandoned treatment with no further information obtained.

2.6. Treatment Outcome, Risk Factors, and Lineage Associations with Genotypic and Phenotypic Resistance

The association between TB treatment outcome and genotypic and phenotypic drug resistance was evaluated. Genotypic p-XDR-TB was significantly associated with mortality (Chi-squared = 4.6231; $p = 0.03$). Both phenotypic resistance (Chi-squared = 7.3445; $p = 0.006$) and genotypic resistance to EMB (Fisher's exact test $p = 0.0007$) were significantly associated with mortality. Another significant association was EMB genotypic resistance (Chi-squared = 24.364; $p = 0.00001$) with the p-XDR-TB genotype. All isolates with a phenotypic EMB available are summarized in Supplementary Table S7. In short, the table shows the available EMB tests of isolates, divided into four groups: EMB genotypically susceptible and resistant and phenotypically susceptible and resistant; each of these groups was also divided into two groups: cured and deceased.

Compared to patients who had any risk factors for developing TB (HIV, diabetes, alcoholism, smoking, and/or illicit drugs abuse) (Supplementary Table S2), among the two outcome groups, deceased and cured, a significant association between the presence of any risk factor and a deceased outcome was observed (Chi-squared = 5.7672; $p = 0.01$).

Upon comparing of the frequency of the presence of drug-resistance-related mutations in Mtb isolates classified as L4.3 (LAM) to those of other sub-lineages, a statistically significantly lower frequency of mutations related to resistance to STR (Chi-squared = 5.3496; $p = 0.02$), CAP (Chi-squared = 6.4498; $p = 0.01$), ETH (Chi-squared = 5.7636; $p = 0.01$), and AMK (Fisher's exact test $p = 0.001$) was associated with the LAM sub-lineage in comparison to non-LAM isolates.

3. Discussion

In this study, we performed a genome-based characterization of the genetic diversity of mainly drug-resistant isolates of Mtb in different regions in Brazil. The analyzed Mtb strains predominantly belonged to the sub-lineage 4.3 (LAM) (52%). This is in accordance with previous studies performed in various Brazilian regions in both susceptible and DR-TB samples [26–29]. L4.3 was predominant in the south, north, southeast and midwest regions. In the northeast, this predominance of L4.3 was not observed, but the number of samples from this region was low. Three isolates (from São Paulo, Rio de Janeiro, and Distrito Federal) belonged to lineage 1, which are known to be restricted mostly to eastern Africa and the south and east of Asia [30], but are apparently described with a certain frequency in the northern region of Brazil [29,31,32], probably imported through slave trade from east Africa [23]. Although L1 is usually not associated with DR, two of the three isolates in this study presented some resistance (one IMR and one Poly-R). This might be due to sampling bias as the CRPHF mainly receives samples from patients suspicious of being DR, either due to treatment failure, treatment abandonment, relapse TB, or contact with TB-resistant patients.

The four pipelines presented a high agreement level ($k = 0.89$), with almost complete agreement among the SNP-based pipelines, but showing divergences mainly in the lineage classifications by RD-Analyzer. The Mtb isolates with high counts of mixed SNP calls may be indicative of a mixed infection, as described by others [33–35]. An analysis of the five

isolates with a high proportion of mixed SNP calls showed four mixed classifications by the in-house WGS pipeline only, while TB Profiler and KvarQ detected only one of the two lineages present in these isolates. RD-Analyzer had some difficulties in classifying lineage in comparison with the SNP-based tools and presented the most divergences among all the tools tested, including 31 isolates with a mixed classification and 6 unidentified despite having a good genome quality and low mixed SNP call counts.

The majority of clusters were identified in Rio de Janeiro, and although sampling from this state was clearly over-represented in our setting, this may represent higher levels of recent infection and drug-resistant TB transmission, differing from TB cases due to the reactivation of a latent infection [36,37]. In this Brazilian state, an increase in primary MDR-TB has been described as being associated with low TB control performance [5]. Studies in other Brazilian states have also shown significant ongoing transmission of MDR-TB, as observed in Sao Paulo [38,39] and Santa Catarina [40], and Rio Grande do Sul in a genome-based study [41]. In the latter, as well as in our study, a significant association between clustering and genotypic MDR-TB and significant ongoing transmission in p-XDR-TB resistance profiles was observed, with practically half of the patients infected with genotypic p-XDR-TB (44.6%; n = 25) in clusters, reflecting an alarming scenario of insufficient DR-TB control [41].

In addition to clusters of isolates derived from residents of particular states, we also observed clusters suggestive of the interstate transmission of particular genotypes that seem to have been circulating in the country for a considerable time. Examples are GC8 (three patients from Rio de Janeiro and one from Ceará) and GC15 (one from Goiás and one from Tocantins) presenting genetic distances of ≤ 12 SNPs. Furthermore, we observed among the isolates of unidentified patients, which were not included in clustering analysis, one isolate from Distrito Federal (isolated in the year 2007) that had a genetic distance of ≤ 12 SNPs to the GC4 isolates that were from Rio de Janeiro (2008–2012), and one isolate from São Paulo (2004) and one isolate from Distrito Federal (2007) that had genetic a distance of 11 SNPs between them. These could point to other two interstate transmission events, requiring further investigation.

From 15 patients, we had several isolates collected during and after treatment. In most of these cases, the infecting lineage strain during follow-up with the patients was the same, and reinfection with another strain and lineage was only observed in one patient (P35). In a retrospective cohort study from China [42], a MIRU-VNTR-based recurrence definition observed a lower frequency of reinfection (n = 21; 36.2%) in comparison to relapse (n = 37; 63.8%). In a population-based study from Malawi [43] using WGS, a lower frequency of reinfection (n = 20; 14.3%) compared with relapse (n = 55; 39.5%) was also observed; 64 (46%) of recurrent patients remained unclassified, showing how challenging it remains to classify such cases.

We observed the emergence of FQ-resistance-conferring mutations among one third of the patients that were followed-up with during treatment. These bacterial populations had likely been selected due non-compliance of drug treatment [44,45]. In addition, the emergence of phenotypic resistance to AMK in patient P39 and phenotypic resistance to KAN and CAP in patient P64 was observed, but without the detection of known mutations in *rrs* and *eis*. This could be caused by not yet known mutations conferring resistance to those drugs and/or by efflux pump activity, not investigated here. Such pumps have been described to be induced by sub-inhibitory levels of antibiotics such as FQ, RIF, and INH in inappropriate treatment regimens [46]. Additionally, the presence of a minor undetected population of drug-resistant bacteria, so-called heteroresistance, can also explain differences between culture and in silico-based drug susceptibility outcomes [47].

Another observation was the emergence of the G406D mutation in *embB* in patient P54 (Table 3); however, a mixed call was detected in this codon and may indicate the selection of the mutant population within a single strain, probably due to inappropriate treatment. Patient P35 presented the emergence of the Q497R mutation in *embB*, and in addition a shift in the RIF resistance genotype, initially presenting the Q432P mutation in *rpoB* and

posteriorly the S450L mutation. Besides that, in the first isolate the sub-lineage detected was L4.1.1 (X) and in the second was L4.3 (LAM), and a genomic difference of 484 SNPs was observed between both, indicating a TB recurrence by reinfection [43].

We observed a statistically significantly lower frequency of mutations related to resistance to STR, CAP, ETH, and AMK in strains belonging to sub-lineage 4.3 (LAM) in comparison with non-LAM sub-lineages. A lower frequency of mutations related to resistance in some lineages has been described, including a lower frequency of S315T mutations in *katG* in the LAM spoligofamily when compared to Haarlem [48]. Another study showed a higher mutation rate in *Mtb* strains of lineage 2 and as such, a faster emergence of resistance-conferring mutations in strains of lineage 2, accompanied by the development of DR-TB [49].

