UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO DE BIOTECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

Epidemiologia molecular e genômica de *Mycobacterium tuberculosis* na Região Sul do Brasil

Tese de Doutorado

Richard Steiner Salvato

Porto Alegre, outubro de 2022

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
CENTRO DE BIOTECNOLOGIA
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

Epidemiologia molecular e genômica de *Mycobacterium tuberculosis* na Região Sul do Brasil

Richard Steiner Salvato

Tese submetida ao Programa de Pós-Graduação em Biologia Celular e Molecular do Centro de Biotecnologia da UFRGS como requisito parcial para a obtenção dotítulo de Doutor em Ciências.

Orientadora: Maria Lucia Rosa Rossetti

Co-orientadores: Afrânio Lineu Kritski e João Perdigão

Salvato, Richard Steiner
Epidemiologia molecular e genômica de Mycobacterium tuberculosis na Região Sul do Brasil / Richard Steiner Salvato. -- 2022.
121 f.
Orientadores: Maria Lucia Rosa Rossetti, Afrânio Lineu Kritski.

Coorientador: João Ruben Lucas Mota Perdigão.

Tese (Doutorado) -- Universidade Federal do Rio Grande do Sul, Centro de Biotecnologia do Estado do Rio Grande do Sul, Programa de Pós-Graduação em Biologia Celular e Molecular, Porto Alegre, BR-RS, 2022.

1. Tuberculose. 2. Epidemiologia. 3. Genômica. I. Rosa Rossetti, Maria Lucia, orient. II. Kritski, Afrânio Lineu, orient. III. Perdigão, João Ruben Lucas Mota, coorient. IV. Título.

Instituições envolvidas

Centro de Biotecnologia/Universidade Federal do Rio Grande do Sul

Centro de Desenvolvimento Científico e Tecnológico/Centro Estadual de Vigilância em Saúde/Secretaria da Saúde do Estado do Rio Grande do Sul

Hospital Sanatório Partenon/ Secretaria da Saúde do Estado do Rio Grande do Sul

Faculdade de Medicina/Universidade Federal do Rio Grande

Faculdade de Medicina/Universidade Federal do Rio de Janeiro

Instituto de Investigação do Medicamento/Faculdade de Farmácia/Universidade de Lisboa

Laboratório de Biologia Molecular Aplicada a Micobactérias/Fundação Oswaldo Cruz (Fiocruz-RJ)

Laboratório de Biologia Molecular Microbiologia e Sorologia/Universidade Federal de Santa Catarina

Laboratório de Biologia Molecular/Universidade Luterana do Brasil

Fontes financiadoras

Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)

Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)

Fundação de Amparo à pesquisa do Estado do Rio Grande do Sul (FAPERGS)

Secretaria da Saúde do Estado do Rio Grande do Sul (SES/RS)

ÍNDICE

LISTA DE ABREVIATURAS, SÍMBOLOS E UNIDADES

TB – Tuberculose

WGS - Sequenciamento do genoma completo, do inglês *whole genome sequencing*

M. tuberculosis - *Mycobacterium tuberculosis*

MIRU-VNTR - *Variable Number of Tandem Repeats of Mycobacterial Intersperced Repetitive Units*

COVID-19 - Doença causada por coronavírus (SARS-CoV-2), do inglês *Coronavírus disease 2019*

HIV - Vírus da imunodeficiência humana, do inglês *human immunodeficiency virus*

TB-RR - Tuberculose resistente à rifampicina

TB-MDR - Tuberculose multidroga-resistente

US$ - Dólar americano

BAAR - Bacilo álcool-ácido resistente

LJ - *Löwenstein-Jensen*

DNA - Ácido desoxirribonucleico, do inglês *Deoxyribonucleic Acid*

PCR - Reação em cadeia da polimerase, do inglês *polymerase chain reaction*

IS6110 – Elemento de inserção encontrado exclusivamente dentro dos membros do Complexo *Mycobacterium tuberculosis*

RFLP - Polimorfismo no Comprimento de Fragmentos de Restrição, do inglês *Restriction fragment length polymorphism*

MLST - Tipagem de sequência multiloco, do inglês *Multilocus sequence typing*

RD - Regiões de Diferença

CMTB - Complexo *Mycobacterium tuberculosis*

LSP - Polimorfismo de Grande Sequência, do inglês *large sequence polymorphism*

SNP – Polimorfismo de nucleotídeo único, do inglês *Single nucleotide polymorphisms*

TB pré-XDR – Tuberculose pré extensivamente resistente

TB-XDR - Tuberculose extensivamente resistente

NGS – Sequenciamento de Nova Geração, do inglês *Next-generation sequencing*

NCBI - *National Center for Biotechnology Information*

LAM - Latino-americana e mediterrânea

GC – Cluster genômico

RNAp – RNA polimerase

Wt – *Wild type*

SIT - *Spoligotype international types*

μg - Micrograma

mL – Mililitro

MIC – Concentração mínima inibitória, do inglês *Minimum inhibitory concentration*

LACEN - Laboratório Central de Saúde Pública

OMS – Organização Mundial da Saúde

TB-TDR - Tuberculose totalmente resistente

RESUMO

A tuberculose (TB) é uma das dez principais causas de morte no mundo e até 2020 tem sido a principal entre aquelas causadas por um único agente infeccioso. Apesar de todos os esforços realizados, a TB permanece como um dos mais importantes desafios de saúde pública em nível mundial. Estima-se que 10 milhões de novos casos ocorram anualmente em todo mundo e aproximadamente 1,4 milhão de mortes pela doença causada pelo agente etiológico *Mycobaterium tuberculosis (M. tuberculosis)*. O sequenciamento do genoma completo (WGS) se tornou nos últimos anos uma importante ferramenta para estimar a relação genética e clonalidade de diferentes isolados clínicos de *M. tuberculosis*, possibilitando a reconstrução das cadeias de transmissão e sua direcionalidade com um poder de resolução não atingível pelos métodos de tipagem clássicos, como *spoligotyping* e MIRU-VNTR. Simultaneamente, o WGS também permite a caracterização completa dos genes associados à resistência às drogas anti-TB. O presente estudo de doutorado avaliou a diversidade genômica dos isolados de *M. tuberculosis* circulantes na Região Sul do Brasil, para: (i) identificar os principais genótipos associados à resistência as drogas anti-TB; (ii) estimar a distância genômica entre isolados para identificar cadeias de transmissão da doença; (iii) elucidar a dinâmica de transmissão de *M. tuberculosis* e evolução da resistência no Sul do Brasil. Na análise dos resultados em isolados clínicos de *M. tuberculosis* provenientes do Rio Grande do Sul e de Santa Catarina, observamos eventos de transmissão contínua de *M. tuberculosis*, principalmente entre cepas resistentes a múltiplas drogas em estabelecimentos prisionais e na comunidade. Tais resultados sinalizam a necessidade da inclusão da vigilância genômica da TB como estratégia permanente dos programas de controle da TB na esfera nacional para interromper a transmissão de *M. tuberculosis* na região.

ABSTRACT

Tuberculosis (TB) is one of the ten leading causes of death in the world and until 2020 it has been the main among those caused by a single infectious agent. Despite all efforts, TB remains as one of the most important public health challenges worldwide. It is estimated that 10 million new TB cases are reported annually worldwide and approximately 1.4 million deaths occur every year from the disease caused by the etiological agent *Mycobaterium tuberculosis* (*M. tuberculosis*). Whole genome sequencing (WGS) has become in recent years an important tool to estimate the genomic relationship between *M. tuberculosis* isolates, allowing the reconstruction of transmission chains and their directionality between clusters that would otherwise be impossible to discriminate by classical typing methods such as spoligotyping and MIRU-VNTR. Simultaneously, the WGS also allows the complete characterization of genes associated with resistance and the detection of resistance-determining variants. In this work, we evaluated the genomic diversity of *M. tuberculosis* isolates circulating in Southern Brazil, to: (i) identify the main genotypes associated with resistance to anti-TB drugs; (ii) estimate the genomic distance between isolates to identify disease transmission chains; (iii) to elucidate the dynamics of *M. tuberculosis* transmission and evolution of resistance in different regions. The analysis results from *M. tuberculosis* strains circulating in Rio Grande do Sul and Santa Catarina in Southern Brazil, showed several events of resistant *M. tuberculosis* strains ongoing transmission, mainly among multidrug resistant strains in prison establishments and community. These results demonstrate the need to include TB genomic surveillance as a permanent strategy for TB control programs at the national level and to interrupt the *M. tuberculosis* transmission in the region.

1 INTRODUÇÃO

1.1 Tuberculose

Não obstante o fato de se tratar de uma doença curável e da existência de regimes terapêuticos eficazes, a tuberculose (TB) persiste como um dos principais desafios de saúde pública mundial, principalmente em países de baixo e médio desenvolvimento. Tratando-se de uma doença transmissível, figura como uma das principais causas de morte em todo o mundo e até o surgimento do SARS-CoV-2 e respectiva doença (COVID-19), a TB era a principal causa de morte de um único agente infeccioso, desde 2015, superando o vírus da imunodeficiência humana (HIV) (WHO, 2021). A TB é uma doença cujas evidências de infecções estavam presentes em registos arqueológicos, designadamente em ossadas humanas datadas em milhares de anos (HERSHKOVITZ *et al.*, 2015) e seu agente causador permaneceu desconhecido até o ano de 1882, quando Robert Koch anunciou a descoberta do bacilo responsável pela doença, mais tarde denominado *Mycobacterium tuberculosis* (*M. tuberculosis*) (SAKULA, 1982). A TB é transmitida por via aérea, quando as pessoas com a doença ativa expelem o bacilo para o ar, principalmente ao tossir e esses bacilos são inalados por outro indivíduo. *M. tuberculosis* pode infectar os pulmões e causar TB pulmonar, mas também outros locais (TB extrapulmonar) como pleura, nódulos linfáticos, abdome, sistema geniturinário, pele e ossos (WHO, 2021).

As partículas aerossóis contendo os bacilos ao serem inaladas pelo indivíduo chegam até os alvéolos pulmonares. Nos alvéolos, *M. tuberculosis* que é uma bactéria intracelular com maior afinidade pelos macrófagos alveolares se aloja. A identificação pelos macrófagos ativa uma resposta imune inata inicial que leva ao recrutamento de células inflamatórias nos pulmões, e nesse caso a evolução clínica da TB pode ocorrer de formas diferentes: o recrutamento de células T, células B, macrófagos e outros leucócitos levando a formação de granulomas, que conseguem conter *M. tuberculosis*, permanecendo assim em um estado latente de infecção, no qual não há sintomas clínicos, ou ocorrer a liberação de *M. tuberculosis* dos granulomas nas vias aéreas e esse se dispersar e replicar levando a TB pulmonar ativa, quando ocorre a infecção e o indivíduo desenvolve os sintomas característicos da doença (NUNES-ALVES *et al*., 2014). Como na TB latente os bacilos não são eliminados, a ativação da TB pode ocorrer em outro momento, geralmente no contexto de imunossupressão, desnutrição ou infecção por outros patógenos. No

entanto, durante a infecção latente, não há transmissão da bactéria, alterações patológicas nos pulmões ou sintomas da doença (Figura 1) (GETAHUN *et al*., 2015).

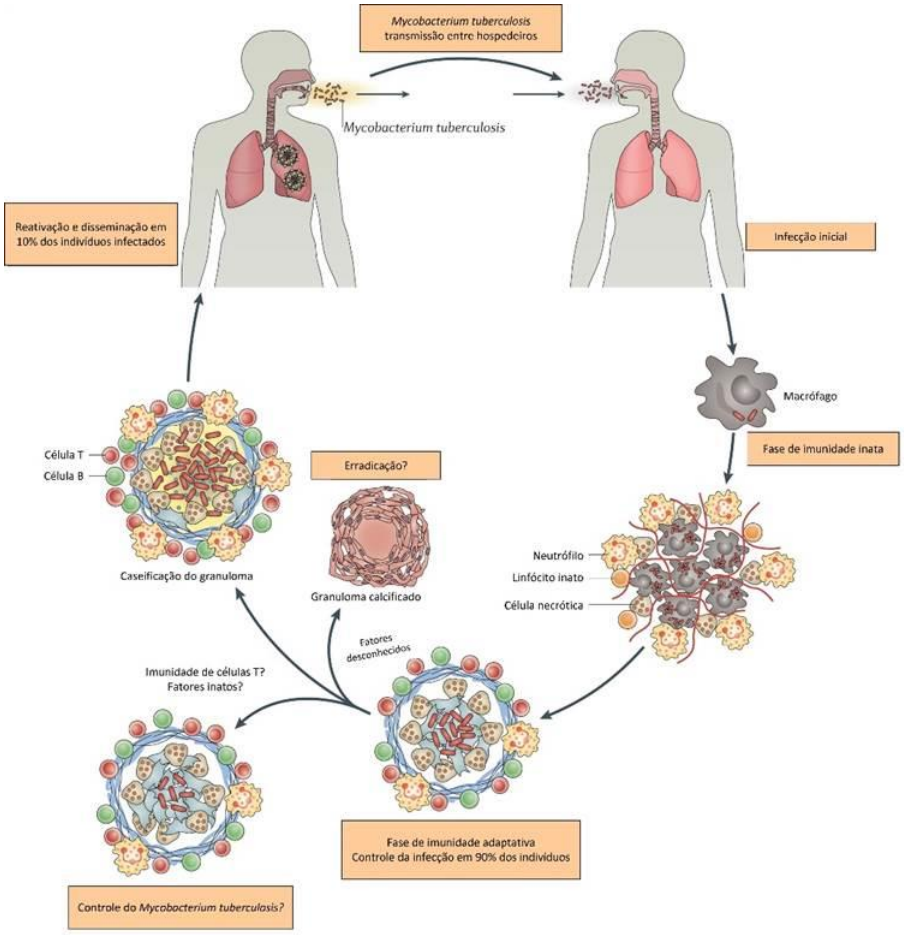


Figura 1. Patogênese da TB. A infecção é iniciada pela inalação de gotículas de aerossol contendo bactérias. Os estágios iniciais da infecção são caracterizados por respostas imunes inatas envolvendo o recrutamento de células inflamatórias para o pulmão. Após a disseminação bacteriana para o linfonodo de drenagem, a apresentação de antígenos bacterianos pelas células dendríticas leva ao *priming* das células T e desencadeia uma expansão das células T antígeno-específicas, que são recrutadas para o pulmão. O recrutamento de células T imunes, células B, macrófagos ativados e outros leucócitos leva ao estabelecimento de granulomas, que podem conter o *M. tuberculosis*. A maioria dos indivíduos infectados permanecerá em um estado de infecção latente, no qual não há sintomas clínicos. Uma pequena porcentagem dessas pessoas eventualmente progredirá e desenvolverá doença ativa, o que pode levar à liberação de *M. tuberculosis* de granulomas rompidos nas vias aéreas. Quando indivíduos com TB ativa tossem, podem gerar gotículas infecciosas que propagam a infecção. Figura traduzida de (NUNES-ALVES *et al*., 2014).

Em regiões onde os Programas de Controle de TB têm elevada qualidade, pacientes com TB droga sensível submetidos ao tratamento anti-TB de 6 meses, aproximadamente 85% deles, evoluem para cura ou tratamento completo. Os tratamentos anti-TB utilizados atualmente começaram a ser desenvolvidos na década de 1940. O tratamento recomendado para pessoas com TB suscetível a drogas consiste em um regime que utiliza quatro drogas de primeira linha: isoniazida, rifampicina, etambutol e pirazinamida. O *Global TB Drug Facility* fornece um regime completo de seis meses de tratamento por cerca de US$ 40 por pessoa. O *Global TB Drug Facility* é um programa desenvolvido pela Parceria das Nações Unidas para o Fim da TB (*Stop TB*) - que foi criada em 2001 visando a eliminação da TB como um problema de saúde pública. Suas 1.500 organizações parceiras incluem organizações internacionais, não governamentais e governamentais e grupos de pacientes - para expandir o acesso e a disponibilidade de medicamentos para TB para tratar até dez milhões de pacientes e ajudar os países a atingir as metas globais de controle da TB propostas pela Organização Mundial da Saúde (OMS) (WHO, 2021).

Quando uma pessoa é diagnosticada com TB resistente à rifampicina (TB-RR) ou TB multidroga-resistente (TB-MDR, definida como resistência à isoniazida e rifampicina, as duas drogas anti-TB mais potentes) o tratamento é mais longo (12 a 18 meses) e pode requerer a utilização de medicamentos de segunda linha, por sua vez de maior custo (≥US$ 1000 por pessoa) e que causam mais efeitos colaterais (WHO, 2021). Tais fatos, tornam a resistência às drogas anti-TB um dos maiores desafios globais para o controle da doença, demandando novas alternativas de diagnóstico rápido da resistência e novas drogas para o tratamento.

1.2 Epidemiologia

A TB é uma doença tão antiga quanto a população humana, e diferentes evidências científicas como alterações morfológicas em esqueletos, presença de DNA e informações filogenéticas de isolados com milhares de anos mostram que a TB co-evoluiu com os humanos por milhares de anos (DONOGHUE *et al*., 2011).

Apesar dos inúmeros avanços em relação ao diagnóstico e tratamento da TB, ela continua como um dos principais problemas de saúde pública, em nível global. A TB acomete indivíduos de todas as classes sociais e em todo o mundo e, apesar dos avanços

científicos e de gestão obtidos nas últimas décadas, nenhum país ainda foi capaz de erradicar a doença. No entanto, mesmo com a presença da doença em todos os continentes, a TB é considerada uma doença associada a pobreza, visto que a grande maioria dos casos acomete pessoas vivendo em países de baixa ou média renda. Os países do BRICs (Brasil, Federação Russa, Índia, China e África do Sul), por exemplo, contabilizam 47% do número total de casos de TB registrados mundialmente (Figura 2) (WHO, 2021).

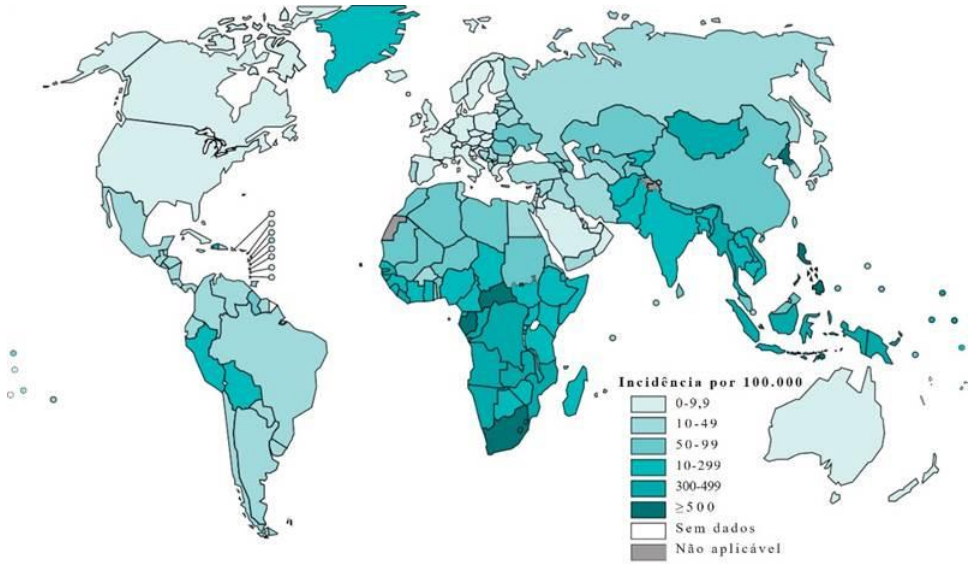


Figura 2. Taxas de incidência de TB por país em 2020, conforme estimado pela OMS. Figura adaptada de (WHO, 2021).

Em 2020, estima-se que 10 milhões de pessoas adoeceram por TB em todo o mundo e os 30 países com maior incidência (incluindo o Brasil) foram responsáveis por 86% dos novos casos da doença. Um total de 1,5 milhão de pessoas morreram de TB em 2020 (incluindo 214.000 pessoas com HIV). Em todo o mundo, a TB é a 13ª principal causa de morte e a segunda principal causa de morte infecciosa após o COVID-19 (WHO, 2021).

Anualmente, cerca de 70.000 novos casos de TB e 4.500 óbitos pela doença são notificados em todo o Brasil, resultando em uma taxa de incidência de 37,1 casos/100.000 habitantes registrados em 2019. Em 2021 essa taxa diminuiu para 32 casos/100.000 habitantes, no entanto, sabe-se que mundialmente houve uma grande queda nas notificações do número de pessoas recém-diagnosticadas com TB em 2020, em

comparação com 2019, reflexo do direcionamento dos serviços de saúde para atendimento da pandemia da COVID-19 (BRASIL, 2022).

Além disso, em nível mundial, atualmente, cerca de meio milhão de pessoas são diagnosticadas com TB resistente à rifampicina ou multidroga-resistente a cada ano, tornando um desafio para os esforços de diagnóstico, prevenção, tratamento e controle da doença de muitos países, principalmente aqueles com alta carga de TB (WHO, 2021). No Brasil, entre 2015 e 2021, foram diagnosticados 6.698 novos casos de TB resistente no país, sendo 848 desses diagnosticados em 2021 (BRASIL, 2022).

Em relação a região sul do Brasil composta por três estados, o Rio Grande do Sul apresenta as maiores incidências de TB e óbitos pela doença. Em 2021, o Rio Grande do Sul teve uma taxa de incidência de 36,5 novos casos por 100 mil habitantes, Santa Catarina uma taxa de 20,6, enquanto o Paraná apresenta uma menor incidência em relação aos dois primeiros (16 casos/100 mil). Quanto a mortalidade por TB, o Rio Grande do Sul também tem a maior taxa dentre os três estados da região sul, registrando coeficiente de mortalidade por TB de 2,4 óbitos por 100 mil habitantes, seguido do Paraná (1,2 óbitos/100 mil) e Santa Catarina (1,1 óbitos/100 mil) (BRASIL, 2022).

1.3 Diagnóstico

O diagnóstico da TB pode ser realizado por meio de diferentes métodos, dentre eles: exames bacteriológicos como baciloscopia e cultura de escarro, exame radiológico de tórax, testes imunológicos e técnicas de identificação molecular.

A baciloscopia é o exame microscópico de escarro, um método de detecção simples, de baixo custo e sua realização é viável na maioria dos laboratórios e unidades de saúde. Este exame foi concebido por Robert Koch em 1882, e consiste na coleta do escarro do paciente e na coloração do mesmo com carbolfucsina ou fluorocromos. Uma vez que *M. tuberculosis* possui propriedade bacilo álcool-ácido resistente (BAAR), ao entrar em contato com um dos corantes empregados na técnica, sua coloração permanece rósea ou avermelhada, não apresentando descoloração sob a ação de álcool e ácidos. A interpretação dos resultados da baciloscopia se dá pela contagem de bacilos por campo visualizado (WHO, 2015).

A cultura de escarro consiste em oferecer um meio propício para o desenvolvimento da micobactéria, oferecendo a nutrição adequada para o crescimento e a multiplicação dos microorganismos. Löwenstein-Jensen (LJ) e Ogawa-Kudoh são os meios sólidos mais utilizados para este fim. Este método de diagnóstico demanda de uma semana (quando realizada em meio líquido) a seis semanas (quando realizada em meio sólido) para obtenção de um resultado, ainda assim, a cultura é a técnica de diagnóstico considerada padrão ouro para a detecção da TB. Por meio da cultura em meio líquido, tornou-se possível também os sistemas de cultura automatizados como o MGIT96 – Mycobacteria Growth Indicator Tube (Becton & Dickinson, Sparks, MD, USA) – que passou a ser amplamente utilizado para o isolamento, detecção de resistência as drogas de primeira linha e, atualmente, também as drogas de segunda linha. A cultura automatizada proporcionou uma importante evolução no diagnóstico da TB, uma vez que é capaz de fornecer um resultado em tempo significativamente menor em relação a cultura em meio sólido (BRASIL, 2019).

Em alternativa a cultura, visando um resultado em tempo mais rápido, a partir de 2010, adotou-se mundialmente os testes rápidos moleculares. Estes testes são baseados na amplificação de ácidos nucleicos para detecção de DNA de *M. tuberculosis* e ainda possibilitam concomitantemente a identificação de cepas resistentes à determinadas drogas pela técnica de reação em cadeia da polimerase (PCR) em tempo real. Esses testes são capazes de fornecer o resultado em aproximadamente duas horas em ambiente laboratorial, e são realizados diretamente de uma amostra clínica. Um desses testes é o Xpert® MTB/RIF (Cepheid, Sunnyvale, CA, USA) que é implementado no Brasil para o diagnóstico da TB, substituindo a baciloscopia, e simultaneamente, detectando a resistência a rifampicina. (BRASIL, 2019).

Para detecção da resistência as drogas, a cultura e teste de sensibilidade as drogas é considerada o padrão ouro, no entanto, como a resistência as drogas em *M. tuberculosis* é adquirida, quase exclusivamente por mutações em genes associados aos mecanismos de ação específicos das drogas, a identificação dessas mutações por métodos moleculares constitui uma abordagem eficaz para a detecção rápida da maioria dos casos de TB resistente a medicamentos. Além disso, as técnicas moleculares podem ser realizadas em questão de horas ou dias e, ao contrário dos ensaios fenotípicos baseados em cultura e teste

de sensibilidade, não dependem de crescimento bacteriano que pode levar várias semanas (PANKHURST *et al*., 2016).

1.4 *Mycobacterium tuberculosis*

M. tuberculosis, agente causador da TB, é um bacilo de crescimento lento, que possui um envelope celular complexo, contribuindo para o carácter crônico da doença, e demanda regimes de tratamento mais longos e com múltiplas drogas. O envelope celular de *M. tuberculosis*, uma bactéria Gram-positiva com genoma rico em guanina e citosina, contém uma camada adicional além do peptidoglicano que é rica em lipídios, glicolipídios e polissacarídeos incomuns. Novas vias biossintéticas geram componentes da parede celular, como ácidos micólicos, ácido micocerósico, fenoltiocerol, lipoarabinomanan e arabinogalactano, e vários deles podem contribuir para a longevidade micobacteriana, desencadear reações inflamatórias no hospedeiro e atuar na patogênese da TB (COLE *et al*., 1998; KOLATTUKUDY *et al*., 1997).

O genoma completo da cepa de referência de *M. tuberculosis* (H37Rv - NC000962.3) foi completamente sequenciado em 1998 e sua análise revelou que o genoma contém 4.411.529 pares de base, incluindo cerca de 4.000 genes, um conteúdo de guanina-citosina médio de 65,4% e a presença de numerosos elementos polimórficos de DNA repetitivo que podem ser usados para tipagem molecular de *M. tuberculosis*. Por exemplo, o elemento de inserção IS6110, amplamente utilizado como marcador genotípico em epidemiologia molecular, é o mais abundante e o mais bem caracterizado e seu número de cópias em cada genoma varia entre as diferentes linhagens (BOTTAI *et al*., 2014; CAMUS *et al*., 2002; COLE *et al*., 1998).

M. tuberculosis foi por muitos anos considerado um patógeno geneticamente monomórfico. Somente após os primeiros métodos de genotipagem, incluindo a detecção do elemento de inserção IS6110 por Polimorfismo no Comprimento de Fragmentos de Restrição (RFLP) e a técnica de *spoligotyping* (KAMERBEEK *et al*., 1997) foi possível revelar a existência de grandes famílias de *M. tuberculosis* (BRUDEY *et al*., 2006; DEMAY *et al*., 2012; VAN EMBDEN *et al*., 1993). No entanto, esses métodos de genotipagem baseados em sequências de DNA repetitivas ou elementos móveis (ao invés de dados de sequenciamento de DNA) tinham limitações para inferência em nível de

linhagens e sublinhagens, e no início da década de 2000, tornou-se possível utilizar técnicas baseadas em tipagem de sequência multiloco (MLST) e sequenciamento (COMAS *et al*., 2009; MAIDEN *et al*., 1998).

O RFLP foi a primeira técnica utilizada como método padrão-ouro para genotipagem de *M. tuberculosis* (SUPPLY *et al*., 2006). Esta técnica emprega uma enzima de restrição *PvuII* associada à hibridização usando uma sequência parcial do elemento de inserção IS6110 como sonda (VAN EMBDEN *et al*., 1993). As sequências de inserção encontram-se dispersas por todo o genoma de *M. tuberculosis* em um número variável de cópias e podem ser usadas como marcador para comparar o número em que podem ser encontradas em cada família (THIERRY *et al*., 1990). No entanto, RFLP é um método trabalhoso, cujos resultados em nível laboratorial são de longa duração, exigindo semanas de incubação para o crescimento apenas de organismos que podem fornecer a grande quantidade de DNA necessária para sua realização (KREMER *et al*., 1999). Além disso, apresentava baixo poder discriminatório em cepas com menos de seis sequências de inserção IS6110 no DNA genômico, exibindo possíveis padrões idênticos mesmo em cepas não relacionadas epidemiologicamente (COWAN *et al*., 2005). Portanto, havia a necessidade de utilizar métodos complementares de genotipagem, como *spoligotyping* e/ou *Variable Number of Tandem Repeats of Mycobacterial Intersperced Repetitive Units* (MIRU-VNTR), que surgiram anos depois e eram baseados em PCR e apresentavam uma execução menos difícil (JONSSON *et al*., 2014; MATHEMA *et al*., 2006).

O método de *spoligotyping* é baseada na amplificação das sequências espaçadoras encontradas entre as regiões de repetição em tandem de 36pb e ocorrendo em número variável dentro do *locus* de repetição direta (DR) de *M. tuberculosis*. A detecção é feita por hibridização reversa com sondas específicas para as 43 sequências espaçadoras colocadas em uma membrana seguida de detecção por ensaio quimioluminescente (KAMERBEEK *et al*., 1997). O padrão de presença ou ausência dessas sequências espaçadoras gera um código binário de 43 elementos que permite a comparação interlaboratorial dos resultados (MATHEMA *et al*., 2006). Os variados padrões de *spoligotyping* apresentam características filogenéticas satisfatórias para diferenciação de isolados entre as diferentes famílias de *M. tuberculosis* (BRUDEY *et al*., 2006; COLL *et al*., 2014; KAMERBEEK *et al*., 1997).

Outra técnica de genotipagem clássica que tem sido utilizada nos últimos anos, baseia-se na detecção de sequências repetitivas que variam de 40 a 100 pares de base, encontradas em repetições em tandem, em diferentes *loci* distribuídos no genoma de *M. tuberculosis*, denominadas *Mycobacterial Interspersed Repetitive Units* (MIRU). Por meio desta técnica, observa-se que as diferentes linhagens apresentam, em diferentes *loci*, polimorfismo no número de repetições desses elementos *Variable Number of Tandem Repeats* (VNTR) (SUPPLY *et al*., 2006). Este método baseia-se na amplificação por PCR de múltiplos *loci* utilizando primers específicos para as regiões flanqueadoras de cada *locus* de repetição e na identificação dos tamanhos dos fragmentos amplificados, por algum método de separação, como migração em gel ou eletrocapilar, e atribuindo a cada tamanho de fragmento um número de repetições correspondente (MAZARS *et al*., 2001). A cada cepa tipada é atribuído um número correspondente ao número de repetições do *locus* MIRU, formando a base de um sistema de codificação. O método tem um poder discriminatório, proporcional ao número de *loci* analisados, em geral, ao utilizar apenas 12 *loci* é menos discriminatório que o RFLP-IS6110, porém, o conjunto de 15 *loci* tem sido proposto como equivalente ao RFLP-IS6110 e o 24 *loci* definido como uma ferramenta de melhor resolução adequada para estudos filogenéticos (SUPPLY *et al*., 2006).