Upon comparing patient treatment outcomes, risk factors for TB development, phenotypic and genotypic resistance, and *Mtb* lineage, we observed a significant association between EMB phenotypic (Chi-squared $p = 0.006$) and genotypic (Fisher's exact test $p = 0.0007$) resistance and TB mortality. EMB resistance has been described as a risk factor for mortality, mainly when associated with MDR pattern [50]. In our study, genotypic p-XDR-TB was significantly associated with EMB genotypic resistance, while the genotypic p-XDR-TB profile was significantly associated with a higher morbidity (Chi-squared $p = 0.03$). In addition, the presence of any of the five risk factors (HIV, diabetes, smoking cigarette, alcohol, and/or illicit drug use) was also significantly associated with morbidity but independent of genotypic or phenotypic EMB resistance. The increased risk of mortality in MDR-TB and p-XDR-TB patients has been described [51,52] and is related to the increasing level of drug resistance, whereas accumulating even more factors increases the probability of an unfavorable outcome [53,54]. EMB resistance could be an important stage of resistance accumulation in *Mtb*.

We observed a similar proportion of main drug-resistance-conferring mutations, frequencies, and affected genes detected by all the pipelines tested, and this is in agreement with other studies [55–59].

Unprecedented was the detection of 20 novel mutations in 32 strains using TB Profiler 5.0, including a novel frameshift mutation *katG_c.2070delC* in an isolate that is likely associated with INH resistance, considering the impact this may have on the catalase-peroxidase structure and activity, accompanied by phenotypic resistance against INH. The same isolate also presented the mutation *ahpC_c.-81C>T*, classified in the WHO catalogue as having uncertain significance, but which may be acting in synergy with the newly described mutation to confer resistance; this needs to be better investigated. This isolate was misclassified as genetically susceptible to INH by all pipelines except TB Profiler, and the same was observed in other isolates with the mutation *ahpC_c.-81C>T*, highlighting the importance of using updated pipelines.

Only 4 of 32 isolates were fully phenotypically characterized for the 9 drugs available in this study. Unfortunately, 13 of the isolates in which we detected a novel mutation were not phenotypically characterized for the drug of interest, making it impossible to establish the genetic and phenotypic correlation. Among the four isolates with the frameshift mutation *pncA_c.193_200dupTCCTCGTC*, two had no PZA DST available and two were PZA-resistant (these two were from the same patient); the mutation *pncA_c.305dupC* was observed in one resistant and one with a DST not available; finally, the mutations *pncA_c.452dupT* and *pncA_c.75_79delCGCGC* were both presents in resistant isolates, all suggesting a resistance association. On the other hand, the frameshift mutation *pncA_c.443_444dupGC* was detected in both susceptible and resistant isolates, making it difficult to interpret. Other phenotypically susceptible isolates observed with novel frameshift mutations in the *pncA* and *katG* genes could be explained by an acquired low-level resistance, under the antibiotic concentration threshold used here or due to some laboratorial error in phenotypic resistance determination. Phenotypically susceptible isolates harboring confident resistance mutations have been reported in other studies [60,61].

Among ten other mutations we detected that were absent in the WHO catalogue (Supplementary Table S5), eight [62–69] in twelve isolates had not yet been reported in Brazil, and four (*katG*_p.Leu634Phe, *pncA*_p.Asp63His, *pncA*_p.Gly23Val, and *pncA*_c.521_522insT) were present in phenotypically resistant isolates in our study. Another mutation associated with resistance was detected for the first time in Brazil, *rpoB*_p.Ile491Phe, a mutation outside of the 81 bp hotspot of the *rpoB* gene, which makes it undetectable by commercial assays such as GeneXpert MTB/RIF and MTBDRplus [70,71]. In Eswatini, for example, the presence of this mutation is a significant problem because of its prevalence of >60% in MDR strains [72]. In our study, this mutation was observed in one isolate that was genotypically and phenotypically MDR and carried the *rpoB*_p.Ser493Leu mutation, an uncertain variant according to the WHO. However, their combination could be acting in synergy because the former is a borderline resistance mutation [73].

The comparison of the pipelines, after evaluating their performance in drug resistance prediction through sensitivity, specificity, accuracy, and positive and negative predictive value calculation, using phenotypic resistance results as a reference, showed a better overall performance in resistance prediction by the in-house WGS pipeline. For all pipelines, the sensitivity of the prediction of resistance to RIF and INH was higher than 80% but the specificity was lower than 95%, and therefore below that of the WHO recommendations [74]. A high positive predictive value (PPV) of >80% was observed in all pipelines used for the detection of mutations conferring resistance to RIF, INH, OFL, AMK, and similar to that described by others [20,75]. A lower PPV was observed for the detection of resistance mutations whose action mechanisms are more complex and for which the genetic bases of resistance are less understood, such as resistance to PZA, KAN, and CAP. Unreliable PZA resistance prediction has been reported in various studies [20,31,75,76], and in order to enhance PZA resistance prediction, the following strategies have been described: (i) improving the mutation library of the pipelines mainly by including the detection of indels in *pncA*; (ii) the “non-wild-type sequence” approach, which consists of interpreting any non-synonymous mutations or indels in *pncA* as genotypic PZA-resistant strains; (iii) manually checking the sequence reads [77,78]. Not surprisingly, the best sensitivity performance for PZA resistance prediction was achieved by TB Profiler 5.0, the most up-to-date pipeline for the detection of DR-associated mutations, accounting for the largest PZA mutation catalogue among all pipelines used and able to detect novel mutations, including fourteen novel indels observed in this study.

Another important finding in our study is that, compared to conventional DST, more than twice as many cases were classified as p-XDR-TB using WGS. This difference in proportion could be beyond the accuracy of genomic tests for second-line resistance detection. However, one should take into account that the criteria for performing phenotypic DST tests for second-line drugs in Brazil cause the underestimation of the detection of resistance to such drugs. These results highlight the importance of using WGS for the epidemiological surveillance and control of DR/MDR/p-XDR-TB, as these discrepancies would not be detected and reported without it use.

Without any doubt, a major limitation of this study is that phenotypic tests for second-line drugs were not conducted for all samples. Among the 161 isolates classified as phenotypically MDR-TB, only 22 were tested for all second-line drugs, and 10 were only tested for 9 drugs; 1 of the 10 Poly-R-TB and 6 of the 27 p-XDR-TB were submitted to the full DST. This further emphasizes the importance of using molecular methods for TB diagnosis and comprehensive resistance analyses. Another limitation is the small sample size from regions outside Rio de Janeiro, Distrito Federal, and Sao Paulo States, not representing the national scenario of DR-TB in Brazil.

In conclusion, the evaluated pipelines for the prediction of MTBC lineage and drug resistance work well in the Brazilian sample studied here, and our data favor the use of WGS in cases without a conventional DST. Although the in-house WGS pipeline performed slightly better in general, all tools performed well in predicting DR to RIF, INH, AMK, KAN, and CAP. Importantly they allowed for the detection of p-XDR-TB strains that otherwise

would probably have been unreported. The analysis of isolates of DR-TB patients that were sampled sometimes years apart demonstrated that several accumulated drug resistance mutations, showing that resistance evolution occurs by the acquisition of mutations in the same strain, probably due to the non-compliance of treatment. Phylogenetic analysis and lineage characterization contribute to better understanding the MTBC genotypes spreading in the population, so WGS improves the knowledge of MTBC population structure and evolution and offers the rapid and reliable assessment of resistance-related mutations, allowing for faster access to effective treatment. A surprisingly low level of reinfection was observed in an area with high TB incidence, even in patients with DR and prolonged treatment. We believe that these findings highlight the importance of the need for active surveillance throughout the national territory in order to avoid further aggravation of the TB scenario in Brazil.