No entanto, o uso de *spoligotyping* e MIRU é limitado para filogenética e classificação de cepas devido à propensão de eventos de evolução convergente envolvendo esses marcadores moleculares, causando a rápida mudança desses padrões de marcadores e ainda possibilitando a existência de mesmos padrões ou padrões semelhantes em isolados de *M. tuberculosis* que não são filogeneticamente relacionadas. Este problema é especialmente relevante para a técnica de spoligotyping dada a unidirecionalidade dos eventos genéticos inerentes ao processo de diversificação dos marcadores empregues. Para contornar este problema, deleções genômicas, também denominadas como Regiões de Diferença (RDs) têm sido usadas como marcadores para classificar cepas de *M. tuberculosis* entre as principais linhagens filogenéticas existentes (COSCOLLA & GAGNEUX, 2014).

Quanto a diversidade genômica, *M. tuberculosis* (também denominado *M. tuberculosis sensu stricto*) pertence ao Complexo *Mycobacterium tuberculosis* (CMTB), que compreende diversas espécies e subespécies bacterianas intimamente relacionadas,

incluindo *M. tuberculosis* e *M. africanum* que são adaptados a humanos, bem como várias formas adaptadas aos animais, i.e. *M. bovis*, *M. caprae*, *M. microti*, *M. pinnipedii*, *M. origis*, *M. mungi*, *M. suricattae*, o bacilo dassie e o bacilo do chimpanzé (COSCOLLA & GAGNEUX, 2014).

O CMTB é dividido em nove linhagens adaptadas aos humanos (linhagens 1, 2, 3, 4, 7 e 8) e *M. tuberculosis* var. *africanum* (linhagens 5, 6 e 9; *M. africanum*), que causam doenças humanas. Linhagens 2, 3 e 4, são conhecidas como linhagens modernas, pois essas linhagens se diversificaram mais recentemente do que as demais linhagens do CMTB. Esta diferenciação se faz pela ausência da região TbD1 nas linhagens modernas, que tem um papel importante na fisiopatologia da doença e contribuiu para a disseminação global e sucesso evolutivo dessas linhagens. As linhagens CMTB variam em sua distribuição geográfica e dispersão, sendo endêmicas em diferentes locais ao redor do globo, levando à hipótese de que os tipos de cepas são especificamente adaptados a diferentes populações humanas (BOTTAI *et al*., 2020; BRITES *et al*., 2018; BRITES & GAGNEUX, 2015; NGABONZIZA *et al*., 2020).

Algumas linhagens, são globalmente mais distribuídas do que outras. Entre estas linhagens mais bem espalhadas mundialmente, estão as linhagens 2 e 4. A linhagem 2 (também conhecida como linhagem do Leste Asiático, inclui a família de cepas Beijing) predomina no leste da Ásia, mas também está presente na Ásia Central, Rússia e África do Sul. A linhagem 4 (também conhecida como linhagem Euro-americana) circula principalmente em populações da Ásia, Europa, África e Américas. As linhagens 1 e 3 tem uma distribuição geográfica mais restrita limitada à África Oriental, Ásia Central, Sul e Sudeste Asiático. As linhagens mais restritas geograficamente são as linhagens 5, 6 e 7, todas associadas a regiões específicas da África. As linhagens 5 e 6 também conhecidas como *M. africanum, West Africa 1* e *West Africa 2*, respectivamente, ocorrem quase exclusivamente na África Ocidental ou em imigrantes recentes dessas regiões. A linhagem 6 ocorre principalmente na parte ocidental da África Ocidental, enquanto a linhagem 5 domina mais ao leste nas regiões que fazem fronteira com o Golfo da Guiné (Figura 3) (BLOUIN *et al*., 2012; COSCOLLA & GAGNEUX, 2014; DE JONG *et al*., 2010; GEHRE *et al*., 2013).

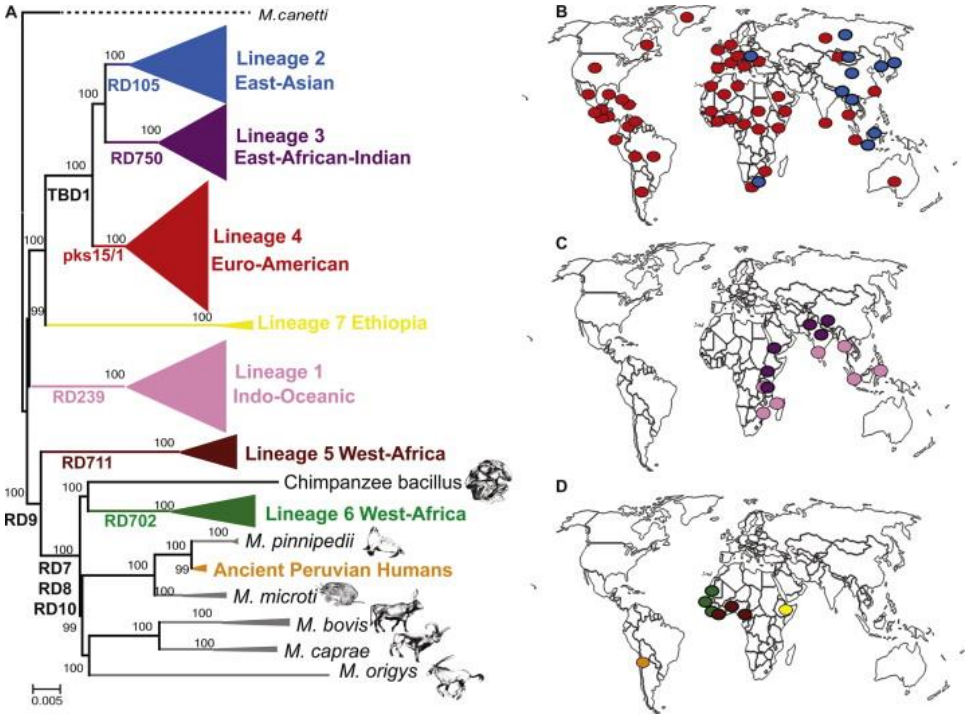


Figura 3. (A) Filogenia de máxima verossimilhança modificada de (BOS *et al*., 2014). O suporte dos ramos com 1.000 replicações de bootstrap é mostrado nos ramos e a árvore é enraizada pelo grupo externo *M. canettii*. Polimorfismos de Grande Sequência (LSPs) descritos em (BROSCH *et al*., 2002) são indicados ao longo dos ramos. A barra de escala indica o número de substituições de nucleotídeos por sítio. As imagens de B a D representam as linhagens do CMTB dominantes por país. Cada ponto corresponde a 1 dos 80 países representados nas 875 cepas do CMTB da coleção global de cepas analisada por (GAGNEUX *et al.,* 2006). O ponto amarelo e laranja representam a linhagem 7 na ETIÓPIA (FIRDESSA *et al*., 2013) e as cepas CMTB extintas do Peru, respectivamente (BOS *et al*., 2014). Em **(B)** mostra as linhagens geograficamente mais difundidas, em **(C)** as linhagens intermediárias e em **(D)** as linhagens mais restritas geograficamente. Figura traduzida de (COSCOLLA & GAGNEUX, 2014).

Em vários estudos foi descrito que as diferenças fenotípicas entre as linhagens do CMTB exercem influência nas características de transmissão, diagnóstico e desfecho clínico da TB pulmonar e extrapulmonar. As variações entre essas diferentes linhagens podem ser observadas consequentemente na manifestação de distintos fenótipos com diferenças no perfil resistência as drogas, transmissibilidade, virulência, resposta imune do indivíduo infectado, local da doença e gravidade. Esses fenótipos conferem vantagens para essas linhagens CMTB, podendo facilitar a disseminação da doença e aumentar as chances de um pior prognóstico para os pacientes (FORD *et al*., 2013; REILING *et al*., 2013).

Atualmente, a tipagem de linhagens de *M. tuberculosis* pode ser realizada baseada em um conjunto de SNPs espalhados por todo genoma, seja *in silico* a partir dos dados de sequenciamento do genoma completo (WGS do inglês *Whole Genome Sequencing*) ou por ensaios laboratoriais para identificação desses SNPs. Recentemente, um conjunto de 90 SNPs tem sido utilizado para tipagem de *M. tuberculosis*, principalmente a partir de dados de WGS, possibilitando a identificação de mais de 85 sublinhagens do CMTB. Isto possibilita uma maior resolutividade na filogenia do CMTB e permite uma melhor compreensão de seus padrões de transmissão, disseminação e evolução (Figura 4) (NAPIER *et al*., 2020).

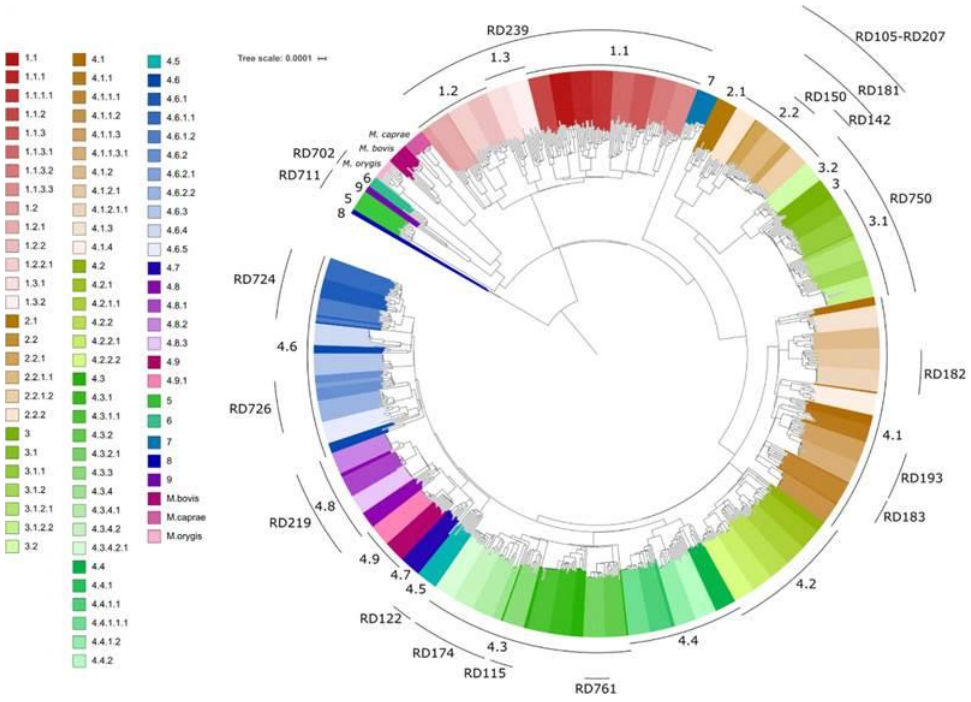


Figura 4. Árvore filogenética de isolados do complexo *Mycobacterium tuberculosis* representativos das diferentes sub-linhagens identificadas por tipagem baseada em SNPs. As Regiões de Diferença (RDs) importantes também são destacadas na árvore. Figura obtida de (NAPIER *et al*., 2020).

Tais métodos de tipagem do CMTB permitem o rastreamento da disseminação das suas diferentes linhagens e sublinhagens e é de extrema importância para o controle da TB. A rápida identificação da linhagem do CMTB permite: a) analisar as associações fenotípicas com os perfis de resistência as drogas, b) identificar a provável fonte de transmissão e c) auxiliar na prevenção e interrupção de cadeias de transmissão da micobactéria (MEEHAN *et al*., 2019). Assim, a tipagem molecular atua como uma

importante ferramenta para o controle da doença e melhor direcionamento das medidas de saúde pública a ela relacionadas.

1.5 Tratamento e resistência as drogas

Na década de 1940 surgiram a estreptomicina e o ácido para-aminossalicílico, as primeiras drogas capazes de tratar a TB. Nas décadas seguintes diversas outras drogas foram se tornando disponíveis para o tratamento da doença e hoje diferentes esquemas de tratamento estão disponíveis a depender do perfil de resistência do bacilo infectante e das condições do indivíduo sendo tratado (LEHMANN, 1946; SCHATZ *et al*., 2005).

Atualmente, o regime de tratamento padrão para casos novos de TB droga sensível, em adultos e adolescentes, consiste em 2 meses de isoniazida, rifampicina, pirazinamida e etambutol (fase intensiva), seguido de 4 meses de isoniazida e rifampicina (fase de manutenção). Quando identificada a resistência a alguma das drogas utilizadas no regime padrão, o esquema deve ser revisado, com a substituição de drogas além de serem adicionadas outras drogas ao regime dependendo da resistência identificada. De acordo com a OMS, os casos em que ocorre a resistência a rifampicina são definidos como TB resistente à rifampicina (TB-RR), enquanto aqueles em que a resistência adicional a isoniazida é identificada denomina-se TB multirresistente (TB-MDR). Os casos de TB-MDR e TB-RR são frequentemente agrupados como TB-MDR/RR e ambos são elegíveis para tratamento com regimes de TB-MDR (WHO, 2022).

As drogas a serem incluídas no tratamento da TB MDR/RR são agrupadas em três diferentes grupos: Grupo A (levofloxacina/moxifloxacina, bedaquilina, linezolida), Grupo B (clofazimina, cicloserina/terizidona) e Grupo C (etambutol, delamanida, pirazinamida, imipenem–cilastatina, meropenem, amicacina, estreptomicina, etionamida/protionamida, ácido p-aminosalicílico), todas essas denominadas drogas de segunda linha. Quando ocorre ainda a resistência a algumas dessas drogas, os casos podem ser também classificados como TB pré extensivamente resistente (TB pré-XDR), causada por cepas de *M. tuberculosis* que preenchem a definição de MDR/RR-TB e que também são resistentes a qualquer fluoroquinolona (levofloxacina e moxifloxacina); e TB extensivamente resistente (TB-XDR) causada por cepas de *M. tuberculosis* que atendem à definição de MDR/RR-TB e que também são resistentes a qualquer fluoroquinolona e pelo menos uma das drogas do

grupo A (levofloxacina ou moxifloxacina, bedaquilina e linezolida). Recentemente, diferentes casos de TB totalmente resistente aos medicamentos (TB-TDR) têm sido descrito e são caracterizados por apresentarem resistência a todos os antimocrobianos testados, além de alguns dos que estão atualmente em processo de descoberta (WHO, 2022).

O mecanismo molecular mais comum e frequente de aquisição de resistência em cepas de *M. tuberculosis* é por meio da aquisição de mutações genômicas como SNPs, deleções ou inserções nos genes que codificam os alvos das drogas, diminuindo ou inibindo a eficácia destas drogas (DOOKIE *et al*., 2018). Essas alterações genômicas ocorrem de forma espontânea e cepas mutantes geralmente apresentam alguma vantagem sobre cepas não mutadas, aumentando sua sobrevivência na presença de drogas anti-TB. A vantagem de cepas carregando essas mutações frente à exposição as drogas colaboram para que esse tipo de cepa venha a se disseminar na população. Além disso, a exposição de *M. tuberculosis* a doses inferiores de droga, pode levar a aquisição de mutações (NGUYEN, 2016). Na tabela 1, são descritos os principais alvos moleculares das principais drogas anti-TB e os mecanismos de aquisição de resistência comumente encontrados mundialmente.

Além da resistência causada por aquisição de mutações, vários mecanismos distintos de resistência inata também são descritos, entre eles: diminuição da permeabilidade do envelope do bacilo, modificação de drogas por enzimas de *M. tuberculosis*, existência de bombas de efluxo removendo drogas que foram capazes de atravessar o envelope celular, e tolerância fenotípica à droga por diminuição da taxa de crescimento e parada metabólica. Todavia esses mecanismos ocorrem em frequência muito inferior aos previamente citados, além de não produzirem resistência estável ou de alto nível (SINGH *et al*., 2020).

Tabela 1. Mecanismo de ação das principais drogas anti-TB, alvos, genes envolvidos e principais mutações associadas a resistência.

Fármaco	Modo de ação	Alvo	Gene	Função do gene	Mutações mais frequentes
Isoniazida	Inibição da síntese do ácido micólico	Ácido micólico	*katG*	Catalase-peroxidase	Ser-315-Thr
			inhA	Enoil-ACP redutase	C-15-T
			ndh	NADH desidrogenase II	Arg-13-Cys, Val-18-Ala
			ahpC	Alquil-hidroperoxidase	C-39-T, G-9-A,
				β-Cetoacil-ACP sintase	Gly-269-Ser
Rifampicina	Inibição da transcrição	RNA polimerase	*rpoB*	Subunidade β da RNA polimerase	Ser-450-Leu
Pirazinamida	Inibição da produção de energia e tradução	Ácido graxo sintase I, proteína ribossômica S1	*pncA*	Pirazinamidase	Asp-12-Ala/Asn,Leu-85-Pro
			rpsA	Proteína Ribosomal S1	Deletion Ala438, Thr-5-Ala
			panD	Aspartato descarboxilase	Ala-128-Ser, Val-138-Aal
Etambutol	Inibição da síntese de arabinogalactano	Arabinosil transferases	*embCAB ubiA/Rv3806c*	Arabinosil transferases	Met-306-Val/Ile/Leu
Estreptomicina	Inibição da síntese de proteínas	Subunidade ribossômica 30S	*rpsL rrs gidB*	Proteína ribossômica S12, RNAr 16S, metiltransferase RNAr 16S	Lis-43-Arg A-1401-G SNP Leu-16-Arg
Amicacina/ Canamicina	Inibição da síntese de proteínas	Subunidade ribossômica 30S	*rrs*	RNAr 16S	A-1401-G SNP
			eis	Aminoglicosídeo acetiltransferase	G-37-T, G-10-A, G-14-T SNPs
Capreomicina	Inibição da síntese de proteínas	Subunidades ribossômicas 30S e 50S	*rrs*	RNAr 16S	A-1401-G SNP
			tylA	Metiltransferase RNAr	G-223-T SNP
				Aminoglicosídeo acetiltransferase	G-37-T, C-12-T SNPs
Etionamida	Inibição da síntese do ácido micolico	Ácido micólico	*ethA*	Monooxigenase	Leu-397-Arg, Leu-328-Met
			inhA	Enoil-ACP redutase	Ile-21-Thr/Val, Ser-94-Ala
			ndh	NADH desidrogenase	Arg-13-Cys, Val-18-Ala
			mshA	Glicosiltransferase	Val-171-Gly, Ala-187-Val

Fluoroquinolonas	Inibição da replicação, transcrição e reparo	DNA girase	*gyrA* *gyrB*	Subunidade A da DNA girase Subunidade B da DNA girase	Ala-90-Val, Asp-94-Gly/Tyr Asn-533-Thr RGM (4)
Ácido p-aminosalicílico	Inibição da síntese de folato	Timidilato sintase, diidrofolato sintase, diidrofolato redutases	*thyA* *folC* *ribD*	Timidilato sintase Diidrofolato sintase Diidrofolato redutases	Thr-202-Ala, Val-261-Gly Glu-153-Aal, Asn-73-Ser G-11-A SNP
Cicloserina	Inibição da síntese de peptidoglicano	Alanina racemase, D-alanina-D-alanina ligase, D-serina /L e D-alanina/glicina/D-cicloserina simportador de prótons, L-alanina desidrogenase	*alr* *ddl* *cycA Ald*	Alanina racemase D-alanina-D-alanina ligase D-serina /L e D-alanina / glicina / D-cicloserina simportador de prótons, L-alanina desidrogenase	G-10-T SNP – Gly-122-Ser

Adaptado de Nasiri *et al*. (2017).

1.6 Epidemiologia genômica

A melhor compreensão da diversidade das linhagens de *M. tuberculosis*, em um nível de resolução proporcionado pela análise de SNPs ao longo do genoma, é capaz de definir se tal diversidade tem impacto na epidemiologia da doença em uma população e se tem relevância para o controle da TB em uma determinada região (COMAS & GAGNEUX, 2009; GAGNEUX & SMALL, 2007).

Tal processo foi facilitado com o surgimento do Sequenciamento de Nova Geração (NGS do inglês Next Generation Sequencing) que proporciona a geração de sequências do genoma completo de *M. tuberculosis* em pouco tempo e com custo cada dia menor. Esse processo pode ser observado no número crescente de sequências de genomas que são depositados em bancos de dados de acesso público como o NCBI (*National Center for Biotechnology Information*) e auxiliam a elucidar aspectos da diversidade genômica do CMTB. O termo "NGS" é comumente utilizado para descrever plataformas de sequenciamento baseadas em métodos mais recentes do que o sequenciamento original baseado na química de terminação da cadeia de Sanger ou o método químico de Maxam e Gilbert (MAXAM & GILBERT, 1977; SANGER *et al*., 1977). Enquanto o método de sequenciamento de Sanger original exigia um primer específico para iniciar a leitura por meio de uma sequência modelo, a estratégia de NGS deriva da abordagem de sequenciamento *shotgun* empregada no Projeto Genoma Humano (ZHANG *et al*., 2011).

As tecnologias de NGS são baseadas na fragmentação genômica, ligação de adaptadores e produção de leituras começando em locais aleatórios por todo o genoma. Os moldes de DNA são, portanto, sequenciados em paralelo em um processo em tempo real, no qual cada molde clonal ou molécula única é sequenciado e analisado individualmente (ZHANG *et al*., 2011). Todas as plataformas de NGS acabam produzindo grandes quantidades de dados em um período relativamente curto e, portanto, o motivo pelo qual NGS também está associado ao termo sequenciamento de alto rendimento. Dependendo da tecnologia, o comprimento de leitura pode variar, pois as primeiras tecnologias NGS começaram produzindo comprimentos de leitura menores aos que eram obtidos pelo sequenciamento de Sanger, no entanto em tecnologias mais recentes é possível produzir longas leituras. Outro aspecto importante a ser considerado no NGS é a profundidade, ou

seja, o número de leituras que se sobrepõem a uma determinada posição de nucleotídeo; essa profundidade permite a realização de análises mais apuradas de eventos de infecção mista ou heteroresistência por exemplo (PAREEK *et al*., 2011).

O uso dessas tecnologias de NGS que hoje possibilitam o sequenciamento do genoma completo auxiliam na melhor compreensão sobre a diversidade genômica de *M. tuberculosis*, produzindo importantes conhecimentos sobre a diversidade, modo de evolução e adaptação ao hospedeiro desse patógeno, e isto em diferentes regiões do mundo. Por meio do WGS também é possível realizar a caracterização completa dos genes associados à resistência a medicamentos e a detecção de variantes determinantes de resistência. Nos últimos anos, principalmente os países de alta renda, estão deixando de utilizar os testes de suscetibilidade fenotípicos e estão migrando para predição de resistência baseada na identificação de mutações associadas a resistência por WGS (MEEHAN *et al*., 2019).

Dada a necessidade de uma melhor compreensão da diversidade genômica dos isolados de *M. tuberculosis* circulando em diferentes regiões do Brasil e a relevância dessa informação para orientar os gestões sobre as melhores medidas de controle da doença em nível local; a geração de sequências genômicas de *M. tuberculosis* e a análise desses dados para suprir as principais demandas de saúde pública mostra-se como uma ferramenta na busca da eliminação da TB proposta pela Organização Mundial da Saúde em 2015 e foi o objetivo desta presente tese.

2 OBJETIVO

O principal objetivo deste trabalho foi caracterizar a diversidade genômica dos isolados de *M. tuberculosis* circulando na região sul do Brasil, e como objetivos específicos se propôs a:

(i) Identificar os principais genótipos e mecanismos associados à resistência as drogas anti-TB bem como a sua prevalência.

(ii) Posicionar num contexto filogenético global as principais famílias ou cepas circulantes na região sul do Brasil.

(iii) Caracterizar a trajetória evolutiva de cepas multirresistentes e dinâmica de aquisição de resistência.

(iv) Elucidar a dinâmica de transmissão de *M. tuberculosis*, identificação de clusters genômicos e dispersão geotemporal.

3 MANUSCRITOS PUBLICADOS

3.1 CAPÍTULO I

O manuscrito que constitui este capítulo, intitulado **"Molecular epidemiology of *Mycobacterium tuberculosis* in Brazil before the whole genome sequencing era: a literature review"** consistiu uma revisão sistemática da literatura sobre o uso do RFLP-IS6110 e MIRU-VNTR no Brasil.

O trabalho foi publicado na revista **Memórias do Instituto Oswaldo Cruz** (https://memorias.ioc.fiocruz.br/), com fator de impacto **JCR 2021 = 2,747**.

O manuscrito está disponível em: https://doi.org/10.1590/0074-02760200517.

Molecular epidemiology of *Mycobacterium tuberculosis* in Brazil before the whole genome sequencing era: a literature review

Emilyn Costa Conceição[1,2,3]+, Richard Steiner Salvato[4,5], Karen Machado Gomes[6], Arthur Emil dos Santos Guimarães[7,8], Marília Lima da Conceição[7,8], Ricardo José de Paula Souza e Guimarães[9], Abhinav Sharma[10], Ismari Perini Furlaneto[8], Regina Bones Barcellos[5], Valdes Roberto Bollela[11], Lívia Maria Pala Anselmo[11], Maria Carolina Sisco[3,12], Cristina Viana Niero[13], Lucilaine Ferrazoli[14], Guislaine Refrégier[15], Maria Cristina da Silva Lourenço[2], Harrison Magdinier Gomes[3], Artemir Coelho de Brito[16], Marcos Catanho[17], Rafael Silva Duarte[12], Philip Noel Suffys[3], Karla Valéria Batista Lima[7,8]

[1]Fundação Oswaldo Cruz-Fiocruz, Instituto Nacional de Infectologia Evandro Chagas, Programa de Pós-Graduação em Pesquisa Clínica e Doenças Infecciosas, Rio de Janeiro, RJ, Brasil
[2]Fundação Oswaldo Cruz-Fiocruz, Instituto Nacional de Infectologia Evandro Chagas, Laboratório de Bacteriologia e Bioensaios em Micobactérias, Rio de Janeiro, RJ, Brasil
[3]Fundação Oswaldo Cruz-Fiocruz, Instituto Oswaldo Cruz, Laboratório de Biologia Molecular Aplicada a Micobactérias, Rio de Janeiro, RJ, Brasil
[4]Universidade Federal do Rio Grande do Sul, Programa de Pós-Graduação em Biologia Celular e Molecular, Porto Alegre, RS, Brasil
[5]Secretaria Estadual de Saúde do Rio Grande do Sul, Centro Estadual de Vigilância em Saúde, Centro de Desenvolvimento Científico e Tecnológico, Porto Alegre, RS, Brasil
[6]Fundação Oswaldo Cruz-Fiocruz, Escola Nacional de Saúde Pública Sergio Arouca, Centro de Referência Professor Hélio Fraga, Laboratório de Referência Nacional para Tuberculose e outras Micobacterioses, Rio de Janeiro, RJ, Brasil
[7]Universidade do Estado do Pará, Instituto de Ciências Biológicas e da Saúde, Pós-Graduação Biologia Parasitária na Amazônia, Belém, PA, Brasil
[8]Instituto Evandro Chagas, Seção de Bacteriologia e Micologia, Ananindeua, PA, Brasil
[9]Instituto Evandro Chagas, Laboratório de Geoprocessamento, Ananindeua, PA, Brasil
[10]International Institute of Information Technology, Department of Data Science, Bangalore, India
[11]Universidade de São Paulo, Departamento de Clínica Médica da Faculdade de Medicina de Ribeirão Preto, Ribeirão Preto, SP, Brasil
[12]Universidade Federal do Rio de Janeiro, Instituto de Microbiologia Paulo de Góes, Laboratório de Micobactérias, Rio de Janeiro, RJ, Brasil
[13]Universidade Federal de São Paulo, Departamento de Microbiologia, Imunologia e Parasitologia, São Paulo, SP, Brasil
[14]Instituto Adolfo Lutz, Centro de Bacteriologia, Núcleo de Tuberculose e Micobacterioses, São Paulo, SP, Brasil
[15]Universit e Paris-Saclay, Ecologie Systematique Evolution, Centre National de la Recherche Scientifique, AgroParisTech, Orsay, France
[16]Coordenação Geral de Vigilância das Doenças de Transmissão Respiratória de Condições Crônicas, Brasília, DF, Brasil
[17]Fundação Oswaldo Cruz-Fiocruz, Instituto Oswaldo Cruz, Laboratório de Genética Molecular de Microrganismos, Rio de Janeiro, RJ, Brasil

Molecular-typing can help in unraveling epidemiological scenarios and improvement for disease control strategies. A literature review of *Mycobacterium tuberculosis* transmission in Brazil through genotyping on 56 studies published from 1996-2019 was performed. The clustering rate for mycobacterial interspersed repetitive units - variable tandem repeats (MIRU-VNTR) of 1,613 isolates were: 73%, 33% and 28% based on 12, 15 and 24-loci, respectively; while for RFLP-IS*6110* were: 84% among prison population in Rio de Janeiro, 69% among multidrug-resistant isolates in Rio Grande do Sul, and 56.2% in general population in São Paulo. These findings could improve tuberculosis (TB) surveillance and set up a solid basis to build a database of *Mycobacterium* genomes.

Key words: tuberculosis - *Mycobacterium tuberculosis* – genotyping - MIRU-VNTR typing - RFLP-IS*6110* – Brazil

Despite being an ancient disease, tuberculosis (TB) is still the leading cause of death among infectious diseases worldwide. From 2016 to 2020, Brazil has been on the World Health Organization (WHO) list of high burden countries for TB and TB/HIV co-infection.[1] In 2014, WHO proposed the End TB Strategy that targets TB prevention, care, control, and together with the Sustainable Development Goals (SDGs) aimed at trying to bring TB incidence and mortality on a global level to those observed in high-income countries.[2,3,4,5,6]

The three pillars of the End TB Strategy are: (i) integrated, patient-centered TB care and prevention, bold policies, and supportive systems (including universal health coverage, social protection, and action on determinants), (ii) intensified research and (iii) innovation. To face this scenario, the main strategy includes milestones (for 2020 and 2025) and quantitative targets (for 2030 and 2035) for three high-level indicators: incidence and mortality rates, as well as the percentage of TB patients and their households.[3,6]

In 2016, a Brazilian report discussing the Global End TB Strategy program was published as a technical report and a national TB research agenda was proposed to the establishment of the National TB Research Strategy

doi: 10.1590/0074-0276020517
Financial support: CAPES (Finance Code 001).
PNS and KVBL contributed equally to supervising this work.
+ Corresponding author: emilyncosta@gmail.com
https://orcid.org/ 0000-0002-7445-6620
Received 01 October 2020
Accepted 11 February 2021

Plan. One of the strategies to address the gap regarding general recommendations was to "create a coordination research group on fundamental and translational research that pursues increased collaboration among various laboratories to better utilise the available knowledge of different groups".[7] One of the key endorsed research areas was regarding the investigation on host-pathogen interaction targeting new genetic, molecular, immunological, or metabolic markers, including the use of genotyping and the "omics" supporting epidemiology, new diagnostics methods, studies about new vaccines and more recently new drugs.[7,8]

The association between specific *M. tuberculosis* strains and the increase of anti-TB drug resistance is one of the major drivers for mortality rate increase. The investigation of transmission sources and monitoring TB strains by molecular epidemiology studies complemented by molecular typing tools is therefore essential to control TB.[9]

Spacer-oligonucleotide-typing (spoligotyping), mycobacterial interspersed repetitive units - variable number tandem repeat (MIRU-VNTR) typing and the insertion sequence *6110* - restriction fragment length polymorphism (RFLP-IS*6110*) are among the most used genotyping methods for *M. tuberculosis* complex (MTBC) strains. However, due to their different resolving power, only the latter two are used to evaluate TB transmission and perform detailed molecular epidemiology.

MIRU-VNTR is the current reference technique due its higher discriminatory power and reproducibility. Except for RFLP-IS*6110*, because of those characteristics and ease of interpretation and storage, both spoligotypes and MIRU-VNTR based genotypes are stored in large international databases that allow inter-laboratory comparison of patterns while RFLP-IS*6110*, although considerable in number, are mostly composing local databases.[10,11]

Through this systematic literature review on the use of RFLP-IS*6110* and MIRU-VNTR, we aimed to (i) describe the Brazilian TB network laboratories structure, (ii) characterise molecular epidemiology studies in Brazil applied to TB; (iii) to study the genetic diversity of *M. tuberculosis* in the country, and (iv) to correlate these data to the national epidemiological scenario of TB.

MATERIALS AND METHODS

Data collection of the Brazilian tuberculosis policies - We have collected the information mainly from the Brazilian National Program of TB Control (NPTC), recently named as General Coordination for the Monitoring of Chronic Conditions Respiratory Transmission Diseases (Coordenação Geral de Vigilância das Doenças de Transmissão Respiratória de Condições Crônicas - CGDR),[12] which is responsible for establishing guidelines for disease control, manuals, and reports. The national recommendations are updated and disclosed in the technical notes of the NPTC and in the publication of the Brazilian Guidelines for Tuberculosis Control (BGTBC), first edited in 2011, and last published in 2019.[13]

Data collection on M. tuberculosis genotyping - Data were collected using PubMed (http://www.ncbi.nlm.nih.gov/pubmed), as well as the Brazilian virtual library BVS (Biblioteca Virtual em Saúde) database. For this, we used the keywords "MIRU-VNTR AND tuberculosis AND Brazil" and "RFLP AND tuberculosis AND Brazil". All papers were downloaded and information such as data, place and date of samples collections, study period, samples characteristics, genotyping techniques used, the year of publication and principal results obtained) were introduced into a Microsoft Office Excel spreadsheet (Albuquerque, United States).

Data analysis - We analysed all papers published until January 28th, 2020 summarised our approach, using the PRISMA flow diagram.[14]

For the present review, for comparison of RFPL-IS*6110*-DNA fingerprints generated in different laboratories and publications, ideally, having access do the DNA patterns together with their respective internal (each lane) or external (each gel) enable us to perform normalisation of the RFLP-IS*6110* patterns appropriate software.[15] However, we only had access to the figures, either in the format of banding patterns or as digitalised patterns; so, we were restricted to perform a qualitative analysis based on the mean results and on conclusions presented in most of the studies.

For reviewing of MIRU-VNTR patterns on the other hand, for each paper we were able to introduce numeric data into a single Excel file simply by reorganising the order of the published 12, 15 and 24 loci presented. The first 12 MIRU loci positions were composed by: MIRU2 (154), MIRU4 (580), MIRU10 (960), MIRU16 (1644), MIRU20 (2059), MIRU23 (2531), MIRU24 (2687), MIRU26 (2996), MIRU27 (3007), MIRU31 (3192), MIRU39 (4348) and MIRU40 (802) was adopted. For the next 12 VNTRs of the 24-MIRU-VNTR patterns, we organised according to the ETR, MTUB and QUB scheme, being ETR-A (2165), ETR-B (2461), ETR-C (577), MTUB 04 (424), MTUB 21 (1955), MTUB 29 (2347), MTUB 30 (2401), MTUB 34 (3171), MTUB 39 (3690), QUB 11 (2163b), QUB 26 (4052) and QUB 4156 (4156). For 15 MIRU-VNTR typing, the same order was adopted but removing nine loci: MIRU2 (154), MIRU20 (2059), MIRU23 (2531), MIRU24 (2687), MIRU27 (3007), MIRU39 (4348), ETR-B (2461), MTUB 04 (424) and MTUB 29 (2347).

Recent transmission was estimated by the N-1 method, using the mathematical model: number of clustered isolates minus (-) number of clusters divided (/) by the total number of isolates.[16] The allelic diversity (*h*) at MIRU-VNTR loci was calculated according to Hunter-Gaston index[17] using Bionumerics (Applied Maths, Sint-Martens-Latem, Belgium) for each set of 12, 15 and 24 loci and defined as "highly discriminant" ($h > 0.6$), "moderately discriminant" ($0.3 \leq h \leq 0.6$) or "poorly discriminant" ($h < 0.3$).

In addition, we have used TBminer (https://info-demo.lirmm.fr/tbminer/)[18] to predict the MTBC lineages from MIRU-VNTR profile.

Disease distribution based on geographic mapping - The boundaries of the regional divisions of Brazil (States and Regions) applied presently were obtained on the website of the Brazilian Institute of Geography and Statistics (IBGE) (https://www.ibge.gov.br/).[19] The coordinates of the institutions were obtained from Google

Maps (https://www.google.com.br/maps). Data processing, interpretation, visualisation, and spatial analysis were performed via ArcGIS software (http://www.arcgis.com/). TB incidence was classified into five levels according to the WHO and being either absence of: no cases (white colour), low (1-10 - green), medium (11-50 - yellow), high (51-100 cases - orange) and very high number (> 100 - red) cases per 100,000 hab.

RESULTS

The Brazilian organisational structure for TB policies - The NPTC is linked to three governmental spheres and coordinated by the so-called Unified Health System (UHS) (Sistema Único de Saúde - SUS) that legally establishes administrative competence at the federal, state, and municipal level. These spheres are composed of the Ministry of Health, the State Health Secretariats (one for each of the 26 states and the Federal District) and the Municipal Health Secretariats, each having their respective technical and administrative sectors.[13]

The National System of Public Health Laboratories (Sistema Nacional de Laboratórios de Saúde Pública - SISLAB) consists of a national network of laboratories, organised in sub-networks in a hierarchical way and with different degrees of complexities of activities related to health surveillance. There are seven laboratory categories[13] as represented in Fig. 1.

The Brazilian Guidelines focus basically on clinical recommendations regarding the standardisation of case finding and treatment actions with little, or no information related to genotyping data generated in Brazilian studies.

For routine TB diagnosis in clinical specimens, besides chest X-ray, collection of sputum samples for acid-fast bacilli staining (Ziehl-Neelsen and/or auramine-rodamine stain) are culture in solid (Lowenstein-Jensen or Ogawa-Kudoh) (AFB) or liquid media (BACTEC MGIT 960) are performed. However, between 2014 and 2015, the Brazilian NPTBC implemented the molecular diagnostics technology Xpert® MTB/RIF (rifampicin) in 92 municipalities with high disease burden.[20] More recently, the identification of MTBC isolates by the rapid immunochromatographic test SD-Bioline TB Ag MPT 64 (Standard Diagnostics, Seoul, South Korea) was implement in Brazil.

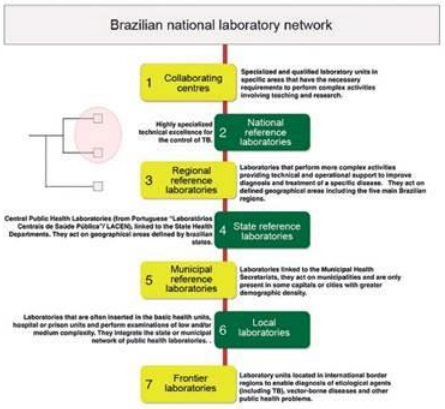


Fig. 1: the Brazilian National System of Public Health Laboratories network classified by degree of complexity highlighting the two levels capable to elaborate a national genetic database for tuberculosis (TB) surveillance.

The phenotypic drug-susceptibility tests (DST) for first line drugs are performed in all State Reference Laboratories named Laboratório Central (LACEN) and are based on the MGIT-960 SIRE kit (MGIT-960; Becton Dickinson Diagnostic Systems, Sparks, MD). At municipality level, the molecular drug-susceptibility test Xpert-Ultra® MTB/RIF (Cepheid, Sunnyvale, EUA) [21] is performed for detection of RIF-R while at the regional reference laboratories, GenoType®MTBDR*plus* and GenoType®MTBDR*sl* (Hain Lifescience GmbH, Nehren, Germany) are used, detecting respectively mutations associated with rifampicin and isoniazid resistance, and mutations associated with fluoroquinolones and second-line injectable drugs.

Currently, the DST for second-line drugs is carried out only in three laboratories in Brazil: at the National Reference Centre (Centro de Referência Professor Hélio Fraga - CRPHF) and at the Laboratório de Bacteriologia e Bioensaios both belonging to the Oswaldo Cruz Foundation (Fundação Oswaldo Cruz - Fiocruz) in Rio de Janeiro, and at the LACEN in São Paulo.

Focusing more on epidemiological surveillance, the Notifiable Diseases Information System for Tuberculosis (SINAN-TB) is the main source for professionals of health surveillance services for data analysis and for planning and monitoring actions towards TB control at the three government levels: federal, state and municipality. A recent study demonstrated the current algorithm used by SINAN-TB, which has a unique identifier per person, integrated with other information systems and built on new technologies, so that TB data transfer and analysis is more streamlined in Brazil. Interestingly, the only results from a molecular diagnostic test that are included in the SINAN-TB are those of the Xpert MTB/RIF assay.[22]

Data analysed - Among a total of 240 manuscripts published between 1996 and 2019, 169 on RFLP-IS*6110* and 71 on MIRU-VNTR, we considered 56 eligible for our study. The BVS database constitutes mostly of PubMed publications along with some duplicated articles within the rest of its own databases. Fig. 2 demonstrates details about the screening process and finally: 17 manuscripts on MIRU-VNTR 17[23-39] and 42 on RFLP-IS*6110* were considered;[23,26,31,32,40-75] three manuscripts considered both methodologies.

In the case of MIRU-VNTR data, some articles were excluded because they only contained information on *M. tuberculosis* var. *bovis* (n = 5);[76,77,78,79,80] no data on genotyping were available (n = 6);[49,66,81,82,83,84] incomplete information on genetic diversity (n = 3) [85,86,87] or genotyping data is mixed with samples from other countries (n = 3).[84,88,89]

Regarding manuscripts on RFLP-IS*6110*, we excluded those with data on *M. bovis* only (n = 4);[90,91,92,93] without visible on genotyping (n = 8)[33,94,95,96] did not present data

low rate of clustering in the *M. tuberculosis* population (0.13 and 0.28, respectively). Additionally, two studies[32,33] investigated isolates from the same patient with up to three different loci and upon further characterisation, such closely related MIRU-VNTR types, demonstrated to belong to the same strain.

The clustering rate according to each MIRU-VNTR set of 12, 15 and 24 loci was 73%, 33% and 28%, respectively (Table II). Among the 24 and 15 loci evaluated, 4052_QUB26, 2163b_QUB11b, 424_Mtub04, 802_MIRU40, 1955_Mtub21, 2696_MIRU26, were considered highly discriminant. Regarding the 24 and 15 loci evaluated, 802_MIRU40, 2696_MIRU26, 2531_MIRU23, were considered highly discriminant and present in 12 and 24 loci analysis, while 802_MIRU40, 2696_MIRU26 and 960_MIRU10 were highly discriminant and commonly present in 12 and 15 loci analysis.

The 802_MIRU40 and 2696_MIRU26 were the highly discriminant among the three dataset and 960_MIRU10 were highly discriminant in 12 and 15 loci and moderately discriminant in 24 loci analysis (Supplementary data II).

The assignation (lineage and/or sublineage) of the strains using TBminer is presented in Supplementary data III. All assignations in green are reliable (at least 2 classifications providing the same result). Lineage 4, mainly Latin-American Mediterranean (LAM) is predominant in all states, but Lineage 1 is mainly isolating from patients in the Pará State, and Lineage 3 is predominantly from Rio Grande do Sul State. Although the observation of the potential concentration *M. bovis* in Goiás, and Lineage 5 (*M. africanum*) in Rio de Janeiro states, there is not a consensus between the two classifications available.

DISCUSSION

This study presents the data on the genetic diversity of *M. tuberculosis* and TB transmission within the new era of global TB monitoring, discussing aspects of TB molecular epidemiology in Brazil previously pointed out in a translational research perspective - "from bench to bedside".[120]

In the light of the Genomic Era, a recent study conducted in England described benefits of TB molecular strain-based cluster investigations (CIs) into a translational approach by identifying new epidemiological links between cases and taking public health action, as well as refuting transmission and saving resources.[121] According to these results, molecular typing is efficient for decreasing transmission and adds value for improving public health in low disease prevalence and high resource setting.

Even though Xpert® MTB/RIF has been implemented in Brazil since 2014, this test does not provide information about MTBC lineages or transmission that could be useful for epidemiological studies and clinical decision-making. Current trends in this direction, point to the use of a new technologies that are able to provide both molecular DST and epidemiological information based on next generation sequencing (NGS) using whole-genome sequencing (WGS).[122]

A recent review[6] showed that between 2009 and 2016, a total of $4.6 billion was invested into TB research, mostly for the development of new diagnostics tools, drugs, and vaccines (61%). Studies on genetic variability are welcome but should go a step further towards translational sciences. Therefore, genotyping tools are important not only to achieve a faster diagnostic and

TABLE I

Summary of restriction fragment length polymorphism-IS*6110* (RFLP-IS*6110*) genotyping publications

	States	No. of studies[a]	%	% by region	No. of IS*6110*	% in cluster
Southeast	SP	11	28.2	74.4	2 to 21	22 to 56
	ES	5	12.5		NR	40 to 48
	RJ	11	28.2		2 to 22	19 to 84[b]
	MG	2	5.1		1 to 18	6 to 25
South	PR	0	0	10.3	-	-
	SC	0	0		-	-
	RS	4	10.3		1 to 18	36 to 69[c]
Central West	GO	2	5.1	10.3	1 to 14	0
	MT	0	0		-	-
	MS	2	5.1		4 to 17	64 to 69[d]
	DF	0	0		-	-
North	TO	0	0	2.6	-	-
	PA	0	0		-	-
	AP	0	0		-	-
	RR	1	2.6		NR	30
	AM	0	0		-	-
	RO	0	0		-	-
	AC	0	0		-	-
Northeast	MA	0	0	2.6	-	-
	PI	0	0		-	-
	CE	0	0		-	-
	RN	0	0		-	-
	PB	0	0		-	-
	PE	0	0		-	-
	AL	0	0		-	-
	SE	0	0		-	-
	BA	1	2.6		2 to 16	27
TOTAL		39	100	100	1 to 22	0 to 84

a: for these calculations, articles that analysed samples from more than one state in the same study were disregarded; *b*: study with inmate population; *c*: tuberculosis multidrug resistant (TB-MDR) population study; *d*: study with indigenous population; NR: not reported.

regarding genetic diversity (n = 9);[24,97,98,99,100,101,102,103] were related to nontuberculous mycobacteria (NTM) (n = 4);[104,105,106,107] were performed in other countries (n = 5);[88,108,109,110,111] did not target IS*6110* for RFLP (n = 7)[112-118] or were data presented as part of a thesis manuscript and had not been peer reviewed (n = 4).

We observed that RFLP-IS*6110* analysis was the first technique to evaluate genetic diversity of *M. tuberculosis* in Brazil over two decades ago, and data using this technique are still being published. This technique has nowadays been substituted almost completely by MIRU-VNTR typing (Supplementary data I).

The geographical map of Brazil with TB incidence, study distribution based on sampling and manuscript authorship is presented by Fig. 3. Fig. 3A demonstrates the spatial location of the country divided in five regions and 27 states demonstrating a considerable difference of incidence per state, with the states of Amazonas (AM - North) and Rio de Janeiro (RJ - Southeast) presenting the highest values (72.9 and 66.3 per 100.000 inhabitants).

Studies using MIRU-VNTR or RFLP-IS*6110* were performed in all regions of Brazil and in 16 (59%) of the states (including the Federal District) but most (n = 40/56) were performed in the Southeast region, including 17 in Rio de Janeiro and 15 in São Paulo states; and in the South, represented by 12 studies from Rio Grande do Sul and three from Paraná states. In the North, three studies were represented from Pará State while the Central West region harbored five studies, all from Goiás State (Fig. 3).

RLFP data-analysis - Among the Brazilian regions, the largest number of studies, using RFLP-IS*6110*, was observed in the Southeast region, with more than 70% of all publications, followed by the South and Central West regions that, together, do not reach 25%. The description of clustering rate and number of IS*6110* copies can be found in Table I, and the Southeast region includes studies comprising all the constituent states with a clustering in the general population ranging from 6 to 56.2%. Among the vulnerable populations studied, transmission rate was observed of 53% among HIV patients (71) and 84% among prisoners.[52]

For São Paulo, the state with the biggest population and economic growth, some specific populations showed high clustering rate, such as patients with resistant TB (52.3%),[46] extensively resistant (52.8%)[54] and prisoners (55.9%).[60] Rio de Janeiro and São Paulo were the municipalities with the highest number of publications using RFLP-IS*6110* (11 each) and demonstrated an increase in the rate of grouping/transmission over time. Among the studies analysed, one study conducted in the central region of São Paulo reported the highest clustering rate (56.2%) and in this study, they sought to verify the impact of migration on the recent transmission of

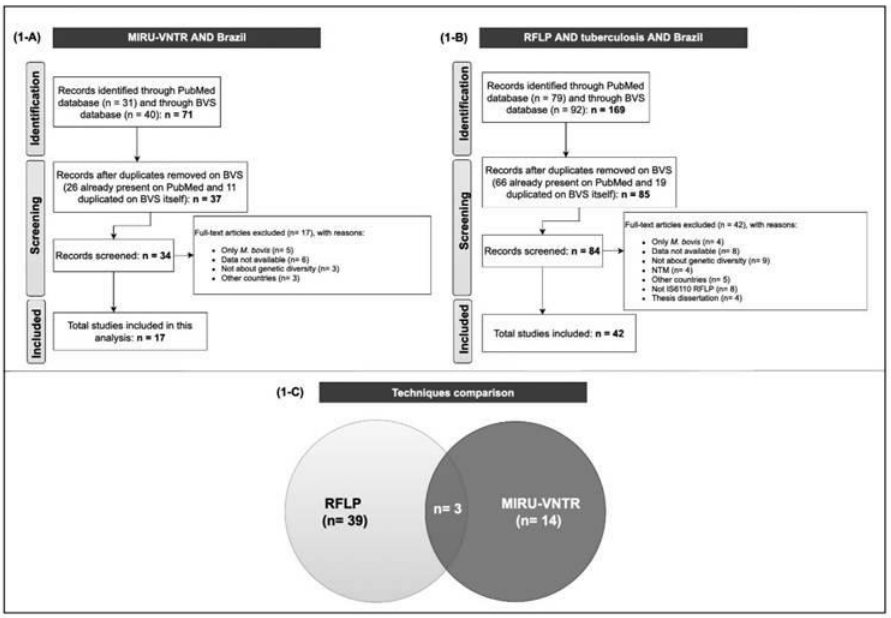
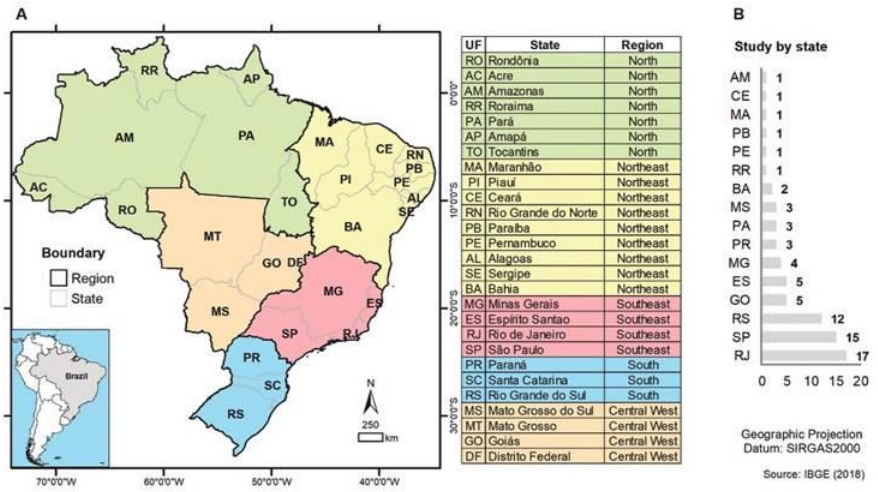


Fig. 2: the PRISMA flow diagram for each genotyping technique demonstrating the total of studies selected for this literature review: 1-A) mycobacterial interspersed repetitive unit-variable variable number tandem repeat (MIRU-VNTR) and 1-B) restriction fragment length polymorphism (RFLP-IS*6110*). 1-C) The 57 studies of *Mycobacterium tuberculosis* genotyping in Brazil and their distribution according to each method.



UF	State	Region
RO	Rondônia	North
AC	Acre	North
AM	Amazonas	North
RR	Roraima	North
PA	Pará	North
AP	Amapá	North
TO	Tocantins	North
MA	Maranhão	Northeast
PI	Piauí	Northeast
CE	Ceará	Northeast
RN	Rio Grande do Norte	Northeast
PB	Paraíba	Northeast
PE	Pernambuco	Northeast
AL	Alagoas	Northeast
SE	Sergipe	Northeast
BA	Bahia	Northeast
MG	Minas Gerais	Southeast
ES	Espírito Santao	Southeast
RJ	Rio de Janeiro	Southeast
SP	São Paulo	Southeast
PR	Paraná	South
SC	Santa Catarina	South
RS	Rio Grande do Sul	South
MS	Mato Grosso do Sul	Central West
MT	Mato Grosso	Central West
GO	Goiás	Central West
DF	Distrito Federal	Central West

Fig. 3: studies distribution based on genotyping by restriction fragment length polymorphism (RFLP-IS*6110*) and mycobacterial interspersed repetitive unit-variable variable number tandem repeat (MIRU-VNTR) in Brazil. (A) Spatial localisation; (B) number of studies by states.

TB. Despite the high percentage found in that study, there seems to be limited contribution of migration in the transmission of TB to Brazilians and vice versa.[50]

For Espírito Santo State, the clustering rate ranged from 40 to 48%; however, as most studies comprised periods of time and/or were overlapping, it was not possible to infer a tendency for increase or decrease in this particular state. However, the incidence of TB in Espirito Santo state seems to be highly influenced by a small set of strains that circulate actively.[53,69]

Data from Minas Gerais State shows that its rate of clustering was the lowest in the Southeast region, ranging between 6.4% and 25.4%, suggesting a low recent rate of TB transmission in this state, including patients with MDR-TB; this might have been influenced however by the low sampling.[26]

Five publications were from the South region, all from the state of Rio Grande do Sul; strains with one to 18 copies of IS*6110* have been reported and the percentage of grouping ranged from 36 to 42.9% for the general population and from 38% to 68.6% for the population with MDR-TB. Despite the small number of publications, studies with MDR-TB patients suggesting an increase of MDR-TB transmission during the studied period, a particular problem of this Southern state and probably associated with HIV infection.

In Mid-West Brazil, more specifically in the Goiás State, polymorphism is observed in TB resistant and TB-MDR strains, suggesting a high rate of primary resistance. Two studies in Mato Grosso State, carried out exclusively upon the indigenous population, demonstrated a high clustering rate of 63.5%[64] and 69%[51] typical for high transmission rates among such populations.

In the Northeast region of the country, the only study we found was that of Silva et al.[119] (evaluating isolates from Bahia State that had been collected between March and June 2008, reporting *M. tuberculosis* with a number of IS*6110* copies ranging between 2 and 16, with a cluster rate of 26.7%.

Similarly, the North region was also represented by a single study conducted in the State of Roraima which borders with Venezuela and Guyana and has an important portion of their TB cases associated with indigenous population, constituting 70% of TB cases (2015-2016) and presenting a clustering rate of 30%.[31] This clustering rate is low when compared to other regions of the country what might be related to a large flow of people, being a border region.

MIRU-VNTR data-analysis - We obtained MIRU-VNTR genotypes from 1,613 MTBC isolates and conducted the analysis using 12, 15 or 24 MIRU-VNTR alleles for constructing genetic patterns. Patterns of 24-MIRU-VNTR were available for 1,041 (64.5%) isolates from all states except from Goiás and Paraná. The data demonstrated that genotypes are not exclusively of a specific state (Fig. 4). Because 24-MIRU-VNTR typing is also considered adequate for phylogenetic analysis, we are puzzled by the bimodal and mostly region-independent grouping within the MST tree with strains from Rio de Janeiro State at the central node (Fig. 4C).

Upon analysis of the number of studies on MIRU-VNTR typing in the country, we observed that the Southeast and South regions present the highest number (82.4%; 14/17) and in general, these studies demonstrated that the MIRU-VNTR present a high discriminatory power. Two of these with the largest sampling[28,29,30] showed a

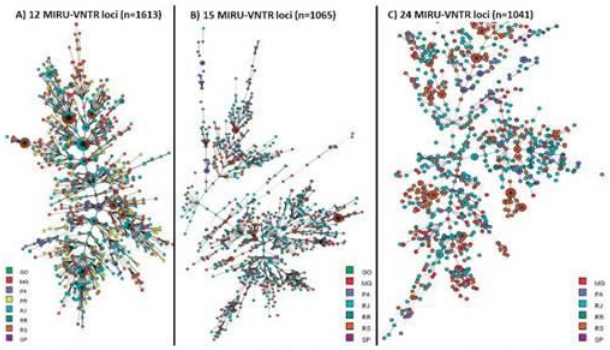


Fig. 4: minimum spanning trees (MST) demonstrating the genetic diversity of *Mycobacterium tuberculosis* in Brazil based on consideration of 12, 15 or 24 mycobacterial interspersed repetitive unit-variable variable number tandem repeat (MIRU-VNTR) alleles and considering a different dataset according to the method's sampling. Samples are coloured according to state origin: Goiás (GO), Minas Gerais (MG), Rio de Janeiro (RJ), Rio Grande do Sul (RS), São Paulo (SP), and Pará (PA).

treatment scheme, but also to be implemented at least at the level of regional reference laboratories as a measure of monitoring and controlling TB.

The Brazilian TB surveillance actions include home visit for new case and summoning of possible cases of TB infection in hospitals and other institutions, as well as follow-up and closure of cases.[13,22] Currently, strain typing (that would be preferably by 24-loci MIRU-VNTR typing) is not part of routine surveillance in any institution in Brazil, only for basic research.

The present study demonstrates that over the last two decades, MIRU-VNTR has been used less for genotyping than the previous gold standard technique (RFLP-IS*6110*), not only because of earlier implementation of the latter (1993 *versus* 2008), but also due to its higher cost. This is of particular importance for a country with a considerable TB incidence such as Brazil (rate of 33.5 per 100,000 inhabitants). However, implementation of such technology to screen only MDR-TB cases would be already a great step ahead.

The use of international databases not only allows local or national genotyping studies, but evaluation of genetic composition of *M. tuberculosis* strains on a larger and even on a global level. In particular for MIRU-VNTR, besides SITIVIT2, there is another international database that allows comparison and classification to the lineage level of local genotypes (MIRU-VNTR*plus* - http://www.miru-vntrplus.org), which has a collection of 186 strains representing the major MTBC lineages. For each strain species, lineage and epidemiologic information is stored together with information regarding the copy numbers of 24 MIRU loci, spoligotyping patterns, regions of difference (RD) profiles, single nucleotide polymorphisms (SNPs), susceptibility data and RFLP-IS*6110* fingerprint images for all isolates.[123,124] However, because this database is limited to the input of genotypes from 500 isolates per analysis and our sampling was composed of 1,613 entries for MIRU-VNTR, we used the Bionumerics v.7.6 software (Applied Maths, Sint-Martens-Latem, Belgium) for analysis.

The limitations of RFLP-IS*6110* are due to the requirement of large amounts of purified DNA in a more complicated methodology with extensive and laborious steps during data analysis for comparison of data generated in different laboratories by considering internal and external molecular weight markers. Comparison of such data requires the use of specialised programs such as Bionumerics and considerable experience on part of the user, for pattern analysis.

Although a considerable number of studies in Brazil performed genotyping by RFLP-IS*6110* have been published, they mostly report on regional patients where a single laboratory analysed the samples without inter-laboratory comparison because of the afore mentioned characteristics of the technique. In addition, no robust international database of RFLP-IS*6110* profiles is accessible. There is one centralised database at the National Institute for Public Health and the Environment (RIVM, Bilthoven, the Netherlands) but only accessible for collaborators.[73,94]

Although this technique has been used to date, MIRU-VNTR analysis has already proved its robustness and equivalence to the results obtained by RFLP- IS*6110*.[125] Moreover, it still has the advantage of allowing the analysis of isolates with fewer copies of IS*6110* which is not considered a good RFLP-IS*6110* method for these cases.

The higher discriminatory power of MIRU-VNTR compared to other genotyping techniques is already widely known,[126] also showing a range of polymorphism, such that loci 10, 23, 26, 31 and 40 have greater discriminatory power than the others. Additionally, its value has been demonstrated for detecting relapse cases, reinfection, and mixed infections.[127]

Comparing the MIRU-VNTR discriminatory power, this study corroborates a recent review evaluating 56 studies (39 from Asia, seven from America, six from Africa, three from Europe and one from a different country),[128] demonstrated that MIRU10, MIRU26, QUB26, MIRU40, QUB11b and Mtub21 was reported to be the loci with the highest discriminatory powers ($h > 0.6$),

14. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol. 2009; 6(7): e1000100.

15. Van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of *Mycobacterium tuberculosis* by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol. 1993; 31(2): 406-9.

16. Glynn JR, Vynnycky E, Fine PEM. Influence of sampling on estimates of clustering and recent transmission of *Mycobacterium tuberculosis* derived from DNA fingerprinting techniques. Am J Epidemiol. 1999; 149(4): 366-71.

17. Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J Clin Microbiol. 1988; 26(11): 2465-6.

18. Azé J, Sola C, Zhang J, Lafosse-Marin F, Yasmin M, Siddiqui R, et al. Genomics and machine learning for taxonomy consensus: the *Mycobacterium tuberculosis* complex paradigm. PLoS One. 2015; 10(7): e0130912.

19. IBGE – Instituto Brasileiro de Geografia e Estatística. Regional divisions of Brazil. The Brazilian Institute of Geography and Statistics [Internet]. 2008 (cited 2021 Jan 06). p. 1. Available from: https://www.ibge.gov.br/en/geosciences/territorial-organization/regional-division/21536-regional-divisions-of-brazil.html?=&t=o-que-e.

20. MS - Ministério da Saúde. Rede de teste rápido para tuberculose no Brasil: primeiro ano da implantação [Internet]. Brasília; 2015 (cited 2021 Jan 06). 63 pp. Available from: http://portalarquivos2.saude.gov.br/images/pdf/2016/janeiro/19/rtr-tb-15jan16-isbn-web.pdf.

21. MS - Ministério da Saúde. OFÍCIO CIRCULAR No. 7/2019/CGDR/.DCCI/SVS/MS - Atualização das recomendações sobre o diagnóstico laboratorial da tuberculose. [Internet]. Brasília; 2019 (cited 2021 Jan 06). Available from: https://central3.to.gov.br/arquivo/472141/.

22. Rocha MS, Bartholomay P, Cavalcante MV, de Medeiros FC, Codenotti SB, Pelissari DM, et al. Sistema de Informação de Agravos de Notificação (Sinan): principais características da notificação e da análise de dados relacionada à tuberculose. Epidemiol Serv Saude [Internet]. 2020 Mar (cited 2020 Dec 11);29(1). Available from: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S2237-96222020000101201&lng=pt&nrm=iso&tlng=pt.

23. Lazzarini LCO, Huard RC, Boechat NL, Gomes HM, Oelemann MC, Kurepina N, et al. Discovery of a novel *Mycobacterium tuberculosis* lineage that is a major cause of tuberculosis in Rio de Janeiro, Brazil. J Clin Microbiol. 2007; 45(12): 3891-902.

24. Noguti EN, Leite CQF, Malaspina AC, Santos ACB, Hirata RDC, Hirata MH, et al. Genotyping of *Mycobacterium tuberculosis* isolates from alow-endemic setting in northwestern state of Paraná in Southern Brazil. Mem Inst Oswaldo Cruz. 2010; 105(6): 779-85.

25. Dalla Costa ER, Vasconcelos SEG, Esteves LS, Gomes HM, Gomes LL, Da Silva PA, et al. Multidrug-resistant *Mycobacterium tuberculosis* of the Latin American mediterranean lineage, wrongly identified as mycobacterium pinnipedii (Spoligotype International Type 863 [SIT863]), causing active tuberculosis in South Brazil. J Clin Microbiol. 2015; 53(12): 3805-11.