4. Materials and Methods

4.1. Sample Collection and Phenotypic Drug Resistance

We used a convenience sample for this study of 298 Mtb clinical isolates (Brazilian SISGEN code: AD20DC4), available at the National Tuberculosis Reference Center Professor Hélio Fraga (CRPHF; Rio de Janeiro, Brazil). The sample set included isolates of patients from 16 different Brazilian states with presumed DR-TB due to a history of treatment failure, treatment abandonment, relapse TB, or contact with DR-TB patients. The isolates were therefore submitted to phenotypic drug susceptibility testing (DST).

DST was performed for the first-line drugs RIF, INH, PZA, and EMB, and to the second-line drugs STR, OFL, AMK, KAN, and CAP, using the liquid MB/BacT system (Organon Teknika Corp., Durham, NC, USA). The following cut-off values were used: RIF (1.0 mg/L), INH (0.1 mg/L), EMB (5.0 mg/L), PZA (100.0 mg/L), STR (1.0 mg/L), OFL (2.0 mg/L), AMK (1.0 mg/L), KAN (4.0 mg/L), and CAP (2.5 mg/L).

4.2. Whole-Genome Sequencing and In-House Pipeline

The Mtb genomic DNA were extracted as described previously [79]. To perform WGS, sequencing libraries were prepared as described previously [80], and sequenced on an Illumina platform at Division of Infectious Diseases and Environmental Health of the Norwegian Institute of Public Health. The sequence reads generated were deposited in the Sequence Read Archive (SRA) of the NCBI under the Bioproject PRJEB27366. To describe the MTBC genetic diversity and genotypic resistance profile, a single-nucleotide-polymorphism (SNP)-based Mtb lineage and sub-lineage and resistance-related genes (11 coding genes and four promoter regions) were assigned, first using an in-house pipeline developed at the Swiss Tropical and Public Health Institute (Swiss TPH) as described before [81], hereafter called the in-house WGS pipeline. In brief, FASTQ files were processed with Trimmomatic v 0.33 (SLIDINGWINDOW:5:20) [82] to remove Illumina adaptors and trim low quality reads. Overlapping reads were then merged with SeqPrep v 1.2. Duplicated reads were marked by the MarkDuplicates module of Picard v 2.9.1. The resulting reads were mapped to a reconstructed ancestral sequence of MTBC as described previously [83] using the BWA-MEM v 0.7.13 algorithm. Pysam v 0.9.0 was used to exclude reads with an alignment score lower than $(0.93 * \text{read_length}) - (\text{read_length} * 4 * 0.07)$; this corresponds to more than seven mismatches per 100 bp. SNPs were called with Samtools v 1.2 mpileup and VarScan v 2.4.1. The SNP annotation used was that of *Mycobacterium tuberculosis* H37Rv reference strain (NC_000962.3). Repetitive regions of the genome as PE/PPE/PGRS were excluded as described before [84].

4.3. Phylogenetic Analysis

A maximum likelihood phylogenetic tree using 100 bootstraps was constructed using an alignment containing all variable positions from all quality filtered genomes with $\geq 15 \times$ average coverage using MEGAX v 10.2 [85], and the resulting tree was rooted using

M. canettii (Genbank accession number: NC_019950.1). The tree was visualized using iTOL v 6.7 [86].

4.4. Lineage Classification Pipelines Comparisons

Mtb lineage classification by the in-house WGS pipeline was compared with three online available pipelines: TB Profiler version 5.0 [21], KvarQ version 0.12.2 [10] but using the SNP barcode used elsewhere [11], and RD-Analyzer version 1.01, a region-of-difference-based analyzer of Mtb from sequence reads [87].

4.5. Genotypic Resistance Detection and Pipelines Evaluation

In order to identify the drug-resistance-related variants among the included Mtb genomes, we used the in-house WGS pipeline and the command-line versions of the pipelines Mykrobe Predictor version 0.3.3 [12], TB Profiler version 0.3.4 [76], KvarQ version 0.12.2 [10], and TB Profiler version 5.0 (<https://github.com/jodyphelan/TBProfiler>, accessed on 28 September 2023), a recent pipeline version that accounts for the updated WHO catalogue of mutations conferring resistance (<https://www.who.int/publications/item/9789240082410>, accessed on 29 January 2024).

4.6. Cluster Analysis and Recent Transmission

Genomic clusters were delineated using the tool *snp-dists* v0.7.0 (<https://github.com/tseemann/snp-dists>, accessed on 5 August 2022). An SNP-based distance matrix was built using the alignment of 170 high-quality genomes not presenting mixed calls. A genomic cluster was defined when two or more isolates presented genomes with ≤ 12 SNPs of difference, which encompasses not only linked cases such as household contacts (genetic distance from zero to five SNPs), but also related cases (five to twelve SNPs). These thresholds are considered as very likely representing recent transmission events, especially in settings using isolates from chronically infected patients and DR-TB [88–90]. To estimate recent transmission, 170 genomes, corresponding to one genome per patient appropriately identified (patient ID), were used. Samples without patient identification and/or duplicated ID were excluded from this analysis. The reconstructed maximum likelihood tree using 100 bootstraps was visualized in GrapeTree [91]. Clustering rate was calculated using the following formula $(nc - c)/n$, where nc = the total isolates in cluster, c = the number of clusters, and n = the total number of isolates.

4.7. Patient Data

The available patient clinical data, including TB treatment outcome, diabetes, alcohol use, illicit drug use, cigarette smoking, HIV coinfection status, and sex, were obtained from the “Sistema de Informação de Agravos de Notificação” (SINAN) and “Sistema de Informação de Tratamentos Especiais de Tuberculose” (SITETB), two national databases for disease surveillance. The project was approved by the Research Ethics Committee of Federal University of Rio de Janeiro (CAEE 10126919.2.0000.5257).

4.8. Statistical Analyses

For the prediction of drug resistance, we calculated sensitivity, specificity, accuracy, and positive and negative predictive values using the statistical software R (Version 3.6.1). First- and second-line anti-TB drug predictions were evaluated for the four pipelines, using MB/BacT DST as the reference standard.

Free marginal Kappa Randolph’s statistics [92] with a 95% confidence interval was used to determine agreement among Mtb lineage classification by the four pipelines. The same procedure was followed to determine agreement among phenotypic drug resistance detected by MB/BacT and the genotypic drug resistance detected by the four pipelines, using four levels of agreement: <0.40 (poor), 0.40 – 0.59 (fair), 0.60 – 0.80 (moderate/good) and >0.80 (excellent).

Pearson's Chi-squared test was used to determine associations among treatment outcome, risk factors, clusters, lineage, genotypic and phenotypic drug resistance using the `chisq.test` function in the statistical software R (Version 3.6.1). In case the expected value is less than 5, Fisher's exact test was calculated using the `fisher.test` function in R.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/antibiotics13060496/s1>. Figure S1: Lineage classification of 298 genome isolates using four pipelines. Figure S2: Venn diagram for the four pipelines used in lineage classification of the 298 genome isolates. Table S1: Phenotypic profile of 294 isolates included in the study. Table S2: Clinical data of 112 patients obtained from SITETB and SINAN databases. Table S3: Resistance-related mutations detected by the in-house WGS pipeline in 298 isolates. Table S4: Agreement among phenotypic DST, in-house WGS pipeline, KvarQ, Mykrobe Predictor, and TB Profiler versions 0.3.4 and 5.0 according to free marginal kappa coefficient. Table S5: Resistance-related mutations detected by TB Profiler version 5.0 in 298 isolates. Table S6: Chronic patients without changes in genotypic profile mutation resistance. Table S7: Genotypic and phenotypic profiles of the 82 patients with cured and deceased treatment outcomes.