26. Dantas NGT, Suffys PN, Carvalho WS, Gomes HM, de Almeida IN, de Assis LJ, et al. Genetic diversity and molecular epidemiology of multidrug-resistant *Mycobacterium tuberculosis* in Minas Gerais State, Brazil. BMC Infect Dis. 2015; 15(1): 1-11.

27. Gomes LL, Vasconcellos SEG, Gomes HM, Elias AR, Rocha AS, Ribeiro SCM, et al. Genetic diversity of the *Mycobacterium tuberculosis* Beijing family in Brazil and Mozambique and relation with infectivity and induction of necrosis in THP-1 cells. Tuberculosis (Edinb). 2015; 95(Suppl. 1): S190-6.

28. Esteves LS, Dalla Costa ER, Vasconcellos SEG, Vargas A, Ferreira Jr SLM, Halon ML, et al. Genetic diversity of *Mycobacterium tuberculosis* isoniazid monoresistant and multidrug-resistant in Rio Grande do Sul, a tuberculosis high-burden state in Brazil. Tuberculosis. 2018; 110: 36-43.

29. Conceição EC, Guimarães AES, Lopes ML, Furlaneto IP, Rodrigues YC, da Conceição ML, et al. Analysis of potential household transmission events of tuberculosis in the city of Belém, Brazil. Tuberculosis. 2018; 113: 125-9.

30. Salvato RS, Schiefelbein S, Barcellos RB, Praetzel BM, Anusca IS, Esteves LS, et al. Molecular characterisation of multidrug-resistant *Mycobacterium tuberculosis* isolates from a high-burden tuberculosis state in Brazil. Epidemiol Infect. 2019; 147: 1-6.

31. Tatara MB, Perdigão J, Viveiros M, Kritski A, da Silva KE, Sacchi FPC, et al. Genetic diversity and molecular epidemiology of *Mycobacterium tuberculosis* in Roraima State, Brazil. Am J Trop Med Hyg. 2019; 101(4): 774-9.

32. Oelemann MC, Gomes HM, Willery E, Possuelo L, Batista Lima KV, Allix-Béguec C, et al. The forest behind the tree: phylogenetic exploration of a dominant *Mycobacterium tuberculosis* strain lineage from a high tuberculosis burden country. PLoS One. 2011; 6(3): 1-10.

33. Alves SLA, Metzker FS, de Araújo-Filho JA, Junqueira-Kipnis AP, Kipnis A. Clinical data and molecular analysis of *Mycobacterium tuberculosis* isolates from drug-resistant tuberculosis patients in Goiás, Brazil. Mem Inst Oswaldo Cruz. 2011; 106(6): 655-61.

34. Valenca MS, da Rocha JZ, Ramis IB, Carrion LL, Madruga C, de Macedo MB, et al. Improving tuberculosis control through the partnership between university and the health system. Rev Soc Bras Med Trop. 2012; 45(4):491-5.

35. Soares RO, de Macedo MB, von Groll A, da Silva PEA. *Mycobacterium tuberculosis* belonging to family LAM and sublineage RDRio: common strains in Southern Brazil for over 10 years. Braz J Microbiol. 2013; 44(4): 1251-5.

36. Pitondo-Silva A, Santos ACB, Jolley KA, Leite CQF, Darini ALC. Comparison of three molecular typing methods to assess genetic diversity for *Mycobacterium tuberculosis*. J Microbiol Methods. 2013; 93(1): 42-8.

37. Ribeiro SCM, Gomes LL, Amaral EP, Andrade MRM, Almeida FM, Rezende AL, et al. *Mycobacterium tuberculosis* strains of the modern sublineage of the beijing family are more likely to display increased virulence than strains of the ancient sublineage. J Clin Microbiol. 2014; 52(7): 2615-24.

38. Machado LNC, Marcondes NR, Leite CQF, Santos ACB, Pavan FR, Baldin VP, et al. First baseline of circulating genotypic lineages of *Mycobacterium tuberculosis* in patients from the brazilian borders with Argentina and Paraguay. PLoS One. 2014; 9(9): 1-9.

39. Vasconcellos SEG, Acosta CC, Gomes LL, Conceição EC, Lima KV, de Araujo MI, 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. 2014; 9(10): 1-14.

40. Filho LA, Kritski AL, Salles CLG, Sardella IG, Silva MG, Fonseca LS, et al. *Mycobacterium tuberculosis* typing: usefulness of DRE-PCR to confirm cross-contamination in the mycobacteriology laboratory of a general reference hospital for AIDS. Int J Tuberc Lung Dis. 2002; 6(2): 150-4.

41. Baptista IMFD, Oelemann MC, Opromolla DVA, Suffys PN. Drug resistance and genotypes of strains of *Mycobacterium tuberculosis* isolated from human immunodeficiency virus-infected and non-infected tuberculosis patients in Bauru, São Paulo, Brazil. Mem Inst Oswaldo Cruz. 2002; 97(8): 1147-52.

TABLE II

Clustering analysis of mycobacterial interspersed repetitive units - variable number tandem repeat (MIRU-VNTR)

Typing methods	n	No. of different patterns	No. of clusters	No. of clusters isolates	No. of unique isolates	% in cluster	Size of clusters
MIRU-VNTR 12 loci	1,613	26	11	40	15	73%	2-7
MIRU-VNTR 15 loci	1,065	821	106	350	715	33%	2-23
MIRU-VNTR 24 loci	1,041	845	91	287	754	28%	2-11

in Brazilian population, we also present MIRU 23 and Mtub04 with high discriminatory power. These eight loci can be considered in studies that need to be faster and less costly. Studies supported by the Brazilian government exploring and describing the MTBC genetic diversity into the five main regions have correlated the emergence of drug resistant-TB to RD^{rio} (LAM sublineage) strains in South and Southeast regions[23,129] and to the T lineage in the North.[130]

Concerning the phylogenetic network, the central position is not meaningful in the context of many diverse isolates. The ancestor that gave rise to all these strains has no good representative today, in this way the centre is highly dependent on the frequency of the samples, and the strains that have by chance quite average values for the genotyped loci. There is little spatial structure in Brazil. This can also be seen in the assignation to lineages as described above. Regarding the lineages, the Brazilian profile observed in this demonstrate the higher frequency of Lineage 4, mainly LAM genotype and the higher frequency of Lineage 1 in Pará State.[130,131,132,133] The presence of *M. tuberculosis* var. *bovis* among human strains was not reported so far, and *M. tuberculosis* var. *africanum* was recently reported as a single isolate from a patient from Pará State.[134]

This study has some limitations, and they are mostly related to data correlation since some articles do not show genotype data (the number of each MIRU-VNTR loci, or RFLP-IS*6110* profile), so they were excluded from the analysis. Some publications repeat data from previous studies without allowing sample identification, so it was not possible to evaluate the real frequency of genotypes isolates per Brazilian state or region.

On December 23rd of 2019, the Brazilian Secretary of Health Surveillance has published the list of approved National and Regional Reference Laboratories for TB and atypical mycobacteria (NTM), aiming at the establishment of the National Network of Public Health Laboratories for the next 5 years. Institutes on the national level are the National Reference Laboratory Professor Hélio Fraga (CRPHF) of the Fundação Oswaldo Cruz. At the regional level are the Regional Reference Laboratories: The Laboratory of Bacteriology and Bioassays of the National Institute of Infectious Diseases Evandro Chagas (INI, FIOCRUZ), the Central Public Health Laboratory of Amazonas (LACEN, AM), the Central Laboratory of Public Health of Espírito Santo (LACEN, ES); and the Central Public Health Laboratory of the Federal District (LACEN, DF).

Regarding the advances on technology evolution, studies have demonstrated that WGS has the greater discriminatory power for epidemiological compared to genotyping methods. For example, to track TB transmission, it was already established that, based on WGS data, a genetic distance from zero to five SNPs separating patient isolates, are present in linked cases such as household contacts; a genetic distance from five to 12 SNPs is for related cases and more than 12 SNPs was defined to classify epidemiologically unrelated cases.[135,136] Besides that, compared to the commercial genotyping methods or Sanger sequencing, analysis based on WGS display a greater panel of mutations associated to drug resistance.

Taking out 23 *M. bovis* genomes, there are few studies in Brazil related to WGS applied to *M. tuberculosis* so far (around 765 genomes): five related to drug resistance characterisation[122,137,138,139,140] and four related to epidemiological approach[29,141,142,143] we did not include them in this current analysis. Such national studies confirm the potential of WGS for molecular epidemiology approach compared to genotyping. In Supplementary data IV there is a list of all published MTBC genomes isolated in Brazil so far, which is the first version of what should become an interactive database of *Mycobacterium* genomes from patients from Brazil, including MTBC, *M. leprae* and NTM presently under construction at http://www.ioc.fiocruz.br/gemibra/. Part of these MTBC genomes are also available at a website http://cplp-tb.ff.ulisboa.pt/, a TB Molecular Epidemiology Database for the Community of Portuguese Speaking Countries (CPLP).[89]

Even through the natural progression towards WGS is going on, applying MIRU-VNTR and creating a national genotyping database for TB surveillance is more feasible, at least for a while, than WGS at the regional and national reference laboratories. However, in parallel, we could give rise to an interactive national database for WGS, focusing on the genetic structure of MTBC in Brazil, for research and for TB surveillance.

Thus, a similar long-term analysis performed in this study could address a better understanding of the TB dynamics in all of Brazil and refocus the attention towards the gold standard of surveillance. This is the same direction that Singapore has taken by demonstrating that there is a large and heterogeneous distribution of MTBC strains. A universal MTBC typing program coupled with enhanced contact investigations may be useful in further understanding the transmission dynamics of TB locally.[144]

In conclusion

Tracing TB cases and their contacts is of vital importance for the control of TB in high burden countries like Brazil. Research on TB genetic diversity and molecular epidemiology in Brazilian territory was more frequent in South and Southeast and it is imperative to reinforce the need of molecular epidemiology surveillance in the central and northern states. This could be achieved by the intensive training of more laboratory professionals and supply of the materials needed to perform the technique. A high but heterogeneous rate of TB transmission was observed in Brazilian regions. This study highlights the importance of including genotypic analysis by MIRU-VNTR in TB surveillance, at least of drug-resistant cases, and of maintaining a hierarchical flow of data between laboratories in the NPCT network. Thus, we propose an implementation of molecular typing techniques for TB transmission detection based initially on MIRU-VNTR towards WGS, as well as the creation of a national database would improve our efforts to decrease the incidence of this challenging disease.

List of abbreviations

AC: Acre; AFB: Acid-fast bacilli; AL: Alagoas; AM: Amazonas; AP: Amapá; BVS: Biblioteca Virtual em Saúde; BA: Bahia; CE: Ceará; DF: Distrito Federal; DR: direct repeat; ES: Espírito Santo; GIS: geographic information system; GO: Goiás; HBCs: high burden countries; HIV: human immunodeficiency virus; LAM: Latin-American; MA: Maranhão; MDR: multidrug resistant; MG: Minas Gerais; MIRU-VNTRs: mycobacterial interspersed repetitive units-variable tandem repeats of DNA tandem repeats; MS: Mato Grosso do Sul; MT: Mato Grosso; MTBC: *Mycobacterium tuberculosis* complex; MST: minimum spanning tree; PA: Pará; PB: Paraíba; PE: Pernambuco; PI: Piauí; PR: Paraná; RFLP: restriction fragment length polymorphism; RJ: Rio de Janeiro; RN: Rio Grande do Norte; RO: Rondônia; RR: Roraima; RS: Rio Grande do Sul; SC: Santa Catarina; SE: Sergipe; SDG: sustainable development goals; SISLAB: Sistema Nacional de Laboratórios de Saúde Pública; SP: São Paulo; SUS: Sistema Único de Saúde; TB: Tuberculosis; TO: Tocantins; UHS: unified health system; WHO: World Health Organization; XDR: extensive drug resistant.

ACKNOWLEDGEMENTS

To all authors who shared their published data in Excel file when they were available only in figures or in text, helping us to speed up the analysis. We are also thankful to Henri Berghs for his great technical assistance with Bionumerics tools; and to Daniel and Edinaldo from DETIN/IOC/FIOCRUZ for technical assistance by hosting the website.

AUTHORS' CONTRIBUTION

ECC conducted project conception and study designing, literature search, reviewed literature, data collection, data analysis, website conception, and was a major contributor in writing the manuscript; RSS, KMG, AESG, MLC, LMPA have performed the literature review, data-analysis, and manuscript edition; RJPSG performed the spatial analysis; IPF performed the statistical analysis; RBB provided epidemiological information from Brazilian database; AS contributed with literature search, writing the manuscript, and performed English review; MCS reviewed the manuscript writing and data-analysis; CVN, LF, MCSL and GR has contributed with their experience through manuscript edition and information regarding molecular techniques; VRB contributed writing the manuscript regarding his experience as a physician and the application of genotyping of MTBC; ACB contributed with information about the Brazilian structure of *Tuberculosis* program; MC developed the website for the whole-genome sequencing database; PNS, HMG, RSD and KVBL were the supervisors. All authors read and approved the final manuscript. The authors declare that they have no competing interests.

REFERENCES

1. WHO - World Health Organization. Global tuberculosis report 2018. Geneva [Internet]; 2018 (cited 2021 Jan 10). Available from: http://apps.who.int/iris/bitstream/handle/10665/274453/9789241565646-eng.pdf?ua=1.
2. Benavente ED, Coll F, Furnham N, McNerney R, Glynn JR, Campino S, et al. PhyTB: phylogenetic tree visualisation and sample positioning for *M. tuberculosis*. BMC Bioinformatics. 2015; 16(1): 155.
3. WHO - World Health Organization. The end TB strategy. Geneva [Internet]; 2015 (cited 2021 Jan 10). Available from: http://www.who.int/tb/End_TB_brochure.pdf?ua=1.
4. Uplekar M, Weil D, Lonnroth K, Jaramillo E, Lienhardt C, Dias HM, et al. WHO's new end TB strategy. Lancet. 2015; 385(9979): 1799-801.
5. Floyd K, Glaziou P, Houben RMGJ, Sumner T, White RG, Raviglione M. Global tuberculosis targets and milestones set for 2016-2035: definition and rationale. Int J Tuberc Lung Dis. 2018; 22(7): 723-30.
6. Floyd K, Glaziou P, Zumla A, Raviglione M. The global tuberculosis epidemic and progress in care, prevention, and research: an overview in year 3 of the end TB era. Lancet Respir Med. 2018; 6(4): 299-314.
7. Kritski A, Barreira D, Junqueira-Kipnis AP, Moraes MO, Campos MM, Degrave WM, et al. Brazilian response to global end TB strategy: the National Tuberculosis Research Agenda. Rev Soc Bras Med Trop. 2016; 9(1): 135-45.
8. Kritski A, Dalcolmo MP, Mello FCQ, Carvalho ACC, Silva DR, de Oliveira MM, et al. The role of the Brazilian tuberculosis research network in national and international efforts to eliminate tuberculosis. J Bras Pneumol. 2018; 44(2): 77-81.
9. Asgharzadeh M, Rashedi J, Poor BM, kafil HS, Zadeh HM, Ahmadpour E. How molecular epidemiology can affect tuberculosis control in the Middle East countries: a systematic review and meta-analysis. Infect Disord - Drug Targets. 2020; 21: 28-37.
10. Couvin D, Rastogi N. The establishment of databases on circulating genotypes of *Mycobacterium tuberculosis* complex and web tools for an effective response to better monitor, understand and control the tuberculosis epidemic worldwide. EuroReference - J Ref. 2014; 12: 36-48.
11. Coscolla M. Biological and epidemiological consequences of MTBC diversity. Adv Exp Med Biol. 2017; 1019: 95-116.
12. MS - Ministério da Saúde. NOTA TÉCNICA No. 31/2019-CGDR/. DCCI/SVS/MS. Brasília [Internet]; 2019 (cited 2020 Dec 09). Available from: http://www.saude.pi.gov.br/uploads/warning_document/file/450/Nota_Técnica_-Regulação.pdf.
13. MS/SVS - Ministério da Saúde/Secretaria de Vigilância em Saúde. Manual de recomendações para o controle da tuberculose no Brasil. 2nd ed. Brasília: Ministério da Saúde/Secretaria de Vigilância em Saúde/Departamento de Vigilância de Doenças Transmissíveis [Internet]; 2011 (cited 2020 Dec 09). 344 p. Available from: https://drive.google.com/file/d/12Sm_BZhR4Cuk5X8YPzgzUXT-G38B5_LA7/view.

72. Ferreira MM, Ferrazoli L, Palaci M, Salles PS, Medeiros LA, Novoa P, et al. Tuberculosis and HIV infection among female inmates in São Paulo, Brazil: a prospective cohort study. J Acquir Immune Defic Syndr Hum Retrovirol. 1996; 13(2): 177-83.

73. Suffys PN, de Araujo MEI, Rossetti ML, Zaha A, Barroso EW, Barreto AMW, et al. Usefulness of IS6110-restriction fragment length polymorphism typing of Brazilian strains of *Mycobacterium tuberculosis* and comparison with an international fingerprint database. Res Microbiol. 2000; 151(5): 343-51.

74. Ferrazoli L, Palaci M, Marques LR, Jamal LF, Afiune JB, Chimara E, et al. Transmission of tuberculosis in an endemic urban setting in Brazil. Int J Tuberc Lung Dis. 2000; 4(1): 18-25.

75. Lourenço MCS, Grinsztejn B, Fandinho-Montes FCO, Silva MG, Saad MHF, Fonseca LS. Genotypic patterns of multiple isolates of *M. tuberculosis* from tuberculous HIV patients. Trop Med Int Health. 2000; 5(7): 488-94.

76. Parreiras PM, Andrade GI, do Nascimento TF, Oelemann MC, Gomes HM, de Alencar AP, et al. Spoligotyping and variable number tandem repeat analysis of *Mycobacterium bovis* isolates from cattle in Brazil. Mem Inst Oswaldo Cruz. 2012; 107(1): 64-73.

77. Ramos DF, Tavares L, da Silva PEA, Dellagostin OA. Molecular typing of *Mycobacterium bovis* isolates: a review. Braz J Microbiol. 2014; 372(2): 365-72.

78. Carvalho RCT, Vasconcellos SEG, Issa MA, Soares Filho PM, Mota PMPC, de Araújo FR, et al. Molecular typing of *Mycobacterium bovis* from cattle reared in Midwest Brazil. PLoS One. 2016; 11(9): e0162459.

79. de Souza-Filho AF, Osório ALAR, Jorge KSG, Araújo FR, Vidal CES, Araújo CP, et al. Genetic profiles of *Mycobacterium bovis* from a cattle herd in southernmost Brazil. Semin Agrar. 2016; 37(5): 3719-26.

80. Furlaneto IP, da Conceição ML, Conceição EC, Lopes ML, Rodrigues YC, Macelino BR, et al. Molecular epidemiology of mycobacteria among herds in Marajó Island, Brazil, reveals strains genetically related and potential zoonotic risk of clinical relevance. Infect Genet Evol. 2020; 1(77): 1-7.

81. Lopes JS, Marques I, Soares P, Nebenzahl-Guimaraes H, Costa J, Miranda A, et al. SNP typing reveals similarity in *Mycobacterium tuberculosis* genetic diversity between Portugal and Northeast Brazil. Infect Genet Evol. 2013; 18: 238-46.

82. Reynaud Y, Millet J, Rastogi N, Van Embden J, Cave M, Crawford J, et al. Genetic structuration, demography and evolutionary history of *Mycobacterium tuberculosis LAM9* sublineage in the Americas as two distinct subpopulations revealed by Bayesian analyses. PLoS One. 2015; 10(10): 1-15.

83. Pontual Y, Pacheco VSS, Monteiro SP, Quintana MSB, Costa MJM, Rolla VC, et al. *ABCB1* gene polymorphism associated with clinical factors can predict drug-resistant tuberculosis. Clin Sci. 2017; 131(15): 1831-40.

84. de Almeida AL, Scodro RBL, de Carvalho HC, Costacurta GF, Baldin VP, Santos NCS, et al. RD Rio *Mycobacterium tuberculosis* lineage in the Brazil/Paraguay/Argentina triple border. Tuberculosis. 2018; 110: 68-78.

85. Franco MMJ, Ribeiro MG, Pavan FR, Miyata M, Heinemann MB, de Souza Filho AF, et al. Genotyping and rifampicin and isoniazid resistance in *Mycobacterium bovis* strains isolated from the lymph nodes of slaughtered cattle. Tuberculosis. 2017; 104: 30-7.

86. Reynaud Y, Zheng C, Wu G, Sun Q, Rastogi N. Bayesian population structure analysis reveals presence of phylogeographically specific sublineages within previously illdefined T group of *Mycobacterium tuberculosis*. PLoS One. 2017; 12(2): 1-131.

87. Ribeiro MG, Lima MCFF, Franco MMJJ, Megid J, Soares LM, Machado LHAA, et al. Pre-multidrug-resistant *Mycobacterium tuberculosis* infection causing fatal enteric disease in a dog from a family with history of human tuberculosis. Trans Emerg Dis. 2017; 64(5): 1-7.

88. Acosta CCD, Russomando G, Candia N, Ritacco V, Vasconcellos SEG, Moreira MBP, et al. Exploring the "Latin American Mediterranean" family and the RDRio lineage in *Mycobacterium tuberculosis* isolates from Paraguay, Argentina and Venezuela. BMC Microbiol. 2019; 19(1): 131-4.

89. Perdigão J, Silva C, Diniz J, Pereira C, Machado D, Ramos J, et al. Clonal expansion across the seas as seen through CPLP-TB database: a joint effort in cataloguing *Mycobacterium tuberculosis* genetic diversity in Portuguese-speaking countries. Infect Genet Evol. 2018; 72: 44-58.

90. Fisanotti JC, Alito A, Bigi F, Latini O, Roxo E, Cicuta E, et al. Molecular epidemiology of *Mycobacterium bovis* isolates from South America. Vet Microbiol. 1998; 60(2-4): 251-7.

91. Zanini MS, Moreira EC, Lopes MTP, Oliveira RS, Leão SC, Fioravanti RL, et al. *Mycobacterium bovis*: polymerase chain reaction identification in bovine lymphonode biopsies and genotyping in isolates from Southeast Brazil by spolygotyping and restriction fragment length polymorphism. Mem Inst Oswaldo Cruz. 2001; 96(6): 809-13.

92. Zanini MS, Moreira EC, Salas CE, Lopes MTP, Barouni AS, Roxo E, et al. Molecular typing of *Mycobacterium bovis* isolates from south-east Brazil by spoligotyping and RFLP. J Vet Med B Infect Dis Vet Public Health. 2005; 52(3): 129-33.

93. Barouni AS, Augusto C, Queiroz MVNP, Lopes MTP, Zanini MS, Salas CE. BCG lymphadenopathy detected in a BCG-vaccinated infant. Braz J Med Biol Res. 2004; 37(5): 697-700.

94. Suffys PN, Rocha AS, de Oliveira M, Campos CED, Barreto AMW, Portaels F, et al. Rapid identification of *Mycobacteria* to the species level using INNO-LiPA *Mycobacteria*, a reverse hybridization assay. J Clin Microbiol. 2001; 39(12): 4477-82.

95. Shaw MA, Atkinson S, Dockrell H, Hussain R, Lins-Lainson Z, Shaw J, et al. An RFLP map for 2q33-q37 from multicase mycobacterial and leishmanial disease families: no evidence for an Lsh/Ity/Bcg gene homologue influencing susceptibility to leprosy. Ann Hum Genet. 1993; 57(4): 251-71.

96. Clemente WT, Lima SSS, Palaci M, Silva MSN, Rodrigues VFS, Dalla Costa ER, et al. Phenotypic and genotypic characterization of drug-resistant *Mycobacterium tuberculosis* strains. Diagn Microbiol Infect Dis. 2008; 62(2): 199-204.

97. Aires RS, Matos MAD, Lopes CLR, Teles SA, Kozlowski AG, Silva AMC, et al. Prevalence of hepatitis B virus infection among tuberculosis patients with or without HIV in Goiânia City, Brazil. J Clin Virol. 2012; 54(4): 327-31.

98. Rocha AS, Leite CC, Torres HM, de Miranda AB, Lopes MQP, Degrave WM, et al. Use of PCR-restriction fragment length polymorphism analysis of the hsp65 gene for rapid identification of mycobacteria in Brazil. J Microbiol Methods. 1999; 37(3): 223-9.

99. Zumarraga MJ, Martin C, Samper S, Alito A, Latini O, Bigi F, et al. Usefulness of Spoligotyping in molecular epidemiology of *Mycobacterium bovis*-related infections in South America. J Clin Microbiol. 1999; 37(2): 296-303.

100. Ribeiro FKC, Lemos EM, Hadad DJ, Leão SC, Viana-Niero C, Dietze R, et al. Evaluation of low-colony-number counts of *Mycobacterium tuberculosis* on solid media as a microbiological marker of cross-contamination. J Clin Microbiol. 2009; 47(6): 1950-2.

101. Huber FD, Sánchez A, Gomes HM, Vasconcellos S, Massari V, Barreto A, et al. Insights into the population structure of *Mycobacterium tuberculosis* using spoligotyping and RDRio in a southeastern Brazilian prison unit. Infect Genet Evol. 2014; 26: 194-202.

42. Borsuk S, Dellagostin MM, Madeira SG, Lima C, Boffo M, Mattos I, et al. Molecular characterization of *Mycobacterium tuberculosis* isolates in a region of Brazil with a high incidence of tuberculosis. Microbes Infect. 2005; 7(13): 1338-44.

43. Cafrune PI, Riley LWW, Possuelo LGG, Valim ARMRM, Borges M, Ribeiro MOO, et al. Recent transmission of tuberculosis involving retired patients. J Infect. 2006; 53(6): 370-6.

44. Calusni ALR, Roscani GN, Villares MCB, Soini H, Graviss EA, Ramos MC. IS*6110* restriction fragment length polymorphism of *Mycobacterium tuberculosis* isolated from patients with pulmonary tuberculosis in Campinas, Brazil. Evidence of intercontinental distribution of strains. Mem Inst Oswaldo Cruz. 2003; 98(5): 655-8.

45. Resende MR, Villares MCB, Ramos MC. Transmission of tuberculosis among patients with human immunodeficiency virus at a University Hospital in Brazil. Infect Control Hosp Epidemiol. 2004; 25(12): 1115-7.

46. Telles MAS, Ferrazoli L, Waldman EA, Giampaglia CMS, Martins MC, Ueki SYM, et al. A population-based study of drug resistance and transmission of tuberculosis in an urban community. Int J Tuberc Lung Dis. 2005; 9(9): 970-6.

47. Valim ARDM, Possuelo LG, Cafrune PI, Borges M, Ribeiro MO, Rossetti MLR, et al. Evaluation and genotyping of multidrug-resistant cases of tuberculosis in southern Brazil. Microb Drug Resist. 2006; 12(3): 186-91.

48. Andrade MKN, Machado SMA, Leite ML, Saad MHF. Phenotypic and genotypic variant of MDR-*Mycobacterium tuberculosis* multiple isolates in the same tuberculosis episode, Rio de Janeiro, Brazil. Braz J Med Biol Res. 2009; 42(5): 433-7.

49. Augusto CJ, Carvalho WS, de Almeida IN, Figueiredo LJA, Dantas NGT, Suffys PN, et al. Comparative study of RFLP-IS6110 and MIRU-VNTR from *Mycobacterium tuberculosis* isolated in the state of Minas Gerais, Brazil. Braz J Microbiol. 2018; 49(3): 641-6.

50. Pescarini JM, Simonsen V, Ferrazoli L, Rodrigues LC, Oliveira RS, Waldman EA, et al. Migration and tuberculosis transmission in a middle-income country: a cross-sectional study in a central area of São Paulo, Brazil. BMC Med. 2018; 16(1): 1-10.

51. Sacchi FPC, Tatara MB, De Lima CC, Da Silva LF, Cunha EA, Simonsen V, et al. Genetic clustering of tuberculosis in an indigenous community of Brazil. Am J Trop Med Hyg. 2018; 98(2): 372-5.

52. Sánchez A, Huber FD, Massari V, Barreto A, Camacho LAB, Cesconi V, et al. Extensive *Mycobacterium tuberculosis* circulation in a highly endemic prison and the need for urgent environmental interventions. Epidemiol Infect. 2012; 140(10): 1853-61.

53. Peres RL, Vinhas SA, Ribeiro FKC, Palaci M, do Prado TN, Reis-Santos B, et al. Risk factors associated with cluster size of *Mycobacterium tuberculosis* (Mtb) of different RFLP lineages in Brazil. BMC Infect Dis. 2018; 18: 1-10.

54. Gallo JF, Pinhata JMW, Simonsen V, Galesi VMN, Ferrazoli L, Oliveira RS. Prevalence, associated factors, outcomes and transmission of extensively drug-resistant tuberculosis among multidrug-resistant tuberculosis patients in São Paulo, Brazil: a cross-sectional study. Clin Microbiol Infect . 2018; 24(8): 889-95.

55. da Costa HCG, Malaspina AC, de Mello FAF, Leite CQF. Tuberculosis in a psychiatric hospital in the state of Goiás, Brazil. J Bras Pneumol. 2019; 32(6): 566-72.

56. Gibson AL, Huard RC, van Pittius NCG, Lazzarini LCO, Driscoll J, Kurepina N, et al. Application of sensitive and specific molecular methods to uncover global dissemination of the major RDRio Sublineage of the Latin American-Mediterranean *Mycobacterium tuberculosis* spoligotype family. J Clin Microbiol. 2008; 46(4): 1259-67.

57. Malaspina AC, Cavalcanti HR, Leite CQF, Machado SMA, Viana BHJ, Silva RMG, et al. Usefulness of *Mycobacterium tuberculosis* molecular typing in a tuberculosis low-endemic agro-industrial setting of Brazil. Jpn J Infect Dis. 2008; 61(3): 231-3.

58. Mendes JM, Machado SMA, Lourenço MC, Ferreira RMC, Fonseca LS, Saad MHF. Molecular diversity of *Mycobacterium tuberculosis* strains in a slum area of Rio de Janeiro, Brazil. J Bras Pneumol. 2008; 34(12): 1063-8.

59. Inumaru VTG, Nogueira PA, Butuem IV, Riley LW, Ferrazoli L. Reactive nitrogen intermediate susceptibility of *Mycobacterium tuberculosis* genotypes in an urban setting. Int J Tuberc Lung Dis. 2009; 13(5): 665-8.

60. Moreira-Oliveira MS, Oliveira HB, Pace F, Stehling EG, Rocha MMM, Aily DCG, et al. Molecular genotyping and epidemiology of *Mycobacterium tuberculosis* isolates obtained from inmates of correctional institutions of Campinas, Southeast Brazil. Braz J Infect Dis. 2008; 12(6): 487-93.

61. Perizzolo PF, Dalla Costa ER, Ribeiro AW, Spies FS, Ribeiro MO, Dias CF, et al. Characteristics of multidrug-resistant *Mycobacterium tuberculosis* in southern Brazil. Tuberculosis (Edinb). 2012; 92(1): 56-9.

62. Vinhas SA, Palaci M, Marques HS, Lobo de Aguiar PP, Ribeiro FK, Peres RL, et al. *Mycobacterium tuberculosis* DNA fingerprint clusters and its relationship with RD(Rio) genotype in Brazil. Tuberculosis. 2013; 93(2): 207-12.

63. Vinhas SA, Jones-López EC, Rodrigues RR, Gaeddert M, Peres RL, Marques-Rodrigues P, et al. Strains of *Mycobacterium tuberculosis* transmitting infection in Brazilian households and those associated with community transmission of tuberculosis. Tuberculosis. 2017; 104: 79-86.

64. Cunha EA, Ferrazoli L, Riley LW, Basta PC, Honer MR, Maia R, et al. Incidence and transmission patterns of tuberculosis among indigenous populations in Brazil. Mem Inst Oswaldo Cruz. 2014;109(1): 108-13.

65. Eyer-Silva WA, de Almeida MR, Martins CJ, Basílio-de-Oliveira RP, de Araujo LF, Basílio-de-Oliveira CA, et al. Antiretroviral therapy-induced paradoxical worsening of previously healed *Mycobacterium haemophilum* cutaneous lesions in advanced HIV infection. 2019; 61: 1-6.

66. Pereira AM, Santos LC, Fernandes HB, Alves SLA, Junqueira-Kipnis AP, Kipnis A. Análise molecular de *Mycobacterium tuberculosis* isolados de pacientes atendidos em Goiânia, Goiás, por meio do rflp-is6110 e do 15 loci miru-vntr. Rev Patol Trop. 2013; 42(3): 275-88.

67. Santos LC, Bousquet HM, Pereira AM, Junqueira-Kipnis AP, Kipnis A. A high prevalence of resistance in new tuberculosis cases of midwestern Brazil. Infect Genet Evol. 2010; 10(7): 1052-7.

68. Gomes T, Vinhas SA, Reis-Santos B, Palaci M, Peres RL, Aguiar PP, et al. Extrapulmonary tuberculosis: *Mycobacterium tuberculosis* strains and host risk factors in a large urban setting in Brazil. PLoS One. 2013; 8(10): 1-10.

69. Ribeiro FKC, Pan W, Bertolde A, Vinhas SA, Peres RL, Riley L, et al. Genotypic and spatial analysis of *Mycobacterium tuberculosis* transmission in a high-incidence urban setting. Clin Infect Dis. 2015; 61(5): 758-66.

70. Fandinho FC, Kritski AL, Hofer C, Júnior Conde H, Ferreira RM, Saad MH, et al. RFLP patterns and risk factors for recent tuberculosis transmission among hospitalized tuberculosis patients in Rio de Janeiro, Brazil. Trans R Soc Trop Med Hyg. 2000; 94(3): 271-5.

71. Ivens-de-Araujo ME, Fandinho FC, Werneck-Barreto AM, Gonçalves-Veloso V, Grinstejn B, Lourenço MK, et al. DNA fingerprinting of *Mycobacterium tuberculosis* from patients with and without AIDS in Rio de Janeiro. Braz J Med Biol Res. 1998; 31(3): 369-72.

102. Prim RI, Schörner MA, Senna SG, Nogueira CL, Figueiredo ACC, de Oliveira JG, et al. Molecular profiling of drug resistant isolates of *Mycobacterium tuberculosis* in the state of Santa Catarina, southern Brazil. Mem Inst Oswaldo Cruz. 2015; 110(5): 618-23.

103. Santos LC, Kipnis APJ, Kipnis A. Métodos aplicados à epidemiologia molecular do *Mycobacterium tuberculosis* TT - Methods employed in the molecular epidemiology of *Mycobacterium tuberculosis*. J Trop Pathol. 2007; 36(1).

104. Oliveira EMD, Morais ZM, Tabata R, Dias RA, de Oliveira RS, Leão SC, et al. Avaliação da virulência em hamsters (Mesocricetus auratus) de estirpes de *Mycobacterium avium* presentes na população de suínos do Sul do Brasil. Braz J Vet Res Anim Sci. 2002; 39(4): 202-7.

105. Chimara E, Giampaglia CMS, Martins MC, Telles MAS, Ueki SYM, Ferrazoli L. Molecular characterization of *Mycobacterium kansasii* isolates in the State of São Paulo between 1995-1998. Mem Inst Oswaldo Cruz. 2004; 99(7): 739-43.

106. Telles MAS, Chimara E, Ferrazoli L, Riley LW. *Mycobacterium kansasii*: antibiotic susceptibility and PCR-restriction analysis of clinical isolates. J Med Microbiol. 2005; 54(10): 975-9.

107. Candido PHC, Nunes LS, Marques EA, Folescu TW, Coelho FS, de Moura VCN, et al. Multidrug-resistant nontuberculous mycobacteria isolated from cystic fibrosis patients. J Clin Microbiol. 2014; 52(8): 2990-7.

108. Shemiakin I, Stepanshina V, Korobova O, Anisimova V, Ivanov I, Lipin M, et al. Genetic typing of *Mycobacterium tuberculosis* strains by spoligotyping and genome fingerprinting techniques. Zh Mikrobiol Epidemiol Immunobiol. 2002; 6: 30-5.

109. Vasconcellos SEG, Huard RC, Niemann S, Kremer K, Santos AR, Suffys PN, et al. Distinct genotypic profiles of the two major clades of *Mycobacterium africanum*. *BMC* Infect Dis. 2010; 10(1): 1-16.

110. Ritacco V, Iglesias M-J, Ferrazoli L, Monteserin J, Dalla Costa ER, Cebollada A, et al. Conspicuous multidrug-resistant *Mycobacterium tuberculosis* cluster strains do not trespass country borders in Latin America and Spain. Infect Genet Evol. 2012; 12(4): 711-7.

111. Weisenberg SA, Gibson AL, Huard RC, Kurepina N, Bang H, Lazzarini LCO, et al. Distinct clinical and epidemiological features of tuberculosis in New York City caused by the RD(Rio) *Mycobacterium tuberculosis* sublineage. Infect Genet Evol. 2012; 12(4): 664-70.

112. Chimara E, Ferrazoli L, Leão SC. *Mycobacterium tuberculosis* complex differentiation using gyrB-restriction fragment length polymorphism analysis. Mem Inst Oswaldo Cruz. 2004; 99(7): 745-8.

113. Marchi AM, Juttel ID, Kawacubo EM, Dalmarco EM, Blatt SL, Cordova CMM. Evaluation of methods for detection and identification of *Mycobacterium* species in patients suspected of having pulmonary tuberculosis. Braz J Microbiol. 2008; 39(4): 613-8.

114. Malaghini M, Brockelt SR, Burger M, Kritski A, Thomaz-Soccol V. Molecular characterization of *Mycobacterium tuberculosis* isolated in the State of Paraná in southern Brazil. Tuberculosis. 2009; 89(1): 101-5.

115. Forestiero FJ, Cecon L, Hirata MH, de Melo FF, Cardoso RF, Cerda A, et al. Relationship of NAT2, CYP2E1 and GSTM1/GSTT1 polymorphisms with mild elevation of liver enzymes in Brazilian individuals under anti-tuberculosis drug therapy. Clin Chim Acta . 2013; 415: 215-9.

116. Riello FN, Brígido RTS, Araújo S, Moreira TA, Goulart LR, Goulart IMB. Diagnosis of mycobacterial infections based on acid-fast bacilli test and bacterial growth time and implications on treatment and disease outcome. BMC Infect Dis. 2016; 16(1): 1-11.

117. Salum KCR, de Castro MCS, Moreira VB, Nani ASF, Kohlrausch FB. Interleukin 1α and 1β gene variations are associated with tuberculosis in silica exposed subjects. Am J Ind Med. 2019; 63(1): 74-84.

118. Santos EA, Gonçalves JS, Fleury MK, Kritski AL, Oliveira MM, Velasque LS, et al. Relationship of anti-tuberculosis drug-induced liver injury and genetic polymorphisms in CYP2E1 AND GST. Braz J Infect Dis. 2019; 23(6): 381-7.

119. Silva ACN, Ferrazoli L, Simonsen V, Reis JN, Pereira SM, Carmo TMA, et al. First insights into the genetic diversity of *Mycobacterium tuberculosis* strains in Salvador, Bahia State, Brazil. Cad Saude Publica. 2011; 27(9): 1859-63.

120. Conceição EC, Lima KVB, Gomes HM, Duarte RS. Molecular epidemiology of tuberculosis in Brazil: a translational research perspective. Rev Panamazonica Saude. 2017; 8(4): 11-3.

121. Hamblion EL, Burkitt A, Lalor MK, Anderson LF, Thomas HL, Abubakar I, et al. Public health outcome of Tuberculosis Cluster Investigations, England 2010-2013. J Infect. 2019; 78(4): 269-74.

122. Feliciano CS, Namburete EI, Plaça JR, Peronni K, Dippenaar A, Warren RM, et al. Accuracy of whole genome sequencing versus phenotypic (MGIT) and commercial molecular tests for detection of drug-resistant *Mycobacterium tuberculosis* isolated from patients in Brazil and Mozambique. Tuberculosis. 2018; 110: 59-67.

123. Allix-Béguec C, Harmsen D, Weniger T, Supply P, Niemann S. Evaluation and strategy for use of MIRU-VNTRplus, a multifunctional database for online analysis of genotyping data and phylogenetic identification of *Mycobacterium tuberculosis* complex isolates. J Clin Microbiol. 2008; 46(8): 2692-9.

124. Weniger T, Krawczyk J, Supply P, Niemann S, Harmsen D. MIRU-VNTRplus: a web tool for polyphasic genotyping of *Mycobacterium tuberculosis* complex bacteria. Nucleic Acids Res. 2010; 38: 326-31.

125. De Beer JL, Van Ingen J, De Vries G, Erkens C, Sebek M, Mulder A, et al. Comparative study of is6110 restriction fragment length polymorphism and variable-number tandem-repeat typing of *Mycobacterium tuberculosis* isolates in the netherlands, based on a 5-year nationwide survey. J Clin Microbiol. 2013; 51(4): 1193-8.

126. Sola C, Filliol I, Legrand E, Lesjean S, Locht C, Supply P, et al. Genotyping of the *Mycobacterium tuberculosis* complex using MIRUs: association with VNTR and spoligotyping for molecular epidemiology and evolutionary genetics. Infect Genet Evol. 2003; 3(2): 125-33.

127. Lymbery AJ, Thompson RCA. The molecular epidemiology of parasite infections: Tools and applications. Mol Biochem Parasitol. 2012; 181(2): 102-16.

128. Ghavidel M, Tadayon K, Mosavari N, Nourian K, Taghanaki HRB, Mohammadi GR, et al. Introducing the best six loci in Mycobacterial Interspersed Repetitive Unit-Variable-Number Tandem Repeat (MIRU-VNTR) typing for *Mycobacterium tuberculosis* genotyping. Rep Biochem Mol Biol. 2019; 8(3): 335-46.

129. Costa ERD, Lazzarini LCO, Perizzolo PF, Díaz CA, Spies FS, Costa LL, et al. *Mycobacterium tuberculosis* of the RDRio genotype is the predominant cause of tuberculosis and associated with multidrug resistance in Porto Alegre City, South Brazil. J Clin Microbiol. 2013; 51(4): 1071-7.

130. Conceição ECEC, Rastogi N, Couvin D, Lopes MLML, Furlaneto IPIP, Gomes HMHM, 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: 62-74.

131. Woodman M, Haeusler IL, Grandjean L. Tuberculosis genetic epidemiology: a latin american perspective. Genes. 2019; 10(1): 1-17.

132. Duarte TA, Nery JS, Boechat N, Pereira SM, Simonsen V, Oliveira M, et al. A systematic review of East African-Indian family of *Mycobacterium tuberculosis* in Brazil. Braz J Infect Dis. 2017; 21(3): 317-24.

133. Brynildsrud OB, Pepperell CS, Suffys P, Grandjean L, Monteserin J, Debech N, et al. Global expansion of *Mycobacterium tuberculosis* lineage 4 shaped by colonial migration and local adaptation. Sci Adv. 2018; 4: 1-12.

134. Rabahi MF, Conceição EC, de Paiva LO, Souto MVML, Sisco MC, de Waard J, et al. Characterization of *Mycobacterium tuberculosis* var. *africanum* isolated from a patient with pulmonary tuberculosis in Brazil. Infect Genet Evol. 2020; 85: 104550.

135. Comas I. Genomic epidemiology of tuberculosis. Adv Exp Med Biol. 2017; 1019: 79-93.

136. Merker M, Kohl TA, Niemann S, Supply P. The evolution of strain typing in the *Mycobacterium tuberculosis* complex. Adv Exp Med Biol. 2017; 1019: 43-78.

137. Salvato RS, Costa ERD, Reis AJ, Schiefelbein SH, Halon ML, Barcellos RB, et al. First insights into circulating XDR and pre-XDR *Mycobacterium tuberculosis* in Southern Brazil. Infect Genet Evol. 2020; 78: 1-14.

138. Guimarães AES, Sharma A, Furlaneto IP, Rutaihwa L, Cardoso JF, da Conceição ML, 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.

139. Feliciano CS, Plaça JR, Peronni K, Silva WA, Bollela VR. Evaluation of resistance acquisition during tuberculosis treatment using whole genome sequencing. Braz J Infect Dis. 2016; 20(3): 290-3.

140. Rossetti ML, Almeida da Silva PE, Salvato RS, Reis AJ, Schiefelbein SH, von Groll A, et al. A highly rifampicin resistant *Mycobacterium tuberculosis* strain emerging in Southern Brazil. Tuberculosis. 2020; 125: 1-13.

141. Verza M, Scheffer MC, Salvato RS, Schorner MA, Barazzetti FH, Machado HM, et al. Genomic epidemiology of *Mycobacterium tuberculosis* in Santa Catarina, Southern Brazil. Sci Rep. 2020; 10(1): 1-8.

142. Hadi SA, Kolte IV, Brenner EP, Cunha EAT, Simonsen V, Ferrazoli L, et al. Identification of a predominant genotype of *Mycobacterium tuberculosis* in Brazilian indigenous population. Sci Rep. 2021; 11: 1-12.

143. Colangeli R, Gupta A, Vinhas SA, Venkata UDC, Kim S, Grady C, et al. *Mycobacterium tuberculosis* progresses through two phases of latent infection in humans. Nat Commun. 2020; 11: 1-10.

144. Lim LKY, Sng LH, Win W, Chee CBE, Hsu LY, Mak E, et al. Molecular epidemiology of *Mycobacterium tuberculosis* complex in Singapore, 2006-2012. PLoS One. 2013; 8(12): 1-9.

3.2 CAPÍTULO II

O manuscrito que constitui este capítulo, intitulado **"Genomic-based surveillance reveals high ongoing transmission of multi-drug-resistant *Mycobacterium tuberculosis* in Southern Brazil"** apresenta uma análise baseada em WGS da diversidade genômica de *M. tuberculosis* circulando no Rio Grande do Sul.

O trabalho foi publicado na revista **International Journal of Antimicrobial Agents** (https://www.journals.elsevier.com/international-journal-of-antimicrobial-agents), com fator de impacto **JCR 2021 = 15.441**.

O manuscrito está disponível em: https://doi.org/10.1016/j.ijantimicag.2021.106401.

International Journal of Antimicrobial Agents 58 (2021) 106401

Contents lists available at ScienceDirect

International Journal of Antimicrobial Agents

journal homepage: www.elsevier.com/locate/ijantimicag

Genomic-based surveillance reveals high ongoing transmission of multi-drug-resistant *Mycobacterium tuberculosis* in Southern Brazil

Richard Steiner Salvato[a,b,*], Ana Júlia Reis[c,†], Sun Hee Schiefelbein[d,†], Michael Andrés Abril Gómez[c], Stéphanie Steiner Salvato[b], Larissa Vitória da Silva[b], Elis Regina Dalla Costa[e], Gisela Unis[f], Claudia Fontoura Dias[f], Miguel Viveiros[g], Isabel Portugal[h], Andrea von Groll[c], Pedro Eduardo Almeida da Silva[c], Afrânio Lineu Kritski[e,‡], João Perdigão[h,‡], Maria Lucia Rosa Rossetti[a,d,‡]

[a] Programa de Pós-graduação em Biologia Celular e Molecular, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil
[b] Centro de Desenvolvimento Científico e Tecnológico, Centro Estadual de Vigilância em Saúde, Secretaria Estadual da Saúde do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil
[c] Medical Microbiology Research Center, Faculdade de Medicina, Programa de Pós-graduação em Ciências da Saúde, Universidade Federal do Rio Grande, Rio Grande, Rio Grande do Sul, Brazil
[d] Programa de Pós-graduação em Biologia Celular e Molecular Aplicada à Saúde, Universidade Luterana do Brasil, Canoas, Rio Grande do Sul, Brazil
[e] Programa Acadêmico de Tuberculose, Faculdade de Medicina e complexo hospitalar HUCFF-IDT, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil
[f] Hospital Sanatório Partenon, Porto Alegre, Rio Grande do Sul, Brazil
[g] Unidade de Microbiologia Médica, Global Health and Tropical Medicine, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisboa, Portugal
[h] iMed.ULisboa – Research Institute for Medicine, Faculdade de Farmácia, Universidade de Lisboa, Lisboa, Portugal

ARTICLE INFO

Article history:
Received 8 April 2021
Accepted 3 July 2021

Editor: Jean-Marc Rolain

Keywords:
Tuberculosis
Drug resistance
Transmission
Whole-genome sequencing

ABSTRACT

Genomic-based surveillance on the occurrence of drug resistance and its transmission dynamics has emerged as a powerful tool for the control of tuberculosis (TB). A whole-genome sequencing approach, phenotypic testing and clinical-epidemiological investigation were used to undertake a retrospective population-based study on drug-resistant (DR)-TB in Rio Grande do Sul, the largest state in Southern Brazil. The analysis included 305 resistant *Mycobacterium tuberculosis* strains sampled statewide from 2011 to 2014, and covered 75.7% of all DR-TB cases identified in this period. Lineage 4 was found to be predominant (99.3%), with high sublineage-level diversity composed mainly of 4.3.4.2 [Latin American and Mediterranean (LAM)/RD174], 4.3.3 (LAM/RD115) and 4.1.2.1 (Haarlem/RD182) sublineages. Genomic diversity was also reflected in resistance of the variants to first-line drugs. A large number of distinct resistance-conferring mutations, including variants that have not been reported previously in any other setting worldwide, and 22 isoniazid-monoresistant strains with mutations described as disputed in the *rpoB* gene but causing rifampicin resistance generally missed by automated phenotypic tests as BACTEC MGIT. Using a cut-off of five single nucleotide polymorphisms, the estimated recent transmission rate was 55.1%, with 168 strains grouped into 28 genomic clusters. The most worrying fact concerns multi-drug-resistant (MDR) strains, of which 73.4% were clustered. Different resistance profiles and acquisition of novel mutations intraclusters revealed important amplification of resistance in the region. This study described the diversity of *M. tuberculosis* strains, the basis of drug resistance, and ongoing transmission dynamics across the largest state in Southern Brazil, stressing the urgent need for MDR-TB transmission control state-wide.

* Corresponding author. Address: Centro Estadual de Vigilância em Saúde, Secretaria Estadual da Saúde do Rio Grande do Sul, Av. Ipiranga 5400, CEP 90610-000, Porto Alegre, Rio Grande do Sul, Brazil. Tel.: +55 51 985767035.
E-mail address: richardsalvato@hotmail.com (R.S. Salvato).
† These authors contributed equally.
‡ These authors supervised this work equally.

https://doi.org/10.1016/j.ijantimicag.2021.106401
0924-8579/

1. Introduction

Resistance to anti-tuberculosis (TB) drugs is one of the main reasons why TB is still one of the leading infectious diseases. In 2019, TB accounted for 1.4 million deaths and 10 million new cases worldwide. Drug resistance increases the standard anti-TB treatment length to at least 1 year, and requires the use of different classes of drugs which are more expensive, less effective and have a lower cure rate: 26% for extensively drug-resistant (XDR) cases [1].

Annually, approximately 73,000 new cases of TB and 4500 deaths due to TB are notified in Brazil, resulting in an incidence rate of 35 cases/100,000 population (2019). From these cases, an important fraction is classified as multi-drug resistant (MDR)/rifampicin (RIF)-resistant: 1119 laboratory-confirmed MDR-TB/RIF-resistant cases were notified in 2017. Nevertheless, the distribution of the disease is heterogeneous across the country, and Rio Grande do Sul, the southernmost Brazilian state and the largest of the three states that compose the south region of Brazil, accounts for 54% of the cases of TB occurring in this region with an incidence rate of 40 cases/100,000 population [2]. Furthermore, Rio Grande do Sul has a high number of cases of MDR/RIF-resistant TB (109 cases in 2017) according to the data from the State Central Laboratory (not shown).

Previous studies on circulating *Mycobacterium tuberculosis* in the south region of Brazil have shown predominance of Lineage 4 strains, but with important diversity at the sublineage level with predominance of 4.3.3 [Latin American and Mediterranean (LAM)], 4.3.4.2 (LAM) and 4.1.2.1 (Haarlem) sublineages [3,4]. Moreover, strains rarely found in other regions, such as SIT863 [5], have been observed in recent studies, coupled with ongoing transmission of highly resistant strains [4,6]. However, previous studies in the region only undertook limited sampling, mainly from MDR strains. To obtain a broader understanding of the DR-TB scenario and aiming to inform DR-TB control efforts in the region, a representative sample of DR *M. tuberculosis* strains circulating in Rio Grande do Sul state between 2011 and 2014 was characterized in terms of their diversity and genomic similarity to estimate ongoing transmission of *M. tuberculosis*.

2. Materials and methods

2.1. Study population

A retrospective population-based study on DR-TB was conducted in Rio Grande do Sul state, Southern Brazil, between 2011 and 2014. Available *M. tuberculosis* clinical strains were collected at the State Central Laboratory (LACEN-RS). LACEN-RS is the reference laboratory in charge of drug susceptibility testing (DST) from TB cases notified state-wide. The study included 305 clinical strains of *M. tuberculosis* that presented resistance to at least one of the following first-line anti-TB drugs: isoniazid (INH), RIF, ethambutol (EMB) and streptomycin (STR).

Rio Grande do Sul, the largest state in Southern Brazil, has an estimated population of 11.3 million people [7], and ranks among the high-burden TB states in Brazil with a high rate of MDR-TB. In 2017, 5031 new TB cases [2] (40 cases/100,000 population) and 90 cases of diagnosed MDR (data from LACEN-RS) were notified from Rio Grande do Sul. Over the 4-year study period, 403 DR-TB samples were identified at LACEN-RS. The study sample includes 305 of these 403 notified DR-TB samples (75.7%). Clinical and epidemiological data from enrolled individuals were obtained from the Brazilian National System for Notifiable Diseases (SINAN) and the Brazilian Special Tuberculosis Treatment Information System (SITE-TB). A confirmed epidemiological link (epi-link) was considered when two individuals lived in the same neighbourhood or spent time at the same prison for any time in the same year.

2.2. Drug susceptibility testing

DST results were obtained for clinical samples from LACEN-RS, where testing was performed using a liquid BACTEC MGIT 960 SIRE Kit for the BACTEC Mycobacteria Growth Indicator Tube 960 (MGIT 960) system (Becton Dickinson Diagnostic Systems, Sparks, MD, USA). Susceptibility evaluation was conducted for the following first-line anti-TB drugs: RIF (breakpoint 1.0 mg/L), INH (breakpoint 0.1 mg/L), EMB (breakpoint 5.0 mg/L) and STR (breakpoint 1.0 mg/L) in accordance with the manufacturer's instructions.

2.3. DNA extraction and whole-genome sequencing

M. tuberculosis genomic DNA was extracted using the cetyltrimethylammonium bromide method, as described by van Embden et al. [8] in a Biosafety Level (BSL) 2 laboratory with BSL-3 safety equipment and work practices. The genomic DNA from the 305 studied *M. tuberculosis* strains was subjected to next-generation sequencing to access its whole-genome sequence. Paired-end sequencing (2 × 150 bp) was performed on an Illumina NextSeq machine using either a 300 cycle v2 mid-output or high-output kit (Illumina, Code FC-404-2003 or Code FC-404-2004), using the standard Illumina procedure as described previously [4].

2.4. Bioinformatic analysis of whole-genome sequencing data

Raw FASTQ files were trimmed to remove adapter sequences and low-quality reads using Trimmomatic [9]. The read data quality was assessed in fastQC (www.bioinformatics.babraham.ac.uk/projects/fastqc/), and then mapped against the reference genome of *M. tuberculosis* H37Rv (GenBank Accession NC000962.3) using the BWA-MEM algorithm [10]. The quality of the resulting mapped BAM files was checked using Qualimap [11]. SAMtools/BCFtools and GATK tools were used for variant calling of single nucleotide polymorphisms (SNPs) and small indels, as described previously [12,13]. Both variant sets called by each tool were combined, and only the concordant set between both callers was retained for downstream analysis. SNP sites with an excess of 10% missing calls were removed from the analysis [14], as well as SNP positions within PE/PPE genes. Variants occurring in drug-resistance-associated genes were retained for increased resolution at the micro-evolutionary level. Retained variants were converted to a FASTA format file and then used to generate an alignment composed of 28,974 high-quality SNP sites.

The final [whole-genome SNP (wgSNP)] alignment was submitted to the JmodelTest tool [15] for best-fit nucleotide substitution model selection under Akaike Information Criterion. Generalized Time-Reversible model with estimated proportion of invariable sites was selected to reconstruct a maximum-likelihood phylogenetic tree using the RAxML tool [16], applying the bootstrap branch support metric, and the resulting tree was annotated in the Interactive Tree of Life online tool [17]. A cut-off of five SNPs was used to delineate genomic clusters [18] among the wgSNP alignment using the *ape* package and *hclust* function implemented in R. Recent transmission among *M. tuberculosis* genomes was estimated from the ratio between the number of clustered strains and the total number of included strains. To analyse the geographical distribution of DR-TB cases, patient addresses were plotted in a map using the online tool Microreact [19]. Pairwise geographical distances between patients were determined using the *Imap* package implemented in R, and the Mann-Whitney *U*-test was applied.

The *fbp*C103 polymorphism (G->A at codon 103) used to differentiate LAM strains from non-LAM strains and genomic variants

underlying drug resistance were identified from VCF files cited previously. The command-line version of TB-Profiler (v2.8.13) [20] was used to determine the *M. tuberculosis* SNP-based type, and SpoTyping (v2.0) [21] was used for prediction of in-silico spoligotypes. The clade and shared international type (SIT) of each spoligotyping pattern were assigned using the SITVIT2 online database [22].

2.5. Resolving discordant phenotypic and whole-genome sequencing results

Discordant results were resolved by repeating the DST on the MGIT 960 system and performing minimum inhibitory concentration (MIC) analyses with the resazurin microtitre assay method, as described previously [23,24], using the breakpoint concentrations of ≤0.25mg/L and ≤0.5 mg/L for INH and RIF, respectively.

3. Results

3.1. Study population and clinical characteristics

From 2011 to 2014, 403 cases of DR-TB were detected statewide and further classified as MDR (n=240), resistant to INH alone (n=122), resistant to RIF alone (n=6), resistant to EMB alone (n=3), resistant to STR alone (n=3), and polyresistant (n=32). Of these 403 DR-TB notified cases, 305 *M. tuberculosis* resistant strains from unique patients were available (75.7% of total from study period) with the following DR profiles: 169 MDR, 111 resistant to INH alone, 19 polyresistant, five resistant to RIF alone, and one resistant to STR alone.

Among the 305 included patients, 76.1% were male and the median age was 40 years (range 13–84 years). For 34.4% of the patients, it was their first TB diagnosis (primary resistance), 29.8% were starting a new treatment period for a TB relapse, 25.3% were starting treatment following loss to follow-up, and 9.2% were starting treatment after failure of a previous treatment. Information was not available for four individuals (1.3%) (according to data obtained from SINAN and SITE-TB). Regarding the main comorbidities and risk factors, 28.5% were infected with human immunodeficiency virus; 8.9% had diabetes mellitus; and 33.1%, 19.7% and 18.4% were alcohol, tobacco and illicit drug users, respectively. In addition, 37 (12.13%) individuals were prison inmates at the time of sample collection. Information related to treatment outcome was obtainable for 254 (83.3%) cases. A favourable outcome (cure or treatment completion) was reported for 49.9% (104/254) of those cases; 35.8% (91/254) of the individuals were lost to follow-up, the current treatment failed for 6.7% (17/254) of the individuals, and 16.5% (42/254) of the individuals died during treatment (see online supplementary material for detailed characteristics of individuals included).

3.2. Drug resistance and associated mutations

From the 111 INH-monoresistant clinical strains, 95 (85.6%) carried a well-described mutation related to resistance. Among the 169 phenotypically MDR strains, only one did not have any known mutations underlying INH and RIF resistance. For 43 samples that presented discordant results for INH and RIF resistance between MGIT testing and the whole-genome sequencing (WGS) prediction, MGIT- and MIC-based DST were performed for the two drugs. This revealed that 15 INH-monoresistant strains on MGIT-SIRE did not have associated resistance mutations, and were, in fact, susceptible. Therefore, for the purposes of evaluation of WGS-based drug-resistance prediction, the susceptibility profile was changed. For the remaining strains, the phenotypic resistance profile remained the same.

Additionally, well-described disputed mutations in the *rpoB* gene causing RIF resistance, generally missed by automated phenotypic DST methods [25], were identified in 22 (22.9%) of the 96 remaining INH-monoresistant strains: D435Y (8/22), L452P (5/22), H445Y (4/22), L430P (2/22), H445G (1/22), H445N (1/22) and D435Y+S431T (1/22). Thus, for the purpose of evaluation of WGS-based drug-resistance prediction, those strains were considered to be RIF resistant based on previous reports of the phenotypic resistance caused by these mutations, and the inability of MGIT to detect it [25]. Concerning clinical evolution for these 22 patients, 17 of them received MDR-TB regimens after a review of their clinical status (non-response or treatment failure to first-line drugs, including RIF). For the remaining five patients, the treatment information was not available, but two of them died from TB. Finally, the overall concordance on WGS-based resistance prediction was 99.3% for INH, 98.7% for RIF, 82.3% for EMB and 86.9% for STR, considering MGIT-based testing as the gold standard (Table 1).

The most common variants sustaining INH resistance were found in the *katG* gene, with the distinct occurrence of some mutations among different resistance profiles (Ser315Thr mainly in MDR and Ser315Asn mainly in INH-monoresistant). In total, 84.5% (240/284) of INH-resistant strains had some related resistance mutation at *katG*, and among these, 53 (18.7%) had additional -15 C>T or T>C -8 changes at the *inhA* promoter. Thirty of 284 (10.6%) INH-resistant strains only carried the *inhA*-promoter -15 C>T mutation, 11/284 (3.9%) only had mutations in the *oxyR-ahpC* regulatory region, and one had the Ser94Ala codon change in *inhA* gene (Figure 1). Regarding RIF resistance, 173 (99.4%) of the 174 RIF-resistant isolates carried mutations in the *rpoB* gene. Amino acid changes in codon 450 were the most frequent, occurring in 73.6% of the resistant isolates. An uncommon insertion (*rpoB* 435 QNNP > QNNPQNNP) [6] was identified in 16/174 isolates (9.2%), and two isolates (1.1%) had deletion of two amino acids (*rpoB* 435 FMD>F) (Figure 1). Disputed mutations occurring in the *rpoB* gene were found in 22 RIF-susceptible (according to MGIT) strains: 14 of them were INH-monoresistant strains, seven were resistant to INH and STR, and one was susceptible to first-line drugs.

Mutations associated with EMB resistance were found in 91.7% (11/12) of the resistant isolates, all carried on the Met306Val amino acid change in the *embB* gene. However, 18.1% (53/293) of the EMB-susceptible strains also had resistance-associated mutations. Among 47 STR-resistant strains, 28 (59.6%) had related resistance mutations identified on the *rrs* (11/47), *gidB* (11/47) and *rpsL* (7/47) genes (Figure 1). Second-line drug-resistance-associated variants were found in nine strains, all of which carried amino acid substitutions at codon 94 on the *gyrA* gene (Asp94His, Asp94Gly, Asp94Asn, Asp94Tyr and Asp94Ala), related to fluoroquinolone resistance. One of these nine strains also had a mutation on the *rrs* gene (1401 A>G), which is a known marker of aminoglycoside resistance (see online supplementary material for complete resistance-associated variant profiles). For one of these nine patients with pre/XDR-TB, the outcome was cure following a change in the anti-TB regimen. In the other eight individuals, an unfavourable outcome was seen: four had treatment failure, three were lost to follow-up (including the one case of XDR-TB), and one died.