Author Contributions: L.S.E., P.N.S., and A.L.K. contributed to conceptualization; L.S.E., L.L.G., D.B., F.C.O.F., M.B., M.A.d.S.P., E.C.C., R.S., B.P.d.C., R.F.d.M.M., P.C.d.S.C., P.R., M.P.D., V.E., S.G., M.L.R., A.L.K. and P.N.S. contributed to methodology; L.S.E., D.B., P.N.S., A.L.K., V.E., and S.G. contributed to formal analysis; L.S.E., L.L.G., D.B., F.C.O.F., M.B., M.A.d.S.P., E.C.C., R.S., B.P.d.C., R.F.d.M.M., P.C.d.S.C., P.R., M.P.D., V.E., S.G., M.L.R., A.L.K. and P.N.S. contributed to investigation; A.L.K., P.N.S., M.L.R., V.E., and S.G. contributed to resources; L.S.E., P.N.S., A.L.K., L.L.G., M.B., and F.C.O.F. contributed to data curation; L.S.E. contributed to writing—original draft preparation; L.S.E., D.B., R.S., P.N.S., A.L.K., M.L.R., V.E., and S.G. contributed to writing—review and editing; P.N.S., A.L.K., M.L.R., and S.G. contributed to supervision; L.S.E. contributed to project administration; P.N.S. and A.L.K. contributed to funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Coordination for the Improvement of Higher Education Personnel (CAPES/MEC/Brazil) and to the National Council for Scientific and Technological Development (CNPq/MCTI/Brazil). Leonardo Souza Esteves was supported by doctorate and sandwich doctorate scholarship 888881.134437/2016-01 (CAPES/Brazil-UFRJ/PPG Clínica Médica, Coordinator: José R. Lapa e Silva), and postdoctoral fellowship by Programa Fiocruz de Fomento à Inovação (INOVA/Fiocruz).

Institutional Review Board Statement: The project was approved by Research Ethics Committee of Federal University of Rio de Janeiro (CAEE 10126919.2.0000.5257).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data are contained within the article.

Acknowledgments: The authors would like to thank to sciCORE from University of Basel for helping in computational processing, the staff from Centro de Referência Professor Hélio Fraga (CRPHF/Brazil/RJ) for helping in samples processing and culture management, Laboratory of Molecular Biology Applied to Mycobacteria at IOC/FIOCRUZ and Laboratory of Infection Biology and Molecular Epidemiology at Swiss TPH and Norwegian Institute of Public Health for the infrastructure.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. WHO. *Global Tuberculosis Report 2023*; WHO: Geneva, Switzerland, 2023; ISBN 9789240083851.
2. Brazil. *Epidemiological Report—Tuberculosis 2021*; Secretaria de Vigilância em Saúde; Ministério da Saúde: Brasília, Brazil, 2021.
3. WHO. *Global Tuberculosis Report 2021*; WHO: Geneva, Switzerland, 2021.
4. WHO. *Global Tuberculosis Report 2019*; WHO: Geneva, Switzerland, 2019.
5. Bhering, M.; Kritski, A. Short Communication Trends in Primary Multidrug-Resistant Tuberculosis in the State of Rio de Janeiro: A Retrospective Study Conducted during 2000–2019. *Rev. Soc. Bras. Med. Trop.* **2021**, *54*, 12–15. [[CrossRef](#)] [[PubMed](#)]
6. Tagliani, E.; Anthony, R.; Kohl, T.A.; De Neeling, A.; Nikolayevskyy, V.; Ködmön, C.; Maurer, F.P.; Niemann, S.; Van Soolingen, D.; Van Der Werf, M.J.; et al. Use of a Whole Genome Sequencingbased Approach for Mycobacterium Tuberculosis Surveillance in Europe in 2017–2019: An ECDC Pilot Study. *Eur. Respir. J.* **2021**, *5*, 2002272. [[CrossRef](#)] [[PubMed](#)]

7. Walker, T.M.; Choisy, M.; Dedicoat, M.; Drennan, P.G.; Wyllie, D.; Yang-Turner, F.; Crook, D.W.; Robinson, E.R.; Walker, A.S.; Smith, E.G.; et al. Mycobacterium Tuberculosis Transmission in Birmingham, UK, 2009–2019: An Observational Study. *Lancet Reg. Health Eur.* **2022**, *17*, 100361. [[CrossRef](#)]
8. Meehan, C.J.; Goig, G.A.; Kohl, T.A.; Verboven, L.; Dippenaar, A.; Ezewudo, M.; Farhat, M.R.; Guthrie, J.L.; Laukens, K.; Miotto, P.; et al. Whole Genome Sequencing of Mycobacterium Tuberculosis: Current Standards and Open Issues. *Nat. Rev. Microbiol.* **2019**, *17*, 533–545. [[CrossRef](#)] [[PubMed](#)]
9. Papaventsis, D.; Casali, N.; Kontsevaya, I.; Drobniowski, F.; Cirillo, D.M.; Nikolayevskyy, V. Whole Genome Sequencing of Mycobacterium Tuberculosis for Detection of Drug Resistance: A Systematic Review. *Clin. Microbiol. Infect.* **2017**, *23*, 61–68. [[CrossRef](#)] [[PubMed](#)]
10. Steiner, A.; Stucki, D.; Coscolla, M.; Borrell, S.; Gagneux, S. KvarQ: Targeted and Direct Variant Calling from Fastq Reads of Bacterial Genomes. *BMC Genom.* **2014**, *15*, 881. [[CrossRef](#)] [[PubMed](#)]
11. Coll, F.; McNerney, R.; Guerra-Assunção, J.A.; Glynn, J.R.; Perdigão, J.; Viveiros, M.; Portugal, I.; Pain, A.; Martin, N.; Clark, T.G. A Robust SNP Barcode for Typing Mycobacterium Tuberculosis Complex Strains. *Nat. Commun.* **2014**, *5*, 4–8. [[CrossRef](#)] [[PubMed](#)]
12. Bradley, P.; Gordon, N.C.; Walker, T.M.; Dunn, L.; Heys, S.; Huang, B.; Earle, S.; Pankhurst, L.J.; Anson, L.; De Cesare, M.; et al. Rapid Antibiotic-Resistance Predictions from Genome Sequence Data for Staphylococcus Aureus and Mycobacterium Tuberculosis. *Nat. Commun.* **2015**, *6*, 10063. [[CrossRef](#)] [[PubMed](#)]
13. Verboven, L.; Phelan, J.; Heupink, T.H.; Van Rie, A. TBProfiler for Automated Calling of the Association with Drug Resistance of Variants in Mycobacterium Tuberculosis. *PLoS ONE* **2022**, *17*, e0279644. [[CrossRef](#)]
14. Schloss, P.D.; Westcott, S.L.; Ryabin, T.; Hall, J.R.; Hartmann, M.; Hollister, E.B.; Lesniewski, R.A.; Oakley, B.B.; Parks, D.H.; Robinson, C.J.; et al. Freedom in Bioinformatics. *Front. Genet.* **2014**, *5*, 259. [[CrossRef](#)] [[PubMed](#)]
15. Smith, D.R. Buying in to Bioinformatics: An Introduction to Commercial Sequence Analysis Software. *Brief. Bioinform.* **2014**, *16*, 700–709. [[CrossRef](#)] [[PubMed](#)]
16. Oppong, Y.E.A.; Phelan, J.; Perdigão, J.; Machado, D.; Miranda, A.; Viveiros, M.; Clark, T.G.; Hibberd, M.L. Genome-Wide Analysis of Mycobacterium Tuberculosis Polymorphisms Reveals Lineage-Specific Associations with Drug Resistance. *BMC Genom.* **2019**, *20*, 252. [[CrossRef](#)] [[PubMed](#)]
17. Joung, B.C.; Hill, P.C.; Aiken, A.; Awine, T.; Antonio, M.; Adetifa, I.M.; Jackson-Sillah, D.J.; Fox, A.; Kathryn, D.; Gagneux, S.; et al. Progression to Active Tuberculosis, but Not Transmission, Varies by M. Tuberculosis Lineage in The Gambia. *J. Infect. Dis.* **2008**, *198*, 1037–1043. [[CrossRef](#)]
18. Portevin, D.; Gagneux, S.; Comas, I.; Young, D. Human Macrophage Responses to Clinical Isolates from the Mycobacterium Tuberculosis Complex Discriminate between Ancient and Modern Lineages. *PLoS Pathog.* **2011**, *7*, e1001307. [[CrossRef](#)]
19. Coscolla, M.; Gagneux, S. Consequences of Genomic Diversity in Mycobacterium Tuberculosis. *Semin. Immunol.* **2014**, *26*, 431–444. [[CrossRef](#)] [[PubMed](#)]
20. Beek, J.V.; Haanper, M.; Smit, P.W.; Mentula, S.; Soini, H. Evaluation of Whole Genome Sequencing and Software Tools for Drug Susceptibility Testing of Mycobacterium Tuberculosis. *Clin. Microbiol. Infect.* **2019**, *25*, 82–86. [[CrossRef](#)] [[PubMed](#)]
21. Phelan, J.E.; O’Sullivan, D.M.; Machado, D.; Ramos, J.; Oppong, Y.E.A.; Campino, S.; O’Grady, J.; McNerney, R.; Hibberd, M.L.; Viveiros, M.; et al. Integrating Informatics Tools and Portable Sequencing Technology for Rapid Detection of Resistance to Anti-Tuberculous Drugs. *Genome Med.* **2019**, *11*, 41. [[CrossRef](#)] [[PubMed](#)]
22. Walker, T.M.; Lalor, M.K.; Broda, A.; Ortega, L.S.; Morgan, M.; Parker, L.; Churchill, S.; Bennett, K.; Golubchik, T.; Giess, A.P.; et al. Assessment of Mycobacterium Tuberculosis Transmission in Oxfordshire, UK, 2007–2012, with Whole Pathogen Genome Sequences: An Observational Study. *Lancet Respir. Med.* **2014**, *2*, 285–292. [[CrossRef](#)]
23. Menardo, F.; Gagneux, S.; Rutaihw, L.K.; Zwyer, M.; Borrell, S.; Comas, I.; Conceição, E.C.; Coscolla, M.; Cox, H.; Joloba, M.; et al. Local Adaptation in Populations of Mycobacterium Tuberculosis Endemic to the Indian Ocean Rim. *F1000Research* **2021**, *10*, 60. [[CrossRef](#)]
24. Gagneux, S.; DeRiemer, K.; Van, T.; Kato-Maeda, M.; De Jong, B.C.; Narayanan, S.; Nicol, M.; Niemann, S.; Kremeri, K.; Gutierrez, M.C.; et al. Variable Host-Pathogen Compatibility in Mycobacterium Tuberculosis. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 2869–2873. [[CrossRef](#)] [[PubMed](#)]
25. Fenner, L.; Egger, M.; Bodmer, T.; Furrer, H.; Ballif, M.; Battegay, M.; Helbling, P.; Fehr, J.; Gsponer, T.; Rieder, H.L.; et al. HIV Infection Disrupts the Sympatric Host-Pathogen Relationship in Human Tuberculosis. *PLoS Genet.* **2013**, *9*, e1003318. [[CrossRef](#)] [[PubMed](#)]
26. Salvato, R.S.; Schiefelbein, S.; Barcellos, R.B.; Praetzel, B.M.; Anusca, I.S.; Esteves, L.S.; Halon, M.L.; Unis, G.; Dias, C.F.; Miranda, S.S.; et al. Molecular Characterisation of Multidrug-Resistant Mycobacterium Tuberculosis Isolates from a High-Burden Tuberculosis State in Brazil. *Epidemiol. Infect.* **2019**, *147*, e216. [[CrossRef](#)] [[PubMed](#)]
27. Esteves, L.S.; Dalla Costa, E.R.; Vasconcellos, S.E.G.; Vargas, A.; Ferreira Junior, S.L.M.; Halon, M.L.; Ribeiro, M.O.; Rodenbusch, R.; Gomes, H.M.; Suffys, P.N.; et al. Genetic Diversity of Mycobacterium Tuberculosis Isoniazid Mono-resistant and Multidrug-Resistant in Rio Grande Do Sul, a Tuberculosis High-Burden State in Brazil. *Tuberculosis* **2018**, *110*, 36–43. [[CrossRef](#)] [[PubMed](#)]
28. Dantas, N.G.T.; Suffys, P.N.; da Silva Carvalho, W.; Gomes, H.M.; de Almeida, I.N.; de Assis, L.J.; Augusto, C.J.; Gomgnimbou, M.K.; Refregier, G.; Sola, C.; et al. Genetic Diversity and Molecular Epidemiology of Multidrug-Resistant Mycobacterium Tuberculosis in Minas Gerais State, Brazil. *BMC Infect. Dis.* **2015**, *15*, 306. [[CrossRef](#)] [[PubMed](#)]