3.3. Genetic diversity of M. tuberculosis

The two typing methods implemented – in-silico spoligotyping and SNP-based typing – assigned 303 (99.3%) clinical strains to *M. tuberculosis* Lineage 4, one strain to Lineage 1 (EAI5) and one strain to *M. bovis* (BOV1) species. SNP-based classification showed predominance of the 4.3.4.2 (LAM) sublineage in 63 (20.7%) strains, 4.3.3 (LAM) in 61 (20%) strains, and 4.1.2.1 (Haarlem) in 54 (17.7%)

Table 1
Agreement between phenotypic testing and genome-based drug resistance prediction with sensitivity and specificity values for whole-genome sequencing (WGS)-based resistance detection.

Drug	DST	Total	WGS (R)	WGS (S)	Sensitivity	Specificity	PVP	PVN
INH	R	284	282	2[a]	0.99	1.00	1.00	0.91
	S	21	0	21				
RIF	R	174	172	1[a]	0.99	0.97	0.98	0.99
	S	131	25[b]	106				
EMB	R	12	11	1	0.92	0.82	0.17	1.00
	S	293	53	240				
STR	R	47	28	19	0.60	0.92	0.57	0.93
	S	258	21	237				

DST, drug-susceptibility testing; R, resistant; S, susceptible; INH, isoniazid; RIF, rifampicin; EMB, ethambutol; STR, streptomycin; PVP, predictive value for a positive test; PVN, predictive value for a negative test.
[a]Mycobacteria Growth Indicator Tube (MGIT) and minimum inhibitory concentration (MIC) determination were performed to confirm phenotypic resistance.
[b] Disputed mutations in the *rpoB* gene were identified in 22 strains. Due to the well-established basis of these mutations, RIF-susceptible samples on MGIT presenting disputed mutations were considered to be phenotypically resistant. The remaining three RIF-susceptible samples on MGIT, carried on the *rpoB* gene and highly related to resistance variants, were submitted to MGIT to confirm phenotypic susceptibility.

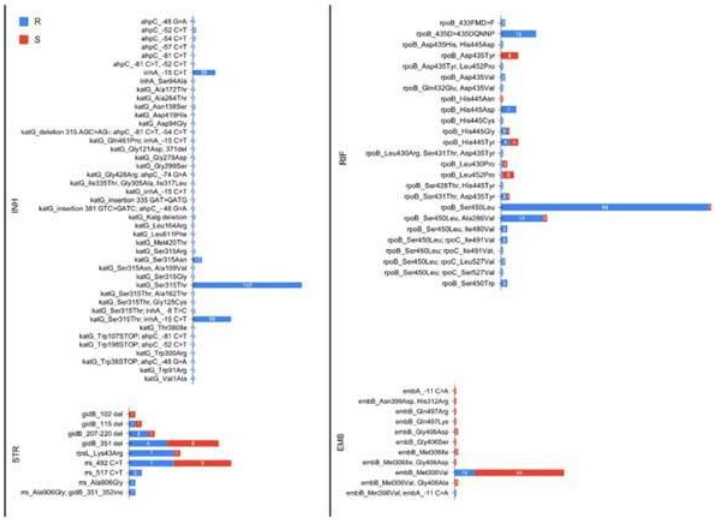


Figure 1. Mutation patterns underlying first-line drug resistance among the 305 drug-resistant *Mycobacterium tuberculosis* strains.

strains. All 188 strains (61.6%) assigned to the 4.3 (LAM) sublineage carried the *fbpC*[103] polymorphism, a known LAM SNP marker; RD174 was found in 64 (34%) of these LAM strains and RD115 was found in 55 (29.3%). In-silico spoligotyping revealed 62 distinct patterns and 48 different SITs among the 305 clinical strains. The most common SITs were: 53/T1 in 40 (13.1%) strains, 93/LAM5 in 26 (8.5%) strains, 65/T1 in 24 (7.9%) strains and 42/LAM9 in 22 (7.2%) strains. The SITVIT2 database did not assign any SITs to 24 (7.9%) strains (Figure 2). SNP barcode-based typing was able to assign a sublineage to 12 strains that were previously unclassified or ill defined by spoligotyping. Disagreements between spoligotyping and SNP-based typing methods were related to the T spoligotype. The complete typing profile for each strain is presented in the online supplementary material (molecular sheet).

3.4. Whole-genome-sequencing-based *M. tuberculosis* transmission analysis

The final wgSNP alignment of the 305 *M. tuberculosis* strains included in this study resulted in 28,974 positions and evidenced an average distance of 498.68 SNPs between strains. Five SNPs was used as the threshold for genomic relatedness detection in *M. tuberculosis* genomes, which could reveal recent or ongoing transmission as proposed by Walker et al. [18]. From 305 strains, 168 (55.1%) were grouped into 28 genomic clusters. Among those, 22 epi-links were identified, involving 77/168 (45.8%) patients: 47 at community level, 30 in prison and two household contacts (see online supplementary material). Among the 169 patients with MDR-TB, 124 (73.4%) were grouped into genomic clusters. The three largest clusters identified (GC1, GC2 and GC3) harboured strains

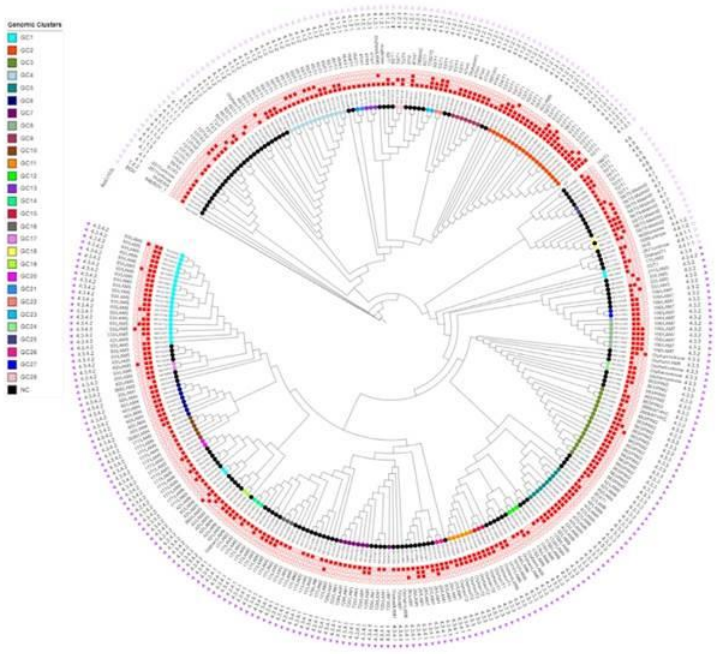


Figure 2. Phylogenetic tree reconstructed using the maximum likelihood approach from genome-wide alignment of the 305 *Mycobacterium tuberculosis* strains with 28,974 high-quality positions. The resulting tree was rooted on *Mycobacterium cannettii* and annotated in Itol [17]. The tips were coloured according to genomic clustering (see legend). The phenotypic drug susceptibility profile for first-line drugs is represented in the red squares (resistant, filled squares; susceptible, empty squares). Spoligotyping clade/shared international type and single-nucleotide-polymorphism-based typing are annotated, and the triangle indicates the presence or absence of fbpC103 polymorphism (presence, filled triangle; absence, empty triangle).

from the 4.3.4.2 (LAM), 4.1.2.1 (Haarlem) and 4.3.3 (LAM) sublineages, respectively. GC1 comprised 25 strains, 19 of which belonged to SIT 93/LAM5. GC2 had 21 strains in total, 20 of which belonged to SIT 53/T1. Of the 17 strains in GC3, 15 had a spoligotyping pattern assigned to SIT 863/PINI2 and 16 had a 12nt insertion at codon 435 of the *rpoB* gene (435QNNP > QNNPQNNP). Furthermore, 14 of the 22 strains harbouring disputed mutations, and susceptible according to the MGIT assay, were grouped into five different clusters, mainly GC4 (9/22).

The largest genomic clusters were mainly composed of MDR strains. However, in nine genomic clusters, the strains presented distinct intracluster drug-resistance patterns. For example, GC1 was composed of 19 MDR strains and five INH-monoresistant strains, and GC4 had a heterogeneous composition (five polyresistant, four INH-monoresistant and four MDR strains). In addition, 13 GCs had intracluster acquisition of additional mutations conferring resistance to first- and second-line drugs (see online supplementary material). The distribution of clustered strains was also analysed in terms of the different cities in Rio Grande do Sul state, and the highest occurrence was found in cities with the highest incidence of TB (mainly the metropolitan area of Porto Alegre and the state capital) [26]. Figure 3 shows the geographical distribution of studied strains, revealing a higher concentration of studied cases and clustered strains (77.4%) in the metropolitan area of Porto Alegre (incidence rate 84.4 cases/100,000 population) [2]. In Porto Alegre, clustered cases were mainly observed in four districts (Santa Tereza, Rubem Berta, Mário Quintana and Sarandi) – disregarding the cases from the city's prison – accounting for 40.3% of clustered cases in the city. The pairwise geographical distance between patient residences within the five larger genomic clusters was statistically lower than observed in non-clustered cases ($P<0.05$).

4. Discussion

Rio Grande do Sul has the fourth highest TB mortality rate among Brazilian states, and recent studies in the region have shown important chains of transmission of DR-*M. tuberculosis* strains that impair disease control [2,4,27,28]. Data are lacking regarding the real incidence of DR-TB in Rio Grande do Sul, as in other Brazilian states, as these data are poorly covered in national reports. In this retrospective study, DR-TB cases identified at Rio Grande do Sul State Central Laboratory were reviewed to obtain the epidemiological scenario of resistance in the state.

Overall, the predicted genome-based resistance showed good concordance with phenotypic resistance testing, ranging from 99.3% and 98.7% resistance to INH and RIF, respectively, to 82.3% and 86.9% resistance to EMB and STR, respectively, in accordance with global data [29,30]. Interestingly, 88.7% of EMB-susceptible strains harbouring resistance-associated mutations were MDR, similar to the results of a previous study in Russia [31]. Lower sensitivity to the detection of STR resistance was found; this has been reported in previous studies with global strain sets, and could be associated with poor genomic predictive performance for STR resistance within Lineage 4 [32,33]. A relevant finding observed from molecular versus phenotypic comparison was the presence

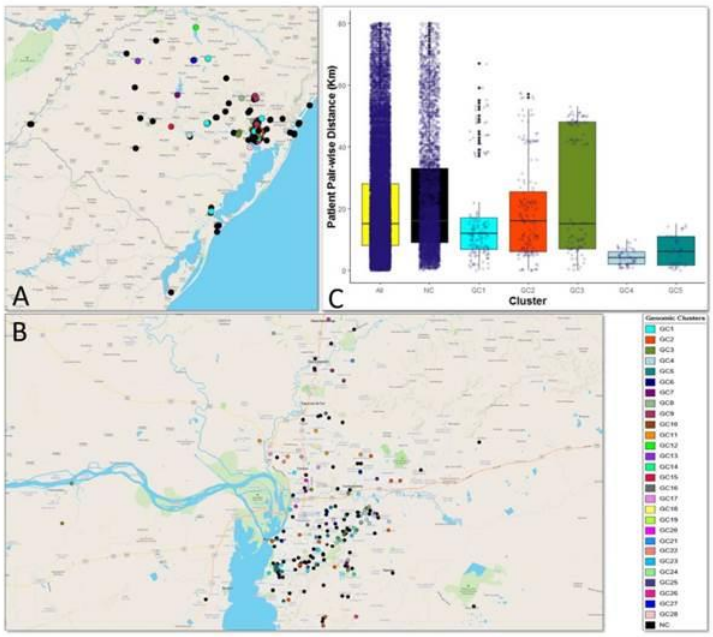


Figure 3. (A) Geographical distribution of the 305 drug-resistant tuberculosis cases in Rio Grande do Sul state. The colours of each marker represent the genomic cluster of each strain (see legend). (B) Zoomed image representing the Porto Alegre metropolitan region. The data were plotted in the map using Microreact [19]. An interactive version is available online at https://microreact.org/project/cZ1SD6nx9LqimJHhNoToG5. (C) Pairwise geographical distances of all patients, and non-clustered and intraclustered groups.

of well-characterized disputed mutations in the *rpoB* gene [25] in 22 strains that were susceptible to RIF on MGIT. In these strains, disputed mutations were found in 21.9% (21/96) of phenotypically INH-monoresistant strains, suggesting potential underestimation of the proportion of MDR cases, and may result in incorrect anti-TB treatment. The fact that 17 patients infected with these strains received an MDR-TB regimen shows the failure of the first-line regimen despite being considered simple INH-monoresistant strains, and two deaths during treatment shows the clinical relevance of these mutations. In addition, the potential for dissemination of these strains is further stressed by the detection of isolates bearing these variants in genomic transmission clusters, particularly GC4. These novel findings in Brazil reveal the importance of tracking strains harbouring disputed mutations for accurate TB surveillance and control in the region.

As revealed by previous studies carried out in the region [4,27,28], the main genotypes associated with first-line drug resistance include high confidence mutations: *katG* S315T (in 50% of resistant strains), *katG* S315T + *inhA* -15 C>T (in 17.6% of resistant strains) and *inhA* -15 C>T (in 10.2% of resistant strains) causing resistance to INH. However, 3.9% (11/284) of INH-resistant strains only carried resistance-conferring mutations on the *ahpC* promoter (-48 G>A, -52 C>T, -54 C>T and -57 C>T), eight of which were resistant to INH alone. Despite limited evidence about the role of mutations in the *ahpC-oxyR* intergenic region conferring resistance to INH, several studies have found a potential association between these variants and INH resistance [29,34], but usually as a compensatory mechanism to mutations at *katG*, mainly non-315 mutations. Overexpression of the *ahpC* gene, producing an alkyl peroxidase that also protects the bacillus against the toxic effects of organic peroxides, may occur due to reduced activity of catalase peroxidase. In addition, this study found mutations in the *ahpC* gene in six strains, which also carried variants in the *katG* gene, all of which were non-315 mutations. Regarding the 11 strains that carried *ahpC* gene variants alone, no other molecular markers that could elucidate the mechanisms causing resistance were identified. In order to further understand the occurrence of resistance in these strains, there is a need to analyse other molecular mechanisms, such as the expression of efflux pumps.

RIF resistance was mainly caused by variants carrying the *rpoB* S450L (in 54.6%) and *rpoB* S450L + A286V (in 10.9%) mutations, and by the 12-nucleotide insertion at codon 435 of the *rpoB* gene (in 9.2%), stressing the significant spread of these highly RIF-resistant strains (MIC ≥32 mg/L) in the population [6]. In the same way, most common resistance-associated mutations were found among clinical strains that were phenotypically resistant to EMB (*embB* M306V in 83.3% of resistant strains) and STR (*rrs* 492 C>T in 14.9% and *rpsL* K43R in 14.9% of resistant strains), similar to results in global collections [14].

Mutations conferring resistance to second-line drugs were found in nine strains: six MDR strains had mutations conferring fluoroquinolone resistance in the *gyrA* gene, one polyresistant and two INH-monoresistant strains had associated resistance mutations in the *gyrA* gene, and one of the latter also had a nucleotide change (1401a>g) in the *rrs* gene that was related to resistance to injectable second-line drugs, characterizing an XDR-TB case (see online supplementary material). According to clinical data, only one of these nine pre/XDR-TB individuals had a favourable out-

come, and three of them were lost to follow-up, which may have contributed to the spread of this resistance. In this study, DST to second-line anti-TB drugs was not performed, precluding comparison with WGS-based predictions. However, a previous work showed good concordance among resistance-conferring mutations and phenotypic resistance to fluoroquinolones amid a subset of clinical strains from the same region [4].

As characterized previously by recent studies in the region, the population structure of *M. tuberculosis* is mainly composed of LAM and Haarlem strains [4,27]. The same predominance of LAM strains occurs in the bordering state of Santa Catarina [35], despite predominance of the 4.3.3 sublineage in the neighbouring state compared with predominance of the 4.3.4.2 sublineage observed in the present study. As well as being able to determine the sublineage of strains not classified by spoligotyping, SNP typing was also able to reclassify a large set of strains previously defined as T sublineage into more plausible sublineages based on their spoligotype pattern. From 93 strains that were assigned as T clade by SITVIT2, only 16 were classified as T clade by SNP-based typing; the remainder were mainly assigned to the 4.1.2.1/Haarlem sublineage (39.8%). This fact was evidenced by 12 strains carrying the fbpC103 variant, a known LAM marker, and by the position in the phylogenetic tree. Moreover, the ill-defined family T assigned by spoligotyping, especially the spoligotype SIT53, has been demonstrated previously as a common pattern for different sublineages in Lineage 4 [36].

Overall, the estimated recent transmission rate of 55.1% among the 305 studied samples, which is higher than that registered in the bordering state of Santa Catarina [35] (mainly susceptible strains) and other regions with a high burden of TB [37] and MDR-TB [38], shows that important transmission chains are feeding the ongoing transmission in the region. However, the most worrying fact revealed in this study concerns MDR strains, 73.4% of which were grouped in genomic clusters. The three larger genomic clusters found in the present analysis consisted, almost exclusively, of MDR strains. GC1 was mainly constituted (24/25) by strains from individuals without a history of imprisonment, with 15 individuals involved in identified community epi-links, representing a large ongoing chain of transmission in the community. In GC2, 10/21 strains were isolated from prison inmates, and in GC3, 8/17 of the clustered strains came from prison inmates or individuals with a recent history of incarceration, along with a prison worker. Thus, GC3 represents an important active transmission chain involving the inmate population and the community [6], similar to that found in other regions in Brazil [39] and worldwide [40].

The larger ongoing transmission chains were found in Porto Alegre (capital and metropolitan area). In Porto Alegre, the four districts accounting for 40.3% of clustered cases in the city presented the lowest Human Development Index [7], reinforcing the need for more robust TB control measures in this region. The presence of MDR strains in genomic clusters composed principally of mono-/polyresistant profiles, along with intracluster acquisition of resistance-related mutations seen in multiple (mainly larger) clusters, indicates the de-novo emergence of MDR-TB in those clusters, leading to the amplification of resistance; this is an additional important area of concern for TB control.

5. Conclusions

These findings reveal important aspects of the molecular basis of drug resistance in *M. tuberculosis* strains circulating in Rio Grande do Sul, Brazil, showing the ability of molecular assays to detect drug resistance, and their importance to detect RIF resistance caused by disputed mutations, thus avoiding missing MDR cases. Multiple ongoing transmission events of DR *M. tuberculosis* strains were identified, mainly MDR strains, stressing the need for measures to interrupt *M. tuberculosis* transmission in the region and the need to improve TB control in prisons.

Acknowledgements

The authors wish to thank TGen, C-Path and ReSeqTB for supporting WGS, and the Brazilian Network of Tuberculosis Research for enabling this partnership. The authors would also like to thank the Mycobacteria Laboratory at LACEN-RS/CEVS/SES-RS for performing and providing information on drug susceptibility testing.

Funding: This study was funded, in part, by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (Grant No. 001) and supported by Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (Grant No. 17/1265–8 INCT-TB). JP is supported by Fundação para a Ciência e Tecnologia through Estímulo Individual ao Emprego Científico (CEECIND/00394/2017).

Competing interests: None declared.

Ethical approval: This project was approved by the Research Ethics Committee of the Fundação Estadual de Produção e Pesquisa em Saúde (Protocol No. 1.587.621 CAAE: 18269313.0.0000.5320).

Data availability: *M. tuberculosis* genome data were deposited in the NCBI BioProject database (IDs: PRJNA535343, PRJNA639713 and PRJNA692642). Individual accession numbers for genomes analysed in this study are given in the online supplementary material.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijantimicag.2021.106401.

References

[1] WHO Global tuberculosis report 2019. Geneva: World Health Organization; 2019.
[2] Secretaria de Vigilância em Saúde, Brasil Ministério da Saúde Boletim Epidemiológico. Brasília: Secretaria de Vigilância em Saúde; 2020.
[3] Perdigão J, Silva C, Diniz J, Pereira C, Machado D, Ramos J, et al. Clonal expansion across the seas as seen through CPLP-TB database: a joint effort in cataloguing Mycobacterium tuberculosis genetic diversity in Portuguese-speaking countries. Infect Genet Evol 2018;72:44–58. doi:10.1016/j.meegid.2018.03.011.
[4] Salvato RS, Costa ERD, Reis AJ, Schiefelbein SH, Halon ML, Barcellos RB, et al. First insights into circulating XDR and pre-XDR Mycobacterium tuberculosis in Southern Brazil. Infect Genet Evol 2020;78:104127. doi:10.1016/j.meegid.2019.104127.
[5] Dalla Costa ER, Vasconcelos SEG, Esteves LS, Gomes HM, Gomes LL, da Silva PA, et al. Multidrug-resistant Mycobacterium tuberculosis of the Latin American Mediterranean lineage, wrongly identified as Mycobacterium pinnipedii (Spoligotype International Type 863 [SIT863]), causing active tuberculosis in South Brazil. J Clin Microbiol 2015;53:3805–11.
[6] Rossetti ML, Almeida da Silva PE, Salvato RS, Reis AJ, Schiefelbein SH, von Groll A, et al. A highly rifampicin resistant Mycobacterium tuberculosis strain emerging in Southern Brazil. Tuberculosis 2020;125:102015.
[7] IBGE. Available at: https://cidades.ibge.gov.br/brasil/panorama [accessed 9 December 2019] 2021.
[8] Van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9.
[9] Bolger AM, Lohse M, Usadel B. Genome analysis Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–20.
[10] Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 2009;25:1754–60.
[11] Cruz LM, Garcı F, Okonechnikov K, Tarazona S, Dopazo J, Meyer TF, et al. Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics 2012;28:2678–9.
[12] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–303.
[13] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–9.
[14] Coll F, Phelan J, Hill-Cawthorne GA, Nair MB, Mallard K, Ali S, et al. Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis. Nat Genet 2018;50:307–16.
[15] Posada D. jModelTest: phylogenetic model averaging. Mol Biol Evol 2008;25:1253–6.

[16] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30:1312–13.
[17] Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 2019;47:W256–9.
[18] Walker TM, Ip CLC, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis 2013;13:137–46.
[19] Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genomics 2016;2:e000093.
[20] Coll F, McNerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. Genome Med 2015;7:51.
[21] Xia E, Teo Y-Y, Ong RT-H. SpoTyping: fast and accurate in silico Mycobacterium spoligotyping from sequence reads. Genome Med 2016;8:19.
[22] Couvin D, David A, Zozio T, Rastogi N. Macro-geographical specificities of the prevailing tuberculosis epidemic as seen through SITVIT2, an updated version of the Mycobacterium tuberculosis genotyping database. Infect Genet Evol 2019;72:31–43.
[23] Palomino J-C, Martin A, Camacho M, Guerra H, Swings J, Portaels F. Resazurin microtiter assay plate: simple and inexpensive method for detection of drug resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother 2002;46:2720–2.
[24] Martin A, Palomino JC. Resazurin microtiter assay (REMA): procedure manual. Belgium: Institute of Tropical Medicine, Mycobacteriology Unit Antwerp; 2009.
[25] Miotto P, Cabibbe AM, Borroni E, Degano M, Cirilloa DM. Role of disputed mutations in the rpoB gene in interpretation of automated liquid MGIT culture results for rifampin susceptibility testing of mycobacterium tuberculosis. J Clin Microbiol 2018;56 e01599-17.
[26] Programa Estadual de Controle da Tuberculose Centro Estadual de Vigilância em Saúde SE da S do RG do S. Tuberculose: Informe Epidemiológico; 2019.
[27] Salvato RS, Schiefelbein S, Barcellos RB, Praetzel BM, Anusca IS, Esteves LS, et al. Molecular characterisation of multidrug-resistant Mycobacterium tuberculosis isolates from a high-burden tuberculosis state in Brazil. Epidemiol Infect 2019;147:e216.
[28] Esteves LS, Dalla Costa ER, Vasconcellos SEG, Vargas A, Ferreira SLM Junior, Halon ML, et al. Genetic diversity of Mycobacterium tuberculosis isoniazid monoresistant and multidrug-resistant in Rio Grande do Sul, a tuberculosis high-burden state in Brazil. Tuberculosis 2018;110:36–43.
[29] Allix-Béguec C, Arandjelovic I, Bi L, Beckert P, Bonnet M, Bradley P, et al. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. N Engl J Med 2018;379:1403–15.
[30] Miotto P, Tessema B, Tagliani E, Chindelevitch L, Starks AM, Emerson C, et al. A standardised method for interpreting the association between mutations and phenotypic drug resistance in Mycobacterium tuberculosis. Eur Respir J 2017;50:1701354.
[31] Mokrousov I, Otten T, Vyshnevskiy B, Narvskaya O. Detection of embB306 mutations in ethambutol-susceptible clinical isolates of Mycobacterium tuberculosis from Northwestern Russia: implications for genotypic resistance testing. J Clin Microbiol 2002;40:3810–13.
[32] Schleusener V, Köser CU, Beckert P, Niemann S, Feuerriegel S. Mycobacterium tuberculosis resistance prediction and lineage classification from genome sequencing: comparison of automated analysis tools. Sci Rep 2017;7:46327.
[33] Mahé P, El Azami M, Barlas P, Tournoud M. A large scale evaluation of tbprofiler and mykrobe for antibiotic resistance prediction in Mycobacterium tuberculosis. PeerJ 2019;7:e6857.
[34] Walker TM, Kohl TA, Omar SV, Hedge J, Del Ojo Elias C, Bradley P, et al. Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: a retrospective cohort study. Lancet Infect Dis 2015;15:1193–202.
[35] Verza M, Scheffer MC, Salvato RS, Schorner MA, Barazzetti FH, Machado H de M, et al. Genomic epidemiology of Mycobacterium tuberculosis in Santa Catarina. Southern Brazil. Sci Rep 2020;10:12891.
[36] Mokrousov I. Ubiquitous and multifaceted: SIT53 spoligotype does not correlate with any particular family of Mycobacterium tuberculosis. Tuberculosis 2021;126:102024.
[37] Miyahara R, Smittipat N, Juthayothin T, Yanai H, Disratthakit A, Imsanguan W, et al. Risk factors associated with large clusters of tuberculosis patients determined by whole-genome sequencing in a high-tuberculosis-burden country. Tuberculosis 2020;125:101991.
[38] Chee CBE, Lim LKY, Ong RTH, Sng LH, Hsu LY, Lee VJM, et al. Whole genome sequencing analysis of multidrug-resistant tuberculosis in Singapore, 2006–2018. Eur J Clin Microbiol Infect Dis 2021;40:1079–83.
[39] Sacchi FPC, Praca RM, Tatara MB, Simonsen V, Ferrazoli L, Croda MG, et al. Prisons as reservoir for community transmission of tuberculosis. Brazil. Emerg Infect Dis 2015;21:452–5.
[40] Umpeleva T, Belousova K, Golubeva L, Boteva T, Morozova I, Vyazovaya A, et al. Molecular characteristics of Mycobacterium tuberculosis in the "closed" Russian town with limited population migration. Infect Genet Evol 2020;79:104174.

3.3 CAPÍTULO III

No manuscrito que constitui este capítulo, intitulado **"A highly rifampicin resistant *Mycobacterium tuberculosis* strain emerging in Southern Brazil"**, descrevemos as características fenotípicas, moleculares e epidemiológicas de uma cepa de *Mycobacterium tuberculosis* altamente resistente à rifampicina emergente no sul do Brasil.

O trabalho foi publicado na revista **Tuberculosis** (https://www.sciencedirect.com/journal/tuberculosis), com fator de impacto **JCR 2021 = 2.973**.

O manuscrito está disponível em: https://doi.org/10.1016/j.tube.2020.102015.

Tuberculosis 125 (2020) 102015

Contents lists available at ScienceDirect

Tuberculosis

journal homepage: http://www.elsevier.com/locate/tube

A highly rifampicin resistant *Mycobacterium tuberculosis* strain emerging in Southern Brazil

Maria Lucia Rossetti [a,d,1], Pedro Eduardo Almeida da Silva [b,1], Richard Steiner Salvato [c,d,*], Ana Júlia Reis [b], Sun Hee Schiefelbein [a,c], Andrea von Groll [a], Regina Bones Barcellos [c,e], Raquel Maschmann [c], Leonardo Souza Esteves [e], Fernanda Spies [c], Rubia Raubach Trespach [a], Elis Regina Dalla Costa [e], Hermes Luís Neubauer de Amorim [f]

[a] Programa de Pós-Graduação em Biologia Celular e Molecular Aplicada a Saúde, Universidade Luterana do Brasil (ULBRA), Canoas, Rio Grande do Sul, Brazil
[b] Núcleo de Pesquisa em Microbiologia Médica (NUPEMM), Faculdade de Medicina, Universidade Federal do Rio Grande, Rio Grande, Rio Grande do Sul, Brazil
[c] Centro de Desenvolvimento Científico e Tecnológico (CDCT), Centro Estadual de Vigilância em Saúde, Secretaria Estadual da Saúde do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil
[d] Programa de Pós-Graduação em Biologia Celular e Molecular, Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, RS, Brazil
[e] Universidade Federal do Rio de Janeiro, Faculdade de Medicina, Programa Acadêmico de Tuberculose, Programa de Pós-Graduação em Clínica Médica, Rio de Janeiro, RJ, Brazil
[f] In Silico Solutions LTDA, Porto Alegre, RS, Brazil

ARTICLE INFO

Keywords:
Mycobacterium tuberculosis
Resistance
Whole genome sequencing
Rifampicin
Transmission

ABSTRACT

Here we described phenotypical, molecular and epidemiological features of a highly rifampicin-resistant *Mycobacterium tuberculosis* strain emerging in Southern Brazil, that carries an uncommon insertion of 12 nucleotides at the codon 435 in the *rpoB* gene. Employing a whole-genome sequencing-based study on drug-resistant *Mycobacterium tuberculosis* strains, we identified this emergent strain in 16 (9.19%) from 174 rifampicin-resistant clinical strains, all of them belonging to LAM RD115 sublineage. Nine of these 16 strains were available to minimum inhibitory concentration determination and for all of them was found a high rifampicin-resistance level (≥to 32 mg/L). This high resistance level could be explained by structural changes into the RIF binding site of RNA polymerase caused by the insertions, and consequent low-affinity interaction with rifampicin complex confirmed through protein modeling and molecular docking simulations. Epidemiological investigation showed that most of the individuals (56.25%) infected by the studied strains were prison inmate individuals or that spent some time in prison. The phylogenomic approach revealed that strains carrying on insertion belonged to same genomic cluster, evidencing a communal transmission chain involving inmate individuals and community. We stress the importance of tuberculosis genomic surveillance and introduction of measures to interrupt *Mycobacterium tuberculosis* transmission chain in this region.

1. Introduction

Tuberculosis (TB) remains a major public health emergency worldwide and current efforts to control the disease have been threatened by the high rates of drug-resistant TB (DR-TB) cases [1]. The overall success rate for TB treatment is 82%, but multidrug resistance (MDR-TB) - resistance to isoniazid (INH) and rifampicin (RIF), the two main anti-TB drugs - is associated with worse treatment outcome, dropping the cure rates to 60% [1]. RIF is one of the most important drugs used in the anti-TB treatment, due to its high bactericidal effects. RIF mechanism action consists of binding at RNA polymerase (RNAp) beta-subunit, encoded by *rpoB* gene, resulting in the inhibition of the bacterial mRNA transcription [2,3]. The occurrence of RIF-resistance in *Mycobacterium tuberculosis* (*M. tuberculosis*) is generally associated to single nucleotide substitutions at *rpoB* gene, mainly at rifampicin resistance determining region (RRDR), an 81-base pair region comprising from

* Corresponding author. Centro Estadual de Vigilância em Saúde, Secretaria Estadual da Saúde do Rio Grande do Sul, Avenida Ipiranga, 5400, Bairro Jardim Botânico, CEP, 90610-030, Porto Alegre, Rio Grande do Sul, Brazil.
E-mail addresses: richardsalvato@hotmail.com, richard.salvato@ufrgs.br (R.S. Salvato).
[1] MLR and PEAS contributed equally.

https://doi.org/10.1016/j.tube.2020.102015
Received 16 June 2020; Received in revised form 8 October 2020; Accepted 18 October 2020
Available online 30 October 2020

codon 426 to 452 [4].