29. Conceição, E.C.; Rastogi, N.; Couvin, D.; Lopes, M.L.; Furlaneto, I.P.; Gomes, H.M.; Vasconcellos, S.E.G.; Suffys, P.N.; Schneider, M.P.C.; de Sousa, M.S.; et al. Genetic Diversity of Mycobacterium Tuberculosis from Pará, Brazil, Reveals a Higher Frequency of Ancestral Strains than Previously Reported in South America. *Infect. Genet. Evol.* **2017**, *56*, 62–72. [[CrossRef](#)] [[PubMed](#)]
30. Baker, L.; Brown, T.; Maiden, M.C.; Drobniowski, F. Silent Nucleotide Polymorphisms and a Phylogeny for Mycobacterium Tuberculosis. *Emerg. Infect. Dis.* **2004**, *10*, 1568–1577. [[CrossRef](#)] [[PubMed](#)]
31. Guimarães, A.E.D.S.; Sharma, A.; Furlaneto, I.P.; Rutaihwa, L.; Cardoso, J.F.; da Conceição, M.L.; Spinassé, L.B.; Machado, E.; Lopes, M.L.; Duarte, R.S.; et al. Evaluation of Drug Susceptibility Profile of Mycobacterium Tuberculosis Lineage 1 from Brazil Based on Whole Genome Sequencing and Phenotypic Methods. *Mem. Inst. Oswaldo Cruz.* **2020**, *115*, e200520. [[CrossRef](#)] [[PubMed](#)]
32. Conceição, E.C.; Refregier, G.; Gomes, H.M.; Olessa-Daragon, X.; Coll, F.; Ratovonirina, N.H.; Rasolofo-Razanamparany, V.; Lopes, M.L.; van Soolingen, D.; Rutaihwa, L.; et al. Mycobacterium Tuberculosis Lineage 1 Genetic Diversity in Pará, Brazil, Suggests Common Ancestry with East-African Isolates Potentially Linked to Historical Slave Trade. *Infect. Genet. Evol.* **2019**, *73*, 337–341. [[CrossRef](#)] [[PubMed](#)]
33. Mallard, K.; McNerney, R.; Crampin, A.C.; Houben, R.; Ndlovu, R.; Munthali, L.; Warren, R.M.; French, N.; Glynn, J.R. Molecular Detection of Mixed Infections of Mycobacterium Tuberculosis Strains in Sputum Samples from Patients in Karonga District, Malawi. *J. Clin. Microbiol.* **2010**, *48*, 4512–4518. [[CrossRef](#)] [[PubMed](#)]
34. Sobkowiak, B.; Glynn, J.R.; Houben, R.M.G.J.; Mallard, K.; Phelan, J.E.; Guerra-assunção, J.A.; Banda, L.; Mzembe, T.; Viveiros, M.; McNerney, R.; et al. Identifying Mixed Mycobacterium Tuberculosis Infections from Whole Genome Sequence Data. *BMC Genom.* **2018**, *19*, 613. [[CrossRef](#)] [[PubMed](#)]
35. Eyre, D.W.; Cule, M.L.; Griffiths, D.; Crook, D.W.; Peto, T.E.A.; Walker, A.S.; Wilson, D.J. Detection of Mixed Infection from Bacterial Whole Genome Sequence Data Allows Assessment of Its Role in Clostridium Difficile Transmission. *PLoS Comput. Biol.* **2013**, *9*, e1003059. [[CrossRef](#)] [[PubMed](#)]
36. Perdígão, J.; Gomes, P.; Miranda, A.; Maltez, F.; Machado, D.; Silva, C.; Phelan, J.E.; Brum, L.; Campino, S.; Couto, I.; et al. Using Genomics to Understand the Origin and Dispersion of Multidrug and Extensively Drug Resistant Tuberculosis in Portugal. *Sci. Rep.* **2020**, *10*, 2600. [[CrossRef](#)] [[PubMed](#)]
37. Wyllie, D.H.; Davidson, J.A.; Grace Smith, E.; Rathod, P.; Crook, D.W.; Peto, T.E.A.; Robinson, E.; Walker, T.; Campbell, C. A Quantitative Evaluation of MIRU-VNTR Typing Against Whole-Genome Sequencing for Identifying Mycobacterium Tuberculosis Transmission: A Prospective Observational Cohort Study. *EBioMedicine* **2018**, *34*, 122–130. [[CrossRef](#)] [[PubMed](#)]
38. Ferrazoli, L.; Palaci, M.; Marques, L.R.M.; Jamal, L.F.; Afiune, J.B.; Chimara, E.; Martins, M.C.; Da Silva Telles, M.A.; Oliveira, C.A.F.; Palhares, M.C.; et al. Transmission of Tuberculosis in an Endemic Urban Setting in Brazil. *Int. J. Tuberc. Lung Dis.* **2000**, *4*, 18–25. [[PubMed](#)]
39. Telles, M.A.S.; Ferrazoli, L.; Waldman, E.A.; Giampaglia, C.M.S.; Martins, M.C.; Ueki, S.Y.M.; Chimara, E.; Silva, C.A.; Cruz, V.; Waldman, C.C.S.; et al. A Population-Based Study of Drug Resistance and Transmission of Tuberculosis in an Urban Community. *Int. J. Tuberc. Lung Dis.* **2005**, *9*, 970–976. [[PubMed](#)]
40. Nogueira, C.L.; Prim, R.I.; Senna, S.G.; Rovaris, D.B.; Maurici, R.; Rossetti, M.L.; Couvin, D.; Rastogi, N.; Bazzo, M.L. First Insight into the Molecular Epidemiology of Mycobacterium Tuberculosis in Santa Catarina, Southern Brazil. *Tuberculosis* **2016**, *97*, 57–64. [[CrossRef](#)] [[PubMed](#)]
41. Salvato, R.; Regina, E.; Costa, D.; Júlia, A.; Hee, S.; Laura, M.; Bones, R.; Unis, G.; Fontoura, C.; Viveiros, M.; et al. Infection, Genetics and Evolution First Insights into Circulating XDR and Pre-XDR Mycobacterium Tuberculosis in Southern Brazil. *Infect. Genet. Evol.* **2020**, *78*, 104127. [[CrossRef](#)] [[PubMed](#)]
42. Zong, Z.; Huo, F.; Shi, J.; Jing, W.; Ma, Y.; Liang, Q.; Jiang, G.; Dai, G.; Huang, H.; Pang, Y. Relapse versus Reinfection of Recurrent Tuberculosis Patients in a National Tuberculosis Specialized Hospital in Beijing, China. *Front. Microbiol.* **2018**, *9*, 1858. [[CrossRef](#)] [[PubMed](#)]
43. Guerra-Assunção, J.A.; Houben, R.M.G.J.; Crampin, A.C.; Mzembe, T.; Mallard, K.; Coll, F.; Khan, P.; Banda, L.; Chiwaya, A.; Pereira, R.P.A.; et al. Recurrence Due to Relapse or Reinfection with Mycobacterium Tuberculosis: A Whole-Genome Sequencing Approach in a Large, Population-Based Cohort with a High HIV Infection Prevalence and Active Follow-Up. *J. Infect. Dis.* **2015**, *211*, 1154–1163. [[CrossRef](#)] [[PubMed](#)]
44. Hughes, D.; Andersson, D.I. Evolutionary Trajectories to Antibiotic Resistance. *Annu. Rev. Microbiol.* **2017**, *71*, 579–596. [[CrossRef](#)] [[PubMed](#)]
45. Anley, D.T.; Akalu, T.Y.; Dessie, A.M.; Anteneh, R.M.; Zemene, M.A.; Bayih, W.A.; Solomon, Y.; Gebeyehu, N.A.; Kassie, G.A.; Mengstie, M.A.; et al. Prognostication of Treatment Non-Compliance among Patients with Multidrug-Resistant Tuberculosis in the Course of Their Follow-up: A Logistic Regression-Based Machine Learning Algorithm. *Front. Digit. Health* **2023**, *5*, 1165222. [[CrossRef](#)] [[PubMed](#)]
46. Kanji, A.; Hasan, R.; Zhang, Y.; Shi, W.; Imtiaz, K.; Iqbal, K.; Shafiq, S.; Hasan, Z. Increased Expression of Efflux Pump Genes in Extensively Drug-Resistant Isolates of Mycobacterium Tuberculosis. *Int. J. Mycobacteriol.* **2016**, *5*, S150. [[CrossRef](#)] [[PubMed](#)]
47. Rigouts, L.; Miotto, P.; Schats, M.; Lempens, P.; Cabibbe, A.M.; Galbiati, S.; Lampasona, V.; de Rijk, P.; Cirillo, D.M.; de Jong, B.C. Fluoroquinolone Heteroresistance in Mycobacterium Tuberculosis: Detection by Genotypic and Phenotypic Assays in Experimentally Mixed Populations. *Sci. Rep.* **2019**, *9*, 11760. [[CrossRef](#)] [[PubMed](#)]

48. Dalla Costa, E.R.; Ribeiro, M.O.; Silva, M.S.; Arnold, L.S.; Rostirolla, D.C.; Cafrune, P.I.; Espinoza, R.C.; Palaci, M.; Telles, M.A.; Ritacco, V.; et al. Correlations of Mutations in KatG, OxyR-AhpC and InhA Genes and in Vitro Susceptibility in Mycobacterium Tuberculosis Clinical Strains Segregated by Spoligotype Families from Tuberculosis Prevalent Countries in South America. *BMC Microbiol.* **2009**, *9*, 39. [[CrossRef](#)] [[PubMed](#)]
49. Ford, C.B.; Shah, R.R.; Maeda, M.K.; Gagneux, S.; Murray, M.B.; Cohen, T.; Johnston, J.C.; Gardy, J.; Lipsitch, M.; Fortune, S.M. Mycobacterium Tuberculosis Mutation Rate Estimates from Different Lineages Predict Substantial Differences in the Emergence of Drug-Resistant Tuberculosis. *Nat. Genet.* **2013**, *45*, 784–790. [[CrossRef](#)] [[PubMed](#)]
50. Gayoso, R.; Dalcolmo, M.; Ueleres, J.; Barreira, D. Predictors of Mortality in Multidrug-Resistant Centers, 2005 to 2012. *Braz. J. Infect. Dis.* **2018**, *22*, 305–310. [[CrossRef](#)] [[PubMed](#)]
51. Zürcher, K.; Reichmuth, M.L.; Ballif, M.; Loiseau, C.; Borrell, S.; Reinhard, M.; Skrivankova, V.; Hömke, R.; Sander, P.; Avihingsanon, A.; et al. Articles Mortality from Drug-Resistant Tuberculosis in High-Burden Countries Comparing Routine Drug Susceptibility Testing with Whole-Genome Sequencing: A Multicentre Cohort Study. *Lancet Microbe* **2021**, *2*, 320–330. [[CrossRef](#)] [[PubMed](#)]
52. Araujo, L.G.; Garcia, M.T.; Zaccarioto, T.R.; MOrretti, M.L.; Levy, C.E.; Resende, M.R. Clinical Outcomes and Molecular Characterization of Drug-Resistant Tuberculosis in Pre- and Extensively Drug-Resistant Disease Based on Line Probe Assays. *Braz. J. Infect. Dis.* **2021**, *5*, 101544. [[CrossRef](#)] [[PubMed](#)]
53. Mitnick, C.D.; Franke, M.F.; Rich, M.L.; Viru, F.A.A.; Appleton, S.C.; Atwood, S.S.; Bayona, J.N.; Bonilla, C.A.; Chalco, K.; Fraser, H.S.F.; et al. Aggressive Regimens for Multidrug-Resistant Tuberculosis Decrease All-Cause Mortality. *PLoS ONE* **2013**, *8*, e58664. [[CrossRef](#)] [[PubMed](#)]
54. Peters, L.; Olson, L.; Khu, D.T.K.; Linnros, S.; Le, N.K.; Hanberger, H.; Hoang, N.T.B.; Tran, D.M.; Larsson, M. Multiple Antibiotic Resistance as a Risk Factor for Mortality and Prolonged Hospital Stay: A Cohort Study among Neonatal Intensive Care Patients with Hospital-Acquired Infections Caused by Gram-Negative Bacteria in Vietnam. *PLoS ONE* **2019**, *14*, e0215666. [[CrossRef](#)] [[PubMed](#)]
55. Murakami, K.S.; Scharf, N.; Stefan, M.; Garcia, G.; Murakami, K. Structural Basis for Rifamycin Resistance of Bacterial RNA Polymerase by the Three Most Clinically Important RpoB Mutations Found in Mycobacterium Tuberculosis. *Mol. Microbiol.* **2018**, *103*, 1034–1045. [[CrossRef](#)]
56. Matsui, T.; Maíra, J.; Pinhata, W.; Christiane, M.; Brandão, A.P.; Ferrazoli, L.; Leão, S.C.; Viana-niero, C. Frequency of First and Second-Line Drug Resistance-Associated Mutations among Resistant Mycobacterium Tuberculosis Clinical Isolates from São Paulo, Brazil. *Mem. Inst. Oswaldo Cruz.* **2020**, *115*, e200055. [[CrossRef](#)] [[PubMed](#)]
57. Daoqun, L.; Yuzhu, S.; Cheng-lin, Z.; Xiaofei, L.; Xueshan, X. Journal of Infection and Public Health Screening Mutations in Drug-Resistant Mycobacterium Tuberculosis Strains in Yunnan, China. *J. Infect. Public Health* **2017**, *10*, 630–636. [[CrossRef](#)] [[PubMed](#)]
58. Dokht, A.; Etemad, N.; Hashemzadeh, M. Journal of Global Antimicrobial Resistance Frequency of Rrs and RpsL Mutations in Streptomycin-Resistant Mycobacterium Tuberculosis Isolates from Iranian Patients. *Integr. Med. Res.* **2017**, *9*, 51–56. [[CrossRef](#)] [[PubMed](#)]
59. Shi, D.; Li, L.; Zhao, Y.; Jia, Q.; Li, H.; Coulter, C.; Jin, Q.; Zhu, G. Characteristics of EmbB Mutations in Multidrug-Resistant Mycobacterium Tuberculosis Isolates in Henan, China. *J. Antimicrob. Chemother.* **2011**, *66*, 2240–2247. [[CrossRef](#)] [[PubMed](#)]
60. Chen, Y.; Takiff, H.E.; Gao, Q. Phenotypic Instability of Mycobacterium Tuberculosis Strains Harboring Clinically Prevalent Drug-Resistant Mutations. *Lancet Microbe* **2023**, *4*, e292. [[CrossRef](#)] [[PubMed](#)]
61. Li, J.; Feng, S.; Chen, Y.; Yu, M.; Wei, W.; Tian, G.B. Antimicrobial Susceptibility Testing in Clinical Mycobacterium Tuberculosis Isolates. *Lancet Microbe* **2023**, *4*, e68. [[CrossRef](#)]
62. Wei, C.; Lei, B.; Musser, J.M.; Tu, S. Isoniazid Activation Defects in Recombinant Mycobacterium Tuberculosis Catalase-Peroxidase (KatG) Mutants Evident in InhA Inhibitor Production. *Antimicrob. Agents Chemother.* **2003**, *47*, 670–675. [[CrossRef](#)]
63. Cheng, S.J.; Thibert, L.; Sanchez, T.; Heifets, L.; Zhang, Y. PncA Mutations as a Major Mechanism of Pyrazinamide Resistance in Mycobacterium Tuberculosis: Spread of a Monoresistant Strain in Quebec, Canada. *Antimicrob. Agents Chemother.* **2000**, *44*, 528–532. [[CrossRef](#)] [[PubMed](#)]
64. Lee, K.W.; Lee, J.-M.; Jung, K.-S. Characterization of PncA Mutations of Pyrazinamide-Resistant Mycobacterium Tuberculosis in Korea. *J. Korean Med. Sci.* **2001**, *16*, 537–543. [[CrossRef](#)]
65. Li, K.; Yang, Z.; Gu, J.; Luo, M.; Deng, J.; Chen, Y. Characterization of PncA Mutations and Prediction of PZA Resistance in Mycobacterium Tuberculosis Clinical Isolates from Chongqing, China. *Front. Microbiol.* **2021**, *11*, 594171. [[CrossRef](#)] [[PubMed](#)]
66. Scorpio, A.; Lindholm-Levy, P.; Heifets, L.; Gilman, R.; Siddiqi, S.; Cynamon, M.; Zhang, Y. Characterization of PncA Mutations in Pyrazinamide-Resistant Mycobacterium Tuberculosis. *Antimicrob. Agents Chemother.* **1997**, *41*, 540–543. [[CrossRef](#)] [[PubMed](#)]
67. Marttila, H.J.; Marjamäki, M.; Vyshnevskaya, E.; Vyshnevskiy, B.I.; Otten, T.F.; Vasilyef, A.V.; Viljanen, M.K. PncA Mutations in Pyrazinamide-Resistant Mycobacterium Tuberculosis Isolates from Northwestern Russia. *Antimicrob. Agents Chemother.* **1999**, *43*, 1764–1766. [[CrossRef](#)] [[PubMed](#)]
68. Morlock, G.P.; Metchock, B.; Sikes, D.; Crawford, J.T.; Cooksey, R.C. EthA, InhA, and KatG Loci of Ethionamide-Resistant Clinical Mycobacterium Tuberculosis Isolates. *Antimicrob. Agents Chemother.* **2003**, *47*, 3799–3805. [[CrossRef](#)] [[PubMed](#)]