Previous studies conducted in Rio Grande do Sul, a high TB burden setting and largest state in Southern Brazil, described an uncommon insertion of 12 nucleotides (12 nt) at the codon 435 (codon 516 in *Escherichia coli* based *rpoB* numbering system) in the *rpoB* gene among MDR *M. tuberculosis* strains. The 12 nt insertion results in a duplication of four amino acids (QNNP) and was firstly described by Perizzolo et al. (2012) [5]. Following studies analyzing MDR strains from Rio Grande do Sul [6] [–] [9] also reported strains carrying on the 12 nt insertion, and to our knowledge, this insertion was not described in any other region worldwide.

Hence, for an enhanced TB control management in Southern Brazil, is need the elucidation concerning molecular and phenotypic consequences of that 12 nt insertion, as well as, comprehension on this strain spreading into population. For this propose, we screened RIF-resistant *M. tuberculosis* strains from Rio Grande do Sul State using a WGS-based population study approach, in order to identify strains carrying on the 12 nt insertion. Afterward, we examine the phylogenetic relationship among these strains, the phenotypic consequences of the insertion into resistance level and biological cost, and also the *in silico* prediction to RIF binding effect.

2. Methods

2.1. Sample collection and drug susceptibility testing (DST)

In total 16 *M. tuberculosis* strains carrying on the 12 nt insertion at *rpoB* gene were identified on a WGS-based state-wide study that is being currently conducted in Rio Grande do Sul, Southern Brazil. The study included 174 RIF-resistant strains collected from 2011 to 2014, and 169 of them presented multidrug resistance. The *M. tuberculosis* clinical strains were from the State Central Laboratory (LACEN - Rio Grande do Sul), the reference laboratory in charge of performing drug susceptibility testing (DST) from TB cases notified statewide. In the studied period around 246 RIF-resistant cases were notified in Rio Grande do Sul according to the SITE-TB website (national database for DR-TB cases reported in Brazil). Thus, our analysis covered 70.73% (174/246) from RIF-resistant cases registered in the period.

The DST was performed at LACEN using the liquid BACTEC™ MGIT™ 960 SIRE Kit for the BACTEC Mycobacteria Growth Indicator Tube 960 (MGIT 960) system (Becton Dickinson Diagnostic Systems, Sparks, MD). The test was conducted for the following first-line anti-TB drugs: RIF (1.0 mg/L), INH (0.1 mg/L), ethambutol (EMB) (5.0 mg/L) and streptomycin (STR) (1.0 mg/L).

2.2. DNA extraction

M. tuberculosis genomic DNA was extracted from sputum culture in Lowenstein–Jensen solid medium using Cetyltrimethylammonium Bromide (CTAB) method, as described by Van Embden et al. (1993) [10].

2.3. Whole genome sequencing

The 174 RIF-resistant clinical strains belonging to our state-wide study were subjected to WGS. Paired-end sequencing (2 × 150 bp) was performed on an Illumina NextSeq machine using either a 300 cycle v2 mid output or high output kit (Illumina, Code FC-404-2003 or Code FC-404-2004) under standard Illumina® procedure as previously described [9]. The *rpoB* RRDR was also accessed for a subset of strains carrying on the 12 nt insertion by Sanger sequencing performed according previously described [8], the AB1 files are available in supplementary material.

2.4. Bioinformatics analysis

Raw reads were submitted to a routine pipeline to genomic variants identification. First, sequence reads were examined using FastQC (v0.11.7) (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and trimmed using Trimmomatic (v0.33) (parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36) [11]. Reads were mapped against *M. tuberculosis* H37Rv reference genome (GenBank accession number: NC_000962.3) under *BWA-MEM* (v0.7.16) algorithm. The quality of the resulting BAM file was checked using *Qualimap* [12] and *sambamba* (v0.6.8) was used to mark read duplicates [13]. *Samtools* (v1.9) [14] and GATK (v3.8) [15] were used to variant calling from sorted mapped sequences. Variants were filtered based on the following criteria: mapping quality ≥50, base alignment quality ≥23 and minimum read depth of 10. Variant functional annotation was performed with *snpEff* (v4.3) [16].

SNP-based *M. tuberculosis* lineages were determined using *TB-Profiler* [17,18] pipeline in command-line version (2.7.4). Phylogenetic analysis was performed using *snippy* pipeline v4.3.6 (https://github.com/tseemann/snippy) for variant calling and alignment of all core-genomes. Variants positions within PE/PPE genes or other repetitive regions associated with low mappability scores were removed. A maximum-likelihood phylogenetic tree was generated using the software RAxML (v8.10.12) [19], applying the generalized time-reversible (GTR) model and 1000 bootstrap replicates. The resulting tree was rooted using *M. canettii* (Genbank accession number: NC_019950.1). A minimum spanning tree was also generated using Phyloviz (http://online2.phyloviz.net/) being implemented the goeBURST algorithm [20]. *In silico* spoligotyping profiles were obtained using *SpoTyping* (v2.0) [21] and the spoligotyping patterns were assigned to lineage using the SITVIT2 web database (http://www.pasteur-guadeloupe.fr:8081/SITVIT2).

2.5. Minimum inhibitory concentration (MIC)

Due to the retrospective nature of this study, some strains were no longer available for additional phenotypic testing. Among the 16 strains carrying on the 12 nt insertion, nine of them were viable to perform susceptibility tests. The minimum inhibitory concentration (MIC) to RIF for these nine available strains was determined using the resazurin microtiter assay (REMA) as previously described [22,23].

2.6. Determination of growth curve by resazurin reduction method

The growth curve for two clinical strains with the 12 nt insertion, three wild type clinical strains and one susceptible control strain (H37Rv) was determined by resazurin reduction method as previously reported by von Groll et al. (2010) [24]. Briefly, cultures were started in 96-well plate containing the bacterial inoculum in triplicate wells as well as the 7H9 broth alone as the blank control. After 48 h incubated at 37 °C, resazurin (viability indicator) was added to all test and blank control wells and re-incubated. The resazurin reduction (the blue indicator color turn to pink) was assessed by measuring every 12 h the optical density (OD) at 620 nm with a plate reader (TECAN Spectrum Classic). Growth curves were constructed plotting the difference in OD between test and control wells every 12 h.

The biological cost of each strain was estimated by the growth index (GI) which was the time needed by each strain to reach an OD of 0.4 starting at an OD of 0.2. This time was calculated from the growth curve considering that all strains were in the logarithmic phase of growth between the two OD values. The biological cost was determined by the fitness relative (FR) which is the ratio of the GI of the insertion in each *rpoB* strains in relation to the H37Rv and in relation to the wild-type. In this case, a biological cost was considered if the FR was >1, since the 12 nt insertion has a slow growth in relation to the susceptible ones.

2.7. Protein insertion modeling and effects on drug binding

The wild-type RNAp (RNAp_wt) structural model was obtained from the crystallographic structure deposited in the RCSB Protein Data Bank

[25,26] (PDB) (www.rcsb.org) under accession code 5ZX3 (resolution 2.75 Å) [27]. The 3D structural models of the two mutants forms of the RNAp were obtained by homology modelling using the SWISS-MODEL Workspace[28]. Only the structures of the beta subunits of each mutant (*rpoB* INS1 and *rpoB* INS2) were modeled. The beta subunit of the structure under accession code 5ZX3 of PDB was used as template. The final structures for each RNAp mutant, here denominated RNAp_INS1 (carrying *rpoB* INS1) and RNAp_INS2 (carrying *rpoB* INS2), were obtained by fitting the coordinates of the modeled subunits with the coordinates of the beta subunit of the 5ZX3 crystallographic structure. The coordinates of the beta subunit of the template structure were subsequently removed, keeping the rest of the polymerase unchanged. The stereochemical quality of the models was checked by the QMEAN-DisCo[29] and QMEAN[30] scoring functions, implemented on the SWISS-MODEL platform [31], and with the analysis tools implemented on the MolProbity platform [32].

Docking simulations were carried out with two programs, AudoDock Vina (Vina) [33], version 1.1.2, and AutoDock [34], version 4.2. These programs use different algorithms and approaches to the searching mechanism and to the scoring function [35]. Reliability of the *in silico* methods increases when it is possible to employ different tools in order to apply a consensus for a given answer. By definition, *in silico* approaches are simulations of reality and therefore will never be completely accurate. When more than one algorithm and approach are combined, more accurate predictions are possible. If two programs that use different data and approaches are consistent, the reliability of the result is better. The errors of a given algorithm/approach must be different from the other algorithm/approach, thus being compensated as the results converge.

All simulations were carried out in triplicate. Targets (RNAp_wt, RNAp_INS1 and RNAp_INS2) and ligand (RIF) were prepared for docking simulations with the AutoDockTools (ADT) [36] interface, version 1.5.6. In all cases, the polymerase was treated as rigid and the ligand (RIF) as flexible. The grid box was centered on the Cδ of the Gln403 amino acid residue of the beta subunit of RNAp. Gasteiger [37] partial charges were calculated after addition of all hydrogens. Nonpolar hydrogens of enzyme and ligand were subsequently merged.

For Vina, a protocol consisting of a cubic box of the 21 × 21 × 21 Å was established. An exhaustiveness of 8 was chosen and the other parameters kept standard. The lowest docking-energy conformation of the set of poses generated by each docking simulation was chosen for analysis.

For AutoDock, a protocol consisting of a cubic box of 80 × 80 × 80 points with a spacing of 0.35 Å between the grid points was used. Global search Lamarckian genetic algorithm (LGA) [38] and local search (LS) pseudo-Solis and Wets [39] methods were applied in the docking search. Each single docking simulation consisted of 100 independent runs. The initial population was 150, the maximum number of generations was 27, 000 and the maximum number of energy evaluations was 2.5 × 106. Default values were selected for other parameters. The resulting docked conformations were clustered into families according to the RMSD. The lowest docking-energy conformation of the cluster with the lowest energy was chosen for analysis. All molecular figures were generated using Chimera software (Pettersen et al., 2004).

2.8. Ethics

This study was approved by the Research Ethics Committee of the Fundação Estadual de Produção e Pesquisa em Saúde (FEPPS/RS), protocol number 1.587.621 CAAE: 18269313.0.0000.5320.

3. Results

3.1. Population and phenotypic drug resistance

Among the 174 RIF-resistant strains from Rio Grande do Sul, collected from 2011 to 2014, we identified 16 (9.19%) strains carrying on a 12 nt insertion at codon 435 of the *rpoB* gene, all of them MDR. Amid those 16 MDR strains, one was additionally resistant to EMB and other to STR, according to DST results. Regarding the clinical and epidemiological characteristics, 14 were male, 14 previously treated for TB and seven HIV-infected. Besides that, unfavorable outcome was seen for 13 patients, the most of them due to treatment abandonment (7/16), failure to current treatment (5/16) and death in one case (Supplementary Table 1).

Nine individuals were prison inmates or spent some time in prison, and two individuals were homeless. The inmate individuals were from four different prison establishments, including the main state prisons: Cadeia Pública de Porto Alegre, Penitenciária Estadual do Jacuí and Penitenciária Modulada Estadual de Charqueadas.

3.2. Molecular characterization of the 12 nt insertion at rpoB gene

Analyzing the *rpoB* gene nucleotide sequence we identified a small difference in the occurrence of the 12 nt insertion. Overall, the 12 nt insertion occurred at the genomic position 761111 corresponding to codon 435 of the *rpoB* gene. However, 12 of the 16 strains with the insertion presented an additional change from adenine to guanine at the position 761110, resulting in the amino acid substitution at codon 435 (GAC > GGC, Asp435Gly). To an easy comprehension, we named the 12 nt insertion with additional polymorphism at genomic position 761110 of "*rpoB* INS1", and that with only the 12 nt insertion of "*rpoB* INS2" (Fig. 1). None of the strains with the 12 nt insertion had additional variants inside RRDR. Additionally, we performed Sanger sequencing for two strains carrying each of the both *rpoB* INS1 and *rpoB* INS2 and observed no discrepancies with NGS-based sequencing. In order to know the possible occurrence of this insertion in another settings worldwide, we performed a review of available data in the web based tools TBDreaMDB [40], MUBII-TB-DB [41], ReSeqTB [42] and studies that covered a large global dataset [43,44], but we did not find another described occurrence.

3.3. Genetic diversity and phylogenetic analysis

All strains presenting the 12 nt insertion were classified as belonging to *M. tuberculosis* lineage 4, sublineage 4.3.3 and to Euro-American family LAM by SNP-based typing. In the same way, all strains bear the RD115 deletion. *In silico* spoligotyping revealed that 15 strains shared the same spoligo pattern, belonging to Spoligo Internacional Type (SIT) 863, previously wrongly identified as *Mycobacterium pinnipedii* [6]. The remaining strain had absence of the 43 spacers within direct repeat (DR) locus and classified as ATYPIC (SIT 2669) on SITVIT2 database.

Besides that, phylogenetic inference including genomes from the seven *M. tuberculosis* lineages, Lineage 4 sublineages and other four SIT 863 strains with absence of the 12 nt insertion, showed the closeness of strains with 12 nt insertion with other SIT 863 strains (Fig. 2). Another important fact revealed by WGS-based phylogenomic is that applying a five SNPs threshold, the 16 strains were grouped into a unique genomic cluster, with a mean pairwise distance of 8.94 SNPs (Fig. 3).

Regarding resistance-related mutations in other genes, all the 16 strains had the Ser315Thr mutation at *katG* gene, three had variants associated to EMB resistance: one in the *embA* gene (−11C > A), one at *embB* (Met306Val) and the remaining a double mutation at *embB* gene (Gly406Asp and Met306Ile). Besides, one strain showed mutation related to ethionamide resistance (an insertion at codons 755/756 of the *ethA* gene).

3.4. In silico prediction of drug binding effect

We modeled the effect of both INS1 and INS2 on the RNAp structure in order to predict the consequence of these polymorphisms for interaction between RIF and the polymerase. At first, we performed the

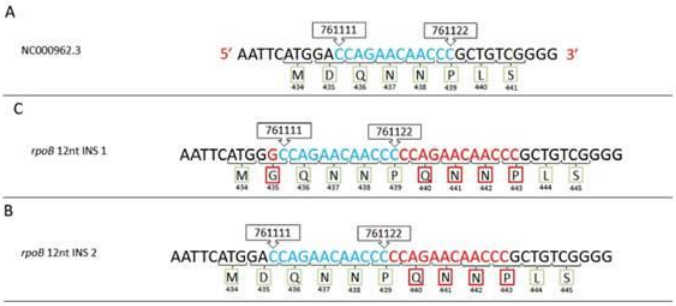


Fig. 1. - Illustration showing the occurrence of the 12 nt insertion. A) wild-type sequence, B) insertion *rpoB* INS1 and C) insertion *rpoB* INS2. The numbers above the sequence refer to the nucleotide genomic position. Bellow of the sequence is annotated the amino acid corresponding to each codon and the codon number on the *rpoB* gene. Colored in blue: nucleotides from wild-type sequence; green: amino acids from wild-type sequence; and red: nucleotides and amino acids resulting from 12 nt insertion.

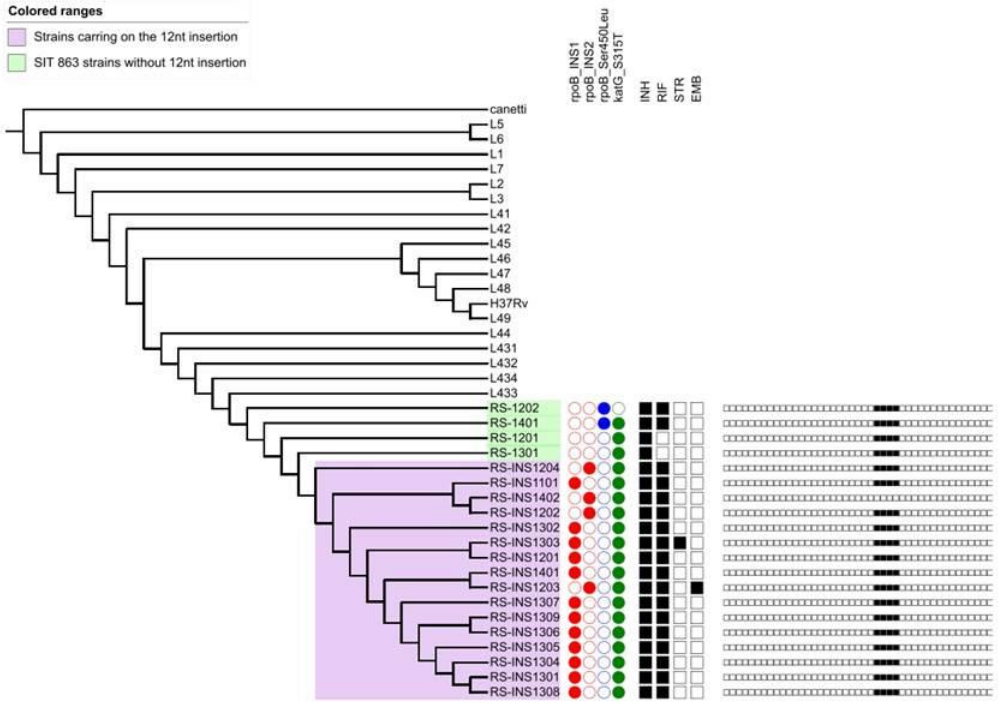


Fig. 2. - Core-genome phylogenetic tree of the 16 strains bearing the 12 nt insertion, four SIT863 strains without the insertion, and representative genomes from *M. tuberculosis* main lineages and Lineage 4 sublineages. The tree was based on 16,566 SNPs and rooted on *Mycobacterium cannettii*. The colored ranges in purple are the strains with the 12 nt insertion, and in green SIT863 strains without the insertion. The presence of mutations related to INH and RIF resistance is annotated with circles (presence: filled circle, absence: empty circle) and phenotypic drug susceptibility data for first-line drugs with the squares (resistant: filled square, susceptible: empty square). Spoligotyping profile is presented on small squares, indicating the presence (filled square) or absence (empty square) of the 43 spacer sequences.

homology modeling followed by model evaluation using stereochemical quality analysis tools (completely showed in Supplementary Information. With the homology models validated, we used docking programs to predict RIF interaction in the modeled structures. In Table 1, it is possible to observe that the docking into the RNAP_wt binding site indicates greater affinity for the RNAP_wt-RIF interaction (average values -8.6 ± 0.0 for Vina and -9.1 ± 0.0 for AutoDock) in comparison with RNAP_INS1-RIF interaction (average values -6.9 ± 0.0 for Vina and -5.4 ± 0.3 for AutoDock) and RNAP_INS2-RIF interaction (average values -6.7 ± 0.0 for Vina and -6.1 ± 0.1 for AutoDock). Fig. 4 allows evaluating the structural effect of insertions on the RIF binding site.

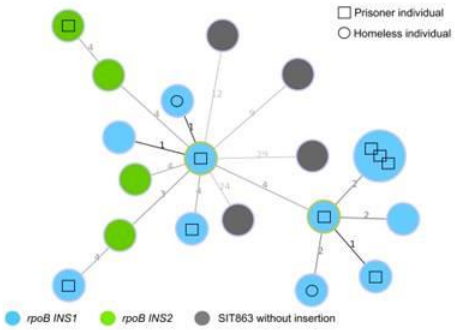


Fig. 3. - Minimum spanning tree for the 12 nt insertion and SIT863 without insertion strains, based on goeBURST algorithm. The circles representing strains from prison inmate individuals are identified with a square and with a circle are flagged the strains from homeless individuals. The color of each circle indicates the insertion type (blue: *rpoB* INS1; green: *rpoB* INS2; grey: SIT863 without insertion). The number in the lines between the circles refers to the core-genome distance amid the strains. The larger circle represents three strains sharing the same core-genome.

Table 1
- Binding scores of RIF interaction with wild-type, INS1 and INS2 RNAp.

Target	Run	Binding score Vina	Binding score Autodock
RNAP_wt	1	−8.6	−9.1
	2	−8.6	−9.1
	3	−8.6	−9.1
RNAP_INS1	1	−6.9	−5.2
	2	−6.9	−5.8
	3	−6.9	−5.3
RNAP_INS2	1	−6.7	−6.0
	2	−6.7	−6.2
	3	−6.7	−6.0

3.5. Resistance level and biological cost

All the nine strains bearing the 12 nt insertion and available for additional phenotypic tests presented a MIC value ≥ to 32 mg/L. The growth index average (GI) was determined for two individual clinical strains with the 12 nt insertion (GI = 38.5 h), for three wild type clinical strains (GI = 20 h) and in one susceptible control strain (H37Rv) (GI = 15). The insertion strains grow 1.95 and 2.5 slower than the wild type and H37Rv, respectively, showing a fitness disadvantage.

4. Discussion

Rio Grande do Sul is a high burden TB state in Southern Brazil with elevated rates of MDR-TB. The state has an estimated population of 11,3 million people [45] and in 2018 were accounted for 4541 new TB cases [46] and around 82 MDR diagnosed (data from LACEN-RS). Our group has been exploring *M. tuberculosis* genetic diversity and its transmission across Rio Grande do Sul State in the last two decades. In 2012, Perizzolo et al. [5] first described the presence of a 12 nt insertion at *rpoB* gene among multi-drug resistant strains collected between 2004 and 2006. In 2015, Dalla Costa et al. [6] described six strains presenting the same insertion, all of them belonging to the SIT 863 and isolated in 2006 and 2010. Latest studies in the region have shown an increased frequency of strains carrying on this insertion [7,9,47], flagging its spreading and ongoing transmission. Currently, we are performing a population-based study on drug-resistant *M. tuberculosis* strains from Rio Grande do Sul, isolated between 2011 and 2014, and identified 16 MDR strains harboring the 12 nt insertion.

WGS-based typing in concert to phylogenetic analysis corroborates with previous data [6] that associated the occurrence of the 12 nt insertion with strains belonging to SIT 863 that presents an important and restricted distribution in Rio Grande do Sul State [9,47]. According to Dalla Costa et al. [6], strains showing the SIT863 pattern, and previously wrongly identified as *Mycobacterium pinnipedii*, are in fact *M. tuberculosis* the Lineage 4, more specifically LAM family (lineage 4.3). Our results are in accordance with that: all strains were classified to sublineage 4.3.3 i.e. RD115 sublineage within the LAM subfamily. These data are also confirmed by phylogenetic inference here conducted that showed the genomic closeness between 12 nt insertion and SIT 863 without the insertion strains, as well as, with 4.3.3 sublineage strains. Although previous studies [5,47] related the presence of 12 nt insertion strains classified as other sublineages, the genotyping techniques previously used, such as RFLP, MIRU-VNTR and spoligotyping, could produce some erroneous classification.

Aiming to identify the transmission scenario among those strains we used a five SNP genome-wide cut-off to define recent transmission, and infer that strains within that maximum genomic distance are likely epidemiologically linked [48]. The 16 strains harbouring the 12 nt insertion at *rpoB* comprised a unique genomic cluster, indicative of recent and ongoing transmission of these strains in recent years. In addition, two SIT 863 strains without the 12 nt insertion had a pairwise distance of 12 SNPs or less from strains bearing the 12 nt insertion, corroborating with the previously cited relationships among these

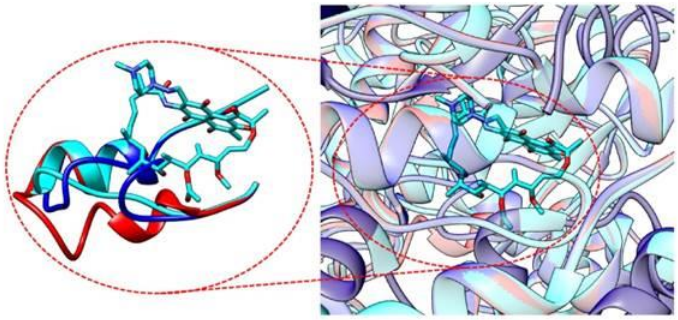


Fig. 4. - Schematic representation of the fitting of RNAP_wt (light blue), RNAP_INS1 (red) and RNAP_INS2 (navy blue) structures with emphasis on the RIF binding site. RIF is shown in its native position.

isolates.

The strains that presented the 12 nt insertion were mostly from inmate individuals (56.25%) from different prison establishments, including the three main state prisons that are widely overcrowded structurally limited establishments. Along to these imprisonment conditions, we observed among the individuals included in this study, the passage of the same inmate individual through different prison establishments in a short period of time, and this fact can be played an important role in the spreading of the 12 nt insertion strain. Besides that, the high number of individuals that abandoned the anti-TB treatment (43.8%) and other risk factors as the homeless condition can be listed as possible in charge for the substantial ongoing transmission of this strain. However, the strains from individuals without imprisonment history within the same cluster those strains from inmate individuals point toward a transmission chain involving the inmate population and the community. The fact that the inmate population may act as a reservoir feeding the TB transmission into the community is well established [49], and requires attention in the public health measures that aim to control the TB transmission in different settings.

In order to obtain a better understanding regarding the 12 nt insertion occurrence, we accessed *rpoB* gene nucleotide sequence and observed that 12 nt insertion occurs with a discrete variance amidst strains. Thus, we obtained RNAp structures carrying on both *rpoB* INS1 and *rpoB* INS2 structures by homology modeling, and docked those structures, as well as the RNAP wt structure with the RIF molecule. The decreased affinity of RIF for the RNAp_INS1 and RNAp_INS2 mutant forms may explain the resistance of the strains that carry the mutations. It is also possible to observe that the insertions promote changes in the binding site, which is probably linked to the lower affinity of RIF interaction with the mutant forms.

We explored the phenotypic impact of the 12 nt insertion concerning resistance level to RIF and to growth rate of the *M. tuberculosis* strains. Using the resistance level classification defined in the classical study of Huitric et al. [50], the nine clinical isolates tested showed a high level of resistance (MIC> 32 μg/mL) to RIF. As stated *in silico* analysis, the insertion of the 12 nucleotides probably causes a low affinity of RIF to its target in the β-subunit of RNA polymerase coding by the *rpoB* gene and reflects in the resistance level. In relation to the fitness, we observed that the two isolated tested presented a slower growth index when compared to wild type clinical isolates and the standard H37Rv. Considering that *rpoB* encode the β-subunit of the RNA polymerase enzyme, that is responsible for transcription from DNA to RNA, which is an essential process in the duplication of bacteria, a further change in its structure may have influenced in this process and consequent growth time. Previous studies showed biological cost associated to *rpoB* mutations [51, 52], however it could be mitigated with compensatory mutations [53]. It is important to point out that spite of the 12 nt insertion may cause a biological cost, is difficult to extrapolate this phenotypic characteristic to an epidemiologic setting, since it is an interplay of environmental, pathogen and host factors.

5. Conclusion

Here we described the key molecular and phenotypic aspects from a *M. tuberculosis* strain emerging in Southern Brazil in recent years that carries an uncommon 12 nt insertion at *rpoB* gene leading high-level RIF-resistance. The phylogenetic approach demonstrated a potential transmission from the prison inmate population to community of the 12 nt insertion strain, demonstrating an need for enhanced surveillance on TB transmission amid the inmate population and the anti-TB treatment abandonment, especially for MDR, in this region. Due to the high level of resistance to RIF - the most important drug in anti-TB treatment - and its established ongoing transmission, it is essential to monitor this strain spreading among the population. Since this is a retrospective study and included clinical samples collected until 2014, it is necessary to take up a more fast and, rather, real-time surveillance for this described strain in order to break up its transmission. For this propose, it would be necessary a rapid PCR-based test, for a fast, easy and low-cost identification of this strain.

Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), (grantnumber: 001), and supported by Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS), (grant number: 17/1265–8 INCT-TB).

Declaration of competing interest

The authors declare no competing interests.

Acknowledgments

We are grateful to TGen, C-Path and ReSeqTB for supporting whole genome sequencing and to Brazilian Network of Tuberculosis Research (REDE-TB) for enabling this partnership. The authors would like to thank Centro de Desenvolvimento Científico e Tecnológico (CDCT)/ Centro Estadual de Vigilância em Saúde/Secretaria Estadual da Saúde do Rio Grande do Sul for the support and infrastructure.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.tube.2020.102015.

Data availability

Mycobacterium tuberculosis genome data were deposited in the NCBI under BioProject IDs: PRJNA 535343 and PRJNA639713, see supplementary spreadsheet.

References

[1] WHO. Global tuberculosis report 2019. 2019.
[2] Cole ST. Rifamycin resistance in mycobacteria. Res Microbiol n.d.;147:48–52.
[3] Campbell EA, Korzheva N, Mustaev A, Murakami K, Nair S, Goldfarb A, et al. Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. Cell 2001;104:901–12. https://doi.org/10.1016/S0092-8674(01)00286-0.
[4] Almeida Da Silva PE, Palomino JC. Molecular basis and mechanisms of drug resistance in Mycobacterium tuberculosis: classical and new drugs. J Antimicrob Chemother 2011;66:1417–30. https://doi.org/10.1093/jac/dkr173.
[5] Perizzolo PF, Dalla Costa ER, Ribeiro AW, Spies FS, Ribeiro MO, Dias CF, et al. Characteristics of multidrug-resistant Mycobacterium tuberculosis in southern Brazil. Tuberculosis 2012;92:56–9. https://doi.org/10.1016/j.tube.2011.09.008.
[6] Dalla Costa ER, Vasconcelos SEG, Esteves LS, Gomes HM, Gomes LL, da Silva PA, et al. Multidrug-resistant Mycobacterium tuberculosis of the Latin American mediterranean lineage, wrongly identified as Mycobacterium pinnipedii (spoligotype International type 863 [SIT863]), causing active tuberculosis in South Brazil. J Clin Microbiol 2015;53:3805–11. https://doi.org/10.1128/JCM.02012-15.
[7] Esteves LS, Dalla Costa ER, Vasconcellos SEG, Vargas A, Ferreira Junior SLM, Halon ML, et al. Genetic diversity of Mycobacterium tuberculosis isoniazid monoresistant and multidrug-resistant in Rio Grande do Sul, a tuberculosis high-burden state in Brazil. Tuberculosis 2018;110:36–43. https://doi.org/10.1016/j.tube.2018.02.009.
[8] Salvato RS, Schiefelbein S, Barcellos RB, Praetzel BM, Anusca IS, Esteves LS, et al. Molecular characterisation of multidrug-resistant Mycobacterium tuberculosis isolates from a high-burden tuberculosis state in Brazil. Epidemiol Infect 2019;147. https://doi.org/10.1017/S0950268819001006.
[9] Salvato RS, Costa ERD, Reis AJ, Schiefelbein SH, Halon ML, Barcellos RB, et al. First insights into circulating XDR and pre-XDR Mycobacterium tuberculosis in Southern Brazil. Infect Genet Evol 2019:104127. https://doi.org/10.1016/j.meegid.2019.104127.
[10] Van Embden JDA, Cave MD, Crawford JT, Dale JW, Eisenach KD, Gicquel B, et al. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: recommendations for a standardized methodology. J Clin Microbiol 1993;31:406–9. https://doi.org/10.1128/JCM.39.4.1683.2001.
[11] Bolger AM, Lohse M, Usadel B. Genome analysis Trimmomatic: a flexible trimmer for Illumina sequence data 2014;30:2114–20. https://doi.org/10.1093/bioinformatics/btu170.