69. Al-Ghafli, H.; Kohl, T.A.; Merker, M.; Varghese, B.; Halees, A.; Niemann, S.; Al-Hajj, S. Drug-Resistance Profiling and Transmission Dynamics of Multidrug-Resistant Mycobacterium Tuberculosis in Saudi Arabia Revealed by Whole Genome Sequencing. *Infect. Drug Resist.* **2018**, *11*, 2219–2229. [[CrossRef](#)] [[PubMed](#)]
70. Sanches-Padilha, E.; Merker, M.; Beckert, P.; Jochims, F.; Dlamini, T.; Kahn, P.; Bonnet, M.; Niemann, S. Detection of Drug-Resistant Tuberculosis by Xpert MTB/RIF in Swaziland. *N. Engl. J. Med.* **2015**, *372*, 1179–1181. [[CrossRef](#)] [[PubMed](#)]
71. Makhado, N.A.; Matabane, E.; Faccin, M.; Pinçon, C.; Jouet, A.; Boutachkourt, F.; Goeminne, L.; Gaudin, C.; Maphalala, G.; Beckert, P.; et al. Outbreak of Multidrug-Resistant Tuberculosis in South Africa Undetected by WHO-Endorsed Commercial Tests: An Observational Study. *Lancet Infect. Dis.* **2018**, *18*, 1350–1359. [[CrossRef](#)] [[PubMed](#)]
72. Beckert, P.; Sanchez-Padilla, E.; Merker, M.; Dreyer, V.; Kohl, T.A.; Utpatel, C.; Köser, C.U.; Barilar, I.; Ismail, N.; Omar, S.V.; et al. MDR M. Tuberculosis Outbreak Clone in Eswatini Missed by Xpert Has Elevated Bedaquiline Resistance Dated to the Pre-Treatment Era. *Genome Med.* **2020**, *12*, 104. [[CrossRef](#)]
73. World Health Organization. *Technical Report on Critical Concentrations for Drug Susceptibility Testing of Isoniazid and the Rifamycins (Rifampicin, Rifabutin and Rifapentine)*; WHO: Geneva, Switzerland, 2021; ISBN 9789240017283.
74. Denking, C.M.; Kik, S.V.; Cirillo, M.; Casenghi, M.; Shinnick, T.; Weyer, K.; Gilpin, C.; Boehme, C.C.; Schito, M.; Kimerling, M.; et al. Defining the Needs for Next Generation Assays for Tuberculosis. *J. Infect. Dis.* **2015**, *211*, 29–38. [[CrossRef](#)] [[PubMed](#)]
75. Macedo, R.; Nunes, A.; Portugal, I.; Duarte, S.; Vieira, L.; Gomes, J.P. Dissecting Whole-Genome Sequencing-Based Online Tools for Predicting Resistance in Mycobacterium Tuberculosis: Can We Use Them for Clinical Decision Guidance? *Tuberculosis* **2018**, *110*, 44–51. [[CrossRef](#)] [[PubMed](#)]
76. Coll, F.; McNerney, R.; Preston, M.D.; Guerra-Assunção, J.A.; Warry, A.; Hill-Cawthorne, G.; Mallard, K.; Nair, M.; Miranda, A.; Alves, A.; et al. Rapid Determination of Anti-Tuberculosis Drug Resistance from Whole-Genome Sequences. *Genome Med.* **2015**, *7*, 51. [[CrossRef](#)] [[PubMed](#)]
77. Coll, F.; Phelan, J.; Hill-Cawthorne, G.A.; Nair, M.B.; Mallard, K.; Ali, S.; Abdallah, A.M.; Alghamdi, S.; Alsomali, M.; Ahmed, A.O.; et al. Genome-Wide Analysis of Multi- and Extensively Drug-Resistant Mycobacterium Tuberculosis. *Nat. Genet.* **2018**, *50*, 307–316. [[CrossRef](#)] [[PubMed](#)]
78. Iwamoto, T.; Murase, Y.; Yoshida, S.; Aono, A.; Kuroda, M.; Sekizuka, T.; Yamashita, A.; Kato, K.; Takii, T.; Arikawa, K.; et al. Overcoming the Pitfalls of Automatic Interpretation of Whole Genome Sequencing Data by Online Tools for the Prediction of Pyrazinamide Resistance in Mycobacterium Tuberculosis. *PLoS ONE* **2019**, *14*, e0212798. [[CrossRef](#)] [[PubMed](#)]
79. van Embden, J.D.A.; Cave, M.D.; Crawford, J.T.; Dale, J.W.; Eisenach, K.D.; Gicquel, B.; Hermans, P.; Martin, C.; Mcadam, R.; Shinnick, T.M.; et al. Strain Identification of Mycobacterium Tuberculosis by DNA Fingerprinting: Recommendations for a Standardized Methodology. *J. Clin. Microbiol.* **1993**, *31*, 406–409. [[CrossRef](#)] [[PubMed](#)]
80. Eldholm, V.; Pettersson, J.H.O.; Brynildsrud, O.B.; Kitchen, A.; Rasmussen, E.M.; Lillebaek, T.; Rønning, J.O.; Crudu, V.; Mengshoel, A.T.; Debech, N.; et al. Armed Conflict and Population Displacement as Drivers of the Evolution and Dispersal of Mycobacterium Tuberculosis. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 13881–13886. [[CrossRef](#)] [[PubMed](#)]
81. Menardo, F.; Loiseau, C.; Brites, D.; Coscolla, M.; Gygli, S.M.; Rutaihua, L.K.; Trauner, A.; Beisel, C.; Borrell, S.; Gagneux, S. Treemmer: A Tool to Reduce Large Phylogenetic Datasets with Minimal Loss of Diversity. *BMC Bioinform.* **2018**, *19*, 164. [[CrossRef](#)]
82. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [[CrossRef](#)]
83. Comas, I.; Chakravarti, J.; Small, P.M.; Galagan, J.; Niemann, S.; Kremer, K.; Ernst, J.D.; Gagneux, S. Human T Cell Epitopes of Mycobacterium Tuberculosis Are Evolutionarily Hyperconserved. *Nat. Genet.* **2010**, *42*, 498–503. [[CrossRef](#)]
84. Stucki, D.; Brites, D.; Jeljeli, L.; Coscolla, M.; Liu, Q.; Trauner, A.; Fenner, L.; Rutaihua, L.; Borrell, S.; Luo, T.; et al. Mycobacterium Tuberculosis Lineage 4 Comprises Globally Distributed and Geographically Restricted Sublineages. *Nat. Genet.* **2016**, *48*, 1535–1543. [[CrossRef](#)] [[PubMed](#)]
85. Kumar, S.; Stecher, G.; Li, M.; Nnyaz, C.; Tamura, K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol. Biol. Evol.* **2018**, *35*, 1547–1549. [[CrossRef](#)]
86. Letunic, I.; Bork, P. Interactive Tree of Life (ITOL) v4: Recent Updates and New Developments. *Nucleic Acids Res.* **2019**, *47*, 256–259. [[CrossRef](#)] [[PubMed](#)]
87. Faksri, K.; Xia, E.; Tan, J.H.; Teo, Y.Y.; Ong, R.T.H. In Silico Region of Difference (RD) Analysis of Mycobacterium Tuberculosis Complex from Sequence Reads Using RD-Analyzer. *BMC Genom.* **2016**, *17*, 847. [[CrossRef](#)] [[PubMed](#)]
88. Walker, T.M.; Ip, C.L.C.; Harrell, R.H.; Evans, J.T.; Kapatai, G.; Dedicoat, M.J.; Eyre, D.W.; Wilson, D.J.; Hawkey, P.M.; Crook, D.W.; et al. Whole-Genome Sequencing to Delineate Mycobacterium Tuberculosis Outbreaks: A Retrospective Observational Study. *Lancet Infect. Dis.* **2013**, *13*, 137–146. [[CrossRef](#)] [[PubMed](#)]
89. Comas, I. Genomic Epidemiology of Tuberculosis. *Adv. Exp. Med. Biol.* **2017**, *1019*, 79–93. [[CrossRef](#)] [[PubMed](#)]
90. Merker, M.; Kohl, T.A.; Niemann, S.; Supply, P. The Evolution of Strain Typing in the Mycobacterium tuberculosis Complex. *Adv. Exp. Med. Biol.* **2017**, *1019*, 43–78. [[PubMed](#)]

-
91. Zhou, Z.; Alikhan, N.F.; Sergeant, M.J.; Luhmann, N.; Vaz, C.; Francisco, A.P.; Carriço, J.A.; Achtman, M. Grapetree: Visualization of Core Genomic Relationships among 100,000 Bacterial Pathogens. *Genome Res.* **2018**, *28*, 1395–1404. [[CrossRef](#)] [[PubMed](#)]
 92. Warrens, M.J. Inequalities between Multi-Rater Kappas. *Adv. Data Anal. Classif.* **2010**, *4*, 271–286. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.