[12] Cruz LM, Garci F, Okonechnikov K, Tarazona S, Dopazo J, Meyer TF, et al. Qualimap: evaluating next-generation sequencing alignment data 2012;28:2678–9. https://doi.org/10.1093/bioinformatics/bts503.
[13] Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Genome analysis Sambamba: fast processing of NGS alignment formats 2015;31:2032–4. https://doi.org/10.5281/zenodo.13200.Contact.
[14] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools 2009;25:2078–9. https://doi.org/10.1093/bioinformatics/btp352.
[15] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–303. https://doi.org/10.1101/gr.107524.110.
[16] Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. 2012 landes bioscience . Do not distribute . © 2012 landes bioscience. Do not distribute; 2012. p. 80–92.
[17] Coll F, Mcnerney R, Preston MD, Guerra-assunção JA, Warry A, Hill-cawthorne G, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. Genome Med 2015;7:1–10. https://doi.org/10.1186/s13073-015-0164-0.
[18] Phelan JE, O'Sullivan DM, Machado D, Ramos J, Oppong YEA, Campino S, et al. Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. Genome Med n.d. doi:10.1186/s13073-019-0650-x.
[19] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30:1312–3. https://doi.org/10.1093/bioinformatics/btu033.
[20] Francisco AP, Bugalho M, Ramirez M, Carriço JA. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinf 2009;10:152. https://doi.org/10.1186/1471-2105-10-152.
[21] Xia E, Teo Y-Y, Ong RT-H. SpoTyping: fast and accurate in silico Mycobacterium spoligotyping from sequence reads. Genome Med 2016;8:19. https://doi.org/10.1186/s13073-016-0270-7.
[22] Palomino J-C, Martin A, Camacho M, Guerra H, Swings J, Portaels F. Resazurin microtiter assay plate: simple and inexpensive method for detection of drug resistance in Mycobacterium tuberculosis. Antimicrob Agents Chemother 2002;46:2720–2. https://doi.org/10.1128/aac.46.8.2720-2722.2002.
[23] Martin A, Palomino JC. Resazurin microtiter assay (REMA): procedure manual. 2009.
[24] von Groll A, Martin A, Portaels F, Silva PEA da, Palomino JC. Growth kinetics of Mycobacterium tuberculosis measured by quantitative resazurin reduction assay: a tool for fitness studies. Braz J Microbiol 2010;41:300–3. https://doi.org/10.1590/S1517-83822010000200006.
[25] Berman HM. The protein Data Bank. Nucleic Acids Res 2000;28:235–42. https://doi.org/10.1093/nar/28.1.235.
[26] Rose PW, Prlic A, Altunkaya A, Bi C, Bradley AR, Christie CH, et al. The RCSB protein data bank: Integrative view of protein, gene and 3D structural information. Nucleic Acids Res 2017;45:D271–81. https://doi.org/10.1093/nar/gkw1000.
[27] Li L, Fang C, Zhuang N, Wang T, Zhang Y. Structural basis for transcription initiation by bacterial ECF σ factors. Nat Commun 2019. https://doi.org/10.1038/s41467-019-09096-y.
[28] Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res 2014;42:W252–8. https://doi.org/10.1093/nar/gku340.
[29] Studer G, Rempfer C, Waterhouse AM, Gumienny R, Haas J, Schwede T. QMEANDisCo-distance constraints applied on model quality estimation n.d. doi:10.1093/bioinformatics/btz828.
[30] Benkert P, Biasini M, Schwede T. Toward the estimation of the absolute quality of individual protein structure models 2011;27:343–50. https://doi.org/10.1093/bioinformatics/btq662.
[31] Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, et al. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res 2018;46. https://doi.org/10.1093/nar/gky427.
[32] Williams CJ, Headd JJ, Moriarty NW, Prisant MG, Videau LL, Deis LN, et al. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci 2018;27:293–315. https://doi.org/10.1002/pro.3330.
[33] Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 2010;31:455–61. https://doi.org/10.1002/jcc.21334.
[34] Morris GM, Huey R, Lindstrom W, Sanner MF, Belew RK, Goodsell DS, et al. AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. J Comput Chem 2009;30:2785–91. https://doi.org/10.1002/jcc.21256.
[35] Forli S, Huey R, Pique ME, Sanner MF, Goodsell DS, Olson AJ. Computational protein-ligand docking and virtual drug screening with the AutoDock suite. Nat Protoc 2016. https://doi.org/10.1038/nprot.2016.051.
[36] Sanner MF. Python: a programming language for software integration and development. J Mol Graph Model 1999;17:57–61.
[37] Gasteiger J, Marsili M. Iterative partial equalization of orbital electronegativity?a rapid access to atomic charges. Tetrahedron 1980;36:3219–28. https://doi.org/10.1016/0040-4020(80)80168-2.
[38] Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, et al. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. J Comput Chem 1998;19:1639–62. https://doi.org/10.1002/(SICI)1096-987X(19981115)19:14<1639::AID-JCC10>3.0.CO;2-B.
[39] Solis FJ, Wets RJ-B. Minimization by random search techniques. Math Oper Res 1981;6:19–30. https://doi.org/10.1287/moor.6.1.19.
[40] Sandgren A, Strong M, Muthukrishnan P, Weiner BK, Church GM, Murray MB. Tuberculosis drug resistance mutation database. PLoS Med 2009;6:e1000002. https://doi.org/10.1371/journal.pmed.1000002.
[41] Flandrois JP, Lina G, Dumitrescu O, MUBII-TB-DB. A database of mutations associated with antibiotic resistance in Mycobacterium tuberculosis. BMC Bioinf 2014;15:107. https://doi.org/10.1186/1471-2105-15-107.
[42] Schito M, Dolinger DL. A collaborative approach for "ReSeq-ing" Mycobacterium tuberculosis drug resistance: convergence for drug and diagnostic developers. EBioMedicine 2015;2:1262–5. https://doi.org/10.1016/j.ebiom.2015.10.008.
[43] Coll F, Phelan J, Hill-Cawthorne GA, Nair MB, Mallard K, Ali S, et al. Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis. Nat Genet 2018;50:307–16. https://doi.org/10.1038/s41588-017-0029-0.
[44] Coll F, McNerney R, Preston MD, Guerra-Assunção JA, Warry A, Hill-Cawthorne G, et al. Rapid determination of anti-tuberculosis drug resistance from whole-genome sequences. Genome Med 2015;7:51. https://doi.org/10.1186/s13073-015-0164-0.
[45] No Title) n.d. https://cidades.ibge.gov.br/brasil/panorama (accessed December 9, 2019).
[46] Brasil M da S-S de V em S. Boletim Epidemiológico: Brasil Livre da Tuberculose: evolução dos cenários epidemiológicos e operacionais da doença. 2019.
[47] Salvato RS, Schiefelbein S, Barcellos RB, Praetzel BM, Anusca IS, Esteves LS, et al. Molecular characterisation of multidrug-resistant *Mycobacterium tuberculosis* isolates from a high-burden tuberculosis state in Brazil. Epidemiol Infect 2019;147:e216. https://doi.org/10.1017/S0950268819001006.
[48] Walker TM, Ip CLC, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, et al. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: a retrospective observational study. Lancet Infect Dis 2013;13:137–46. https://doi.org/10.1016/S1473-3099(12)70277-3.
[49] Sacchi FPC, Praca RM, Tatara MB, Simonsen V, Ferrazoli L, Croda MG, et al. Prisons as reservoir for community transmission of tuberculosis, Brazil. Emerg Infect Dis 2015;21:452–5. https://doi.org/10.3201/eid2103.140896.
[50] Huitric E, Werngren J, Juréen P, Hoffner S. Resistance levels and rpoB gene mutations among in vitro-selected rifampin-resistant Mycobacterium tuberculosis mutants. Antimicrob Agents Chemother 2006;50:2860–2. https://doi.org/10.1128/AAC.00303-06.
[51] Mariam DH, Mengistu Y, Hoffner SE, Andersson DI. Effect of rpoB mutations conferring rifampin resistance on fitness of Mycobacterium tuberculosis. Antimicrob Agents Chemother 2004;48:1289–94. https://doi.org/10.1128/AAC.48.4.1289-1294.2004.
[52] 80- Gagneux S, Long CD, Small PM, Van T, Schoolnik GK, Bohannan BJM. The competitive cost of antibiotic resistance in Mycobacterium tuberculosis. Science 2006;312. https://doi.org/10.1126/science.1124410. 1944–6.
[53] Comas I, Borrell S, Roetzer A, Rose G, Malla B, Kato-Maeda M, et al. Whole-genome sequencing of rifampicin-resistant Mycobacterium tuberculosis strains identifies compensatory mutations in RNA polymerase genes. Nat Genet 2012;44:106–10. https://doi.org/10.1038/ng.1038.

3.4 CAPÍTULO IV

O manuscrito que constitui este capítulo, intitulado **"Genomic epidemiology of *Mycobacterium tuberculosis* in Santa Catarina, Southern Brazil"** apresenta uma análise baseada em WGS da diversidade genômica de *M. tuberculosis* circulando em Santa Catarina.

O trabalho foi publicado na revista **Scientific Reports** (https://www.nature.com/srep/), com fator de impacto **JCR 2021 = 4.996**.

O manuscrito está disponível em: https://doi.org/10.1038/s41598-020-69755-9.

SCIENTIFIC REPORTS
nature research

Check for updates

OPEN

Genomic epidemiology of *Mycobacterium tuberculosis* in Santa Catarina, Southern Brazil

Mirela Verza[1,2,8✉], Mara Cristina Scheffer[2,8], Richard Steiner Salvato[3,4✉], Marcos André Schorner[2], Fernando Hartmann Barazzetti[2], Hanalydia de Melo Machado[2], Taiane Freitas Medeiros[2], Darcita Buerger Rovaris[5], Isabel Portugal[6], Miguel Viveiros[7], João Perdigão[6], Afrânio Kritski[1] & Maria Luiza Bazzo[2]

***Mycobacterium tuberculosis* (*M.tb*), the pathogen responsible for tuberculosis (TB) poses as the major cause of death among infectious diseases. The knowledge about the molecular diversity of *M.tb* enables the implementation of more effective surveillance and control measures and, nowadays, Whole Genome Sequencing (WGS) holds the potential to produce high-resolution epidemiological data in a high-throughput manner. Florianópolis, the state capital of Santa Catarina (SC) in south Brazil, shows a high TB incidence (46.0/100,000). Here we carried out a WGS-based evaluation of the *M.tb* strain diversity, drug-resistance and ongoing transmission in the capital metropolitan region. Resistance to isoniazid, rifampicin, streptomycin was identified respectively in 4.0% (n = 6), 2.0% (n = 3) and 1.3% (n = 2) of the 151 studied strains by WGS. Besides, resistance to pyrazinamide and ethambutol was detected in 0.7% (n = 1) and reistance to ethionamide and fluoroquinolone (FQ) in 1.3% (n = 2), while a single (0.7%) multidrug-resistant (MDR) strain was identified. SNP-based typing classified all isolates into *M.tb* Lineage 4, with high proportion of sublineages LAM (60.3%), T (16.4%) and Haarlem (7.9%). The average core-genome distance between isolates was 420.3 SNPs, with 43.7% of all isolates grouped across 22 genomic clusters thereby showing the presence of important ongoing TB transmission events. Most clusters were geographically distributed across the study setting which highlights the need for an urgent interruption of these large transmission chains. The data conveyed by this study shows the presence of important and uncontrolled TB transmission in the metropolitan area and provides precise data to support TB control measures in this region.**

Tuberculosis (TB) remains a major public health threat and is currently the tenth leading cause of death worldwide and the leading cause of death by a single infectious microorganism. In 2018, 1.5 million people died from TB and 10 million new cases are estimated to have occurred worldwide. Brazil, along with the Russian Federation, India, China and South Africa—the BRICS countries—account for more than 40% of the global TB disease burden in incidence and deaths, and about 58% of the global burden of drug-resistant TB. In Brazil, 90,527 TB cases were reported (45 per 100,000 population) in 2018[1]. While the state of Santa Catarina in Southern Brazil is characterized by an intermediate TB incidence rate—23.7 per 100,000 population, the state capital Florianópolis shows a higher incidence rate of 46.0 per 100,000 population along with high rates of TB-HIV co-infection[2].

[1]Programa Acadêmico de Tuberculose da Faculdade de Medicina e Complexo Hospitalar HUCFF-IDT, Universidade Federal do Rio de Janeiro (UFRJ), Rua Prof. Rodolpho Paulo Rocco, 255, Ilha do Fundão, Rio de Janeiro, RJ 21941-590, Brazil. [2]Laboratório de Biologia Molecular, Microbiologia e Sorologia, Centro de Ciências da Saúde, Universidade Federal de Santa Catarina (UFSC), Florianópolis, Santa Catarina, Brazil. [3]Programa de Pós-graduação em Biologia Celular e Molecular, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brazil. [4]Centro de Desenvolvimento Científico e Tecnológico (CDCT), Centro Estadual de Vigilância em Saúde, Secretaria Estadual da Saúde do Rio Grande do Sul, Av. Ipiranga, 5400, Porto Alegre, RS 90610-000, Brazil. [5]Setor de Bacteriologia da Tuberculose, Laboratório Central do Estado de Santa Catarina (LACEN-SC), Florianópolis, Santa Catarina, Brazil. [6]iMed.Ulisboa-Research Institute for Medicines, Faculdade de Farmácia, Universidade de Lisboa, Lisboa, Portugal. [7]Unidade de Microbiologia Médica, Global Health and Tropical Medicine, Instituto de Higiene e Medicina Tropical, Universidade Nova de Lisboa, Lisboa, Portugal. [8]These authors contributed equally: Mirela Verza and Mara Cristina Scheffer. ✉email: mimivz@gmail.com; richardsalvato@hotmail.com

Worldwide, the asymmetrical distribution of the disease is also linked with distinctive lineages and genotypes of the *Mycobacterium tuberculosis* (*M.tb*) Complex species, the etiological agent of TB[3]. Early strain typing methods greatly boosted our understanding of TB transmission dynamics while simultaneously providing additional tools and perspectives on the phylogeographic landscape of such strains at a global level[3]. These so-called classical typing methods, i.e., those based on the characterization of genetic repetitive elements, have been successful in informing public health interventions due to the inability of traditional contact tracing approaches to reconstruct complex transmission chains[4–6]. In fact, population-based typing of *M.tb* clinical isolates has been important to infer on the degree of recent infection and highly relevant in the identification of specific risk factors and geospatial hotspots for intensified TB transmission. Two classical examples of such studies are those conducted in San Francisco and New York in the late 1980s and early 1990s that enabled the correlation of specific ethnicities, drug resistance and/or residence in specific areas with recent TB transmission as assessed by genetic clustering based on profiles obtained with Restriction Fragment Length Polymorphisms with IS*6110*[6,7]. In the next decades, classical genotyping methods such as Spoligotyping and Mycobacterial Interspersed Repetitive Unit—Variable Number of Tandem Repeat (MIRU-VNTR) have been widely used for molecular epidemiology studies[8]. While these methods have been an integral part of several national TB control plans, most of the genomic diversity and distinctiveness of each isolate is overlooked by typing methods that are only able to interrogate a minor fraction of an isolate's genome[9–11]. In contrast, Whole Genome Sequencing (WGS) has emerged as the leading typing strategy to study the dissemination and transmission dynamics at a genome-wide SNP-based resolution that clearly outperforms classical typing methods and, coupled with molecular evolutionary approaches, also holds the potential to retrace the evolutionary history of locally circulating strains[12,13]. This level of resolution has enabled the reconstruction of the microevolutionary trajectory of circulating extensively drug resistant strains in Portugal and South Africa along with the breakdown into genomic clusters that are proposed to reflect epidemiological links between patients[14–16].

In Santa Catarina state, in Brazil, it is known that the *M.tb* population structure is mostly dominated by Latin American and Mediterranean (LAM) strains, followed by the "ill-defined" T strains as assessed by spoligotyping[17]. However, nothing is known regarding the genomic diversity associated with this region and reports using higher resolution typing methods are limited to 12-loci MIRU-VNTR[18]. To address this gap, herein we report the preliminary results from a genomic epidemiological study in the metropolitan area, aiming to unravel the *M.tb* genotypes circulating in this region, infer on existing and active transmission clusters in the community, and examine the distribution of drug resistance while exploring possible associations between unfavorable treatment outcomes and SNP-based lineages.

Results

Study sample and epidemiology. The study encompasses 151 patients diagnosed with TB that started on anti-TB treatment between May 2014 and May 2016 in the metropolitan area. Ninety-six (63.6%) patients were male and 49% of all patients were aged between 26 and 45 years. Regarding the risk factors and comorbidities, 13 (8.6%) patients had diabetes mellitus, 47 (31.1%) had alcohol abuse history, 85 (56.3%) were tobacco users and 47 (31.1%) illicit drug users. Twenty-seven (17.9%) patients were co-infected with human immunodeficiency virus (HIV) (Supplementary Table S1). In addition, nine (6.0%) individuals were homeless and four (2.6%) spent time in a prison establishment.

Drug resistance and molecular basis. Among the 151 isolates obtained for the study, genotypic-based prediction for drug resistance enabled the identification of six (4.0%) isoniazid resistant isolates, three (2.0%) with rifampicin resistance, two (1.3%) streptomycin resistant isolates and, resistance to pyrazinamide and ethambutol was detected in one isolate (0.7%). Besides, resistance to second-line drugs were detected for ethionamide (n = 2; 1.3%) and for the FQ (n = 2; 1.3%). One MDR isolate was identified (Fig. 1).

The detected mutations underpinning these genotypic-based predictions are listed in Table 1 with the Ser315Thr at the *kat*G gene being the most frequent mutation associated with isoniazid resistance (n = 4/6; 66.7%). Two additional *inhA* promoter mutations (*inhA* C-15T and T-8C) were putatively associated with isoniazid resistance and ethionamide resistance. Resistance to rifampicin was driven by two distinct *rpoB* mutations: Ser450Leu (n = 2; 1.3%) and Ser441Leu (n = 1; 0.7%). Two putative compensatory mutations were detected in the single MDR isolate found: *ahpC* – 48G>A and *rpoC* Phe452Ser[19]. The complete resistance-related mutation profile of all isolates is shown in the Supplementary spreadsheet.

Genomic population structure and phylogeny. All strains had been previously characterized by classical spoligotyping by membrane reverse-hybridization methods, revealing a population structure composed mainly of LAM strains (n = 91; 60.3%), followed by the ill-defined T strains (n = 25; 16.4%) and Haarlem (n = 12; 7.9%). In total, 38 different SITs were found; the SIT 216/LAM5 (13.2%) was the most prevalent, followed by SIT 42/LAM9 (7.3%), SIT 17/LAM2 (7.3%) and SIT 64/LAM6 (7.3%) (Fig. 1). As all clinical isolates were subsequently subjected to WGS, and consistent with the spoligotyping-based classification, all 151 isolates were classified as *M.tb* Lineage 4 (Euro-American) across 13 of its sub-lineages with sub-lineages 4.3.3 (30.5%), 4.3.4.1 (19.9%) and 4.3.4.2 (13.2%) being the most frequent. Moreover, screening for the *fbpC*[103] SNP marker for LAM confirmed the spoligotyping-based classification of all LAM strains while three isolates initially classified as belonging to the T family were found to bare *fbpC*[103] and, from a phylogenetic standpoint, should therefore be considered as LAM strains (SC61, SC115, SC134 [SIT823/T1]). Eleven additional isolates without clade assignment or of unknown profile on SITVIT2 also carry this marker (ten Orphans and one SIT2752) (Fig. 1). A genome-wide phylogeny based on 17,027 core SNPs displayed a topology congruent with the distribution of the different sub-lineages found where the LAM strains herein assessed based on the *fbpC*[103] polymorphism formed

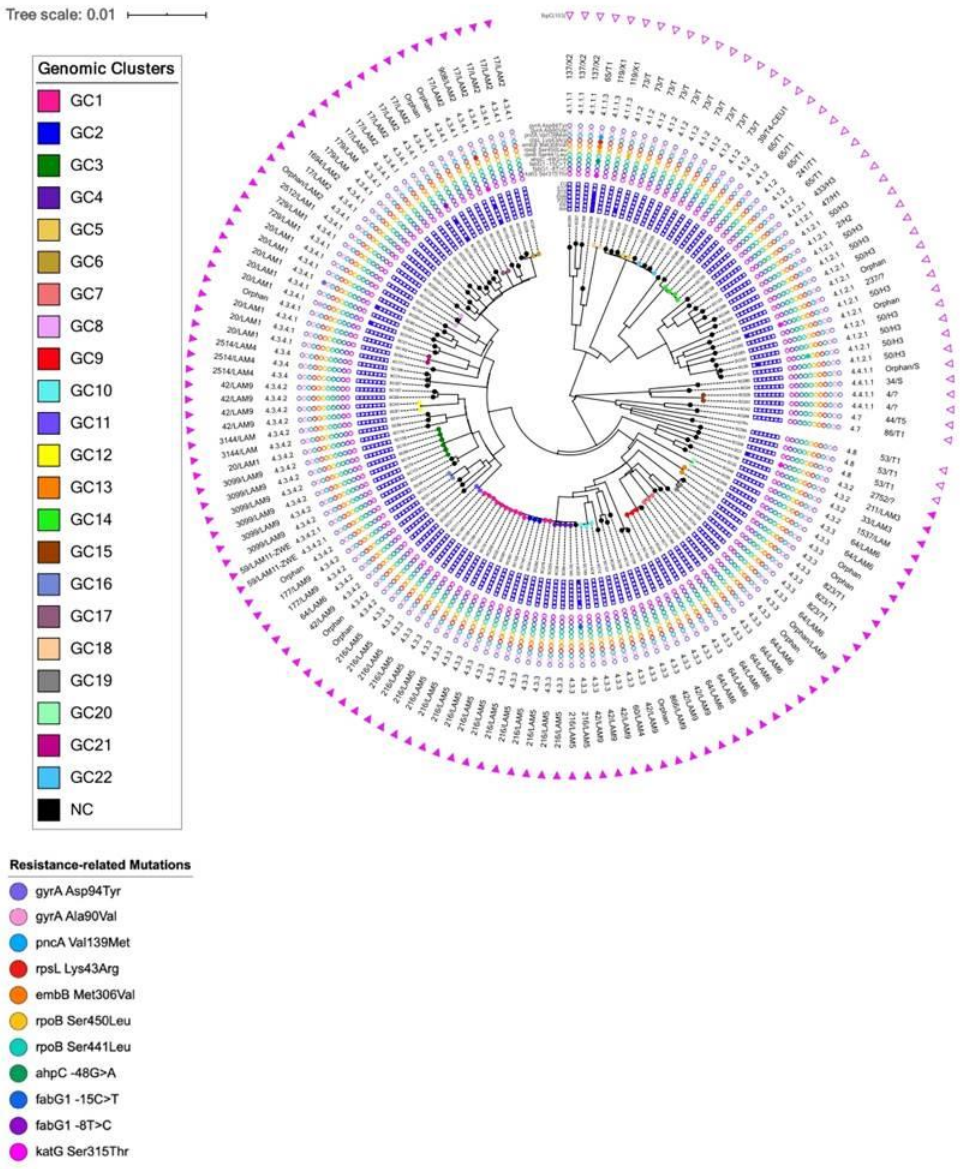


Figure 1. Maximum likelihood phylogenetic tree for the 151 *M.tb* isolates included in the study. This tree was constructed based on 17,027 core SNPs and is shown annotated with (from the outside to the inside): presence or absence of *fbpC*[103] LAM marker; SIT and spoligotyping clade; SNP-based sub-lineage; drug resistance associated mutations; and genotypic-based prediction for drug susceptibility. Tips are shown coloured according to the genomic cluster (see legend). *NC* not clustered, *GC* genomic cluster.

	Resistant strain	n	Related resistance mutations
Isoniazid	6 (4.0%)	4	*kat*G Ser315Thr
		1	*fab*G1 −15C>T
		1	*fab*G1 −8T>C
		1	*ahp*C −48G>A
Rifampicin	3 (2.0%)	1	*rpo*B Ser441Leu
		2	*rpo*B Ser450Leu
Pyrazinamide	1 (0.7%)	1	*pnc*A Val139Met
Ethambutol	1 (0.7%)	1	*emb*B Met306Val
Ethionamide	2 (1.3%)	1	*fab*G1 −15C>T
		1	*fab*G1 −8T>C
Streptomycin	2 (1.3%)	2	*rps*L Lys43Arg
Fluoroquinolones	2 (1.3%)	1	*gyr*A Asp94Tyr
		1	*gyr*A Ala90Val

Table 1. Mutations detected in genes associated with drug resistance across the 151 M.*tb* isolates included in the study.

a deeply rooted monophyletic clade. All drug resistant isolates detected were scattered across the phylogenetic tree suggesting independent emergence of drug resistance associated mutations.

SNP-based clustering and M.tb transmission. The average distance amid the core-genome of the 151 isolates was 420.3 SNPs. To delineate genomic clusters that might reflect recent transmission underpinned by putative epidemiologically linked patients, isolates within a maximum distance of 5 SNPs were assigned to genomic clusters (GC). Using this cutoff criterion, a total of 66 (43.7%) isolates were grouped into 22 clusters. The largest cluster (GC1) involved 12 isolates including one isolate obtained from a patient that spent time in a prison. The second largest cluster (GC3) involved six patients while the third largest clusters (GC14 and GC4) included four isolates each; GC14 with one isolate from an individual with incarceration history and GC4 including two epidemiologically-linked patients that shared the same household (pairwise distance: 0 SNPs). The remaining 18 genomic clusters included 42 isolates: one cluster included one isolate from an individual that spent time in a prison establishment; another cluster included a homeless patient; and, one other cluster was composed of two household contacts (pairwise distance: 0 SNPs; Supplementary Table S2). All clusters except GC1, GC8 and GC22 formed monophyletic branches in the phylogenetic tree suggesting further circulation and differentiation of strains within the same branch. The GC2 branch stemmed from within the GC1 clade. GC5 (n = 2) and GC22 (n = 2) isolate appear to have emerged from a broader late-branching phylogenetic SIT73/T (sub-lineage 4.1.2) clade that may represent a group of a mostly non-clustered strain that underwent extensive circulation in the community. A minimum spanning tree (MST) obtained for all isolates confirmed the integrity of all clusters highlighting a more extensive dissemination of GC1 isolates along with the observation of independently generated drug resistance associated nodes in the tree (Fig. 2, Supplementary Figure S1).

In the present sampling, the geographical distribution of patients across metropolitan region by residence location showed a higher concentration of cases among patients living at the central region of Florianópolis. Nevertheless, genome clustered cases did not show any particular pattern of geographical clustering as patients whose isolates were clustered showed a wide geographical dispersion in the studied region (Fig. 3). Comparing the pairwise geographic distances between patients' residence (Fig. 3, Supplementary Figure S2) no statistically significant difference was observed between non-clustered patients (average pairwise distance: 13.7 km; range 0–45 km) and patients within the seven largest clusters (n ≥ 3 patients) except for: patients in GC2 (average pairwise distance: 3.7 km; range 1–5 km) which showed a statistically lower pairwise distance ($p = 0.0302$, Wilcoxon Rank Sum Test); patients in GC4 (average pairwise distance: 2.2 km; range 0–5 km) which also showed a statistically lower average pairwise geographic distance ($p = 0.0004$, Wilcoxon Rank Sum Test); and, likewise for patients in GC10 (average pairwise distance: 2.2 km; range 0–5 km; $p = 0.0004$, Wilcoxon Rank Sum Test). The geographical distribution of the analysed cases is similar to the 263 cases reported for Greater Florianópolis in the same period (Supplementary Figure S3).

Treatment outcome and statistical association. According to the data collected, 107 (70.9%) patients were declared cured, treatment failure occurred in four (2.6%) patients and six (4.0%) patients died. Additionally, treatment dropout was observed for 33 (21.9%) patients and no information on the treatment outcome was available for one (0.7%) patient (Table1).

We found a statistically significant association between TB outcome and HIV coinfection ($p = 0.022$, Fisher's exact Test) since the coinfected patients had a statistically lower cure rate (75% vs 94.5%, respectively) and a higher death rate (18.7% vs 2.2%, respectively). No statistical association between TB outcome and the others risk factors (alcohol intake, illicit drugs usage and clustering) was found ($p \geq 0.05$; Table 2).

Regarding treatment dropout and risk factors, we found that: higher treatment dropout rates were statistically associated with excessive alcohol intake ($p = 0.009$, Chi-square Test) and HIV coinfection ($p = 0.009$, Chi-square Test); HIV coinfection was associated with illicit drug usage ($p \leq 0.001$, Chi-square Test), alcohol

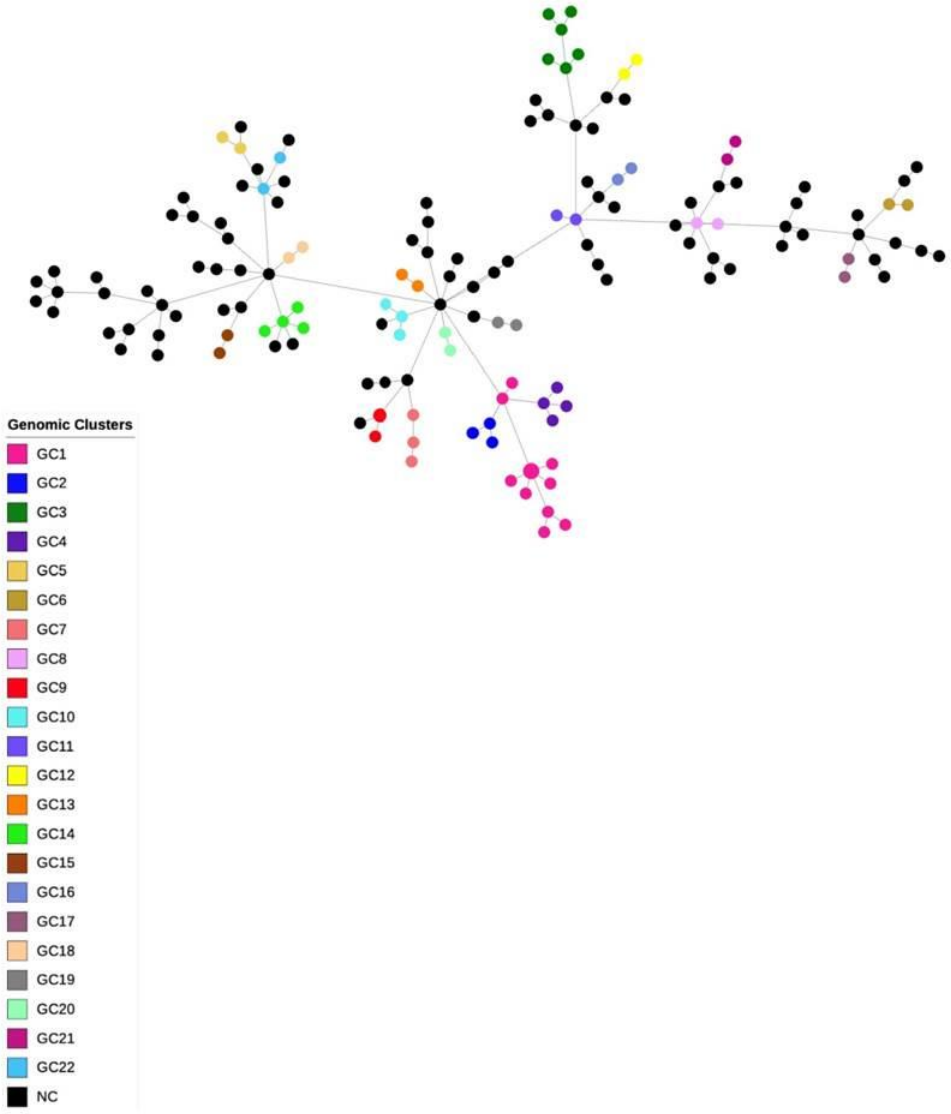


Figure 2. Minimum spanning tree (MST) of the 151 *M.tb* clinical isolates included in the present study. This MST is based on 17,027 core SNPs and nodes are shown coloured in function of the associated genomic cluster.

intake ($p \leq 0.001$, Chi-square test) and lower rates of hemoptysis ($p = 0.011$, Chi-square Test; Table 2). Illicit drug usage and excessive alcohol intake were also found to be associated ($p \leq 0.001$, Chi-square Test). No statistically significant association was found between clinical characteristics and SNP-based *M.tb* lineages. We tested also the hypothesis of a possible association between the outcomes and/or clinical characteristics and clustering, however, no association was found. Likewise, no significant association was found between risk factors and clustering (data not showed).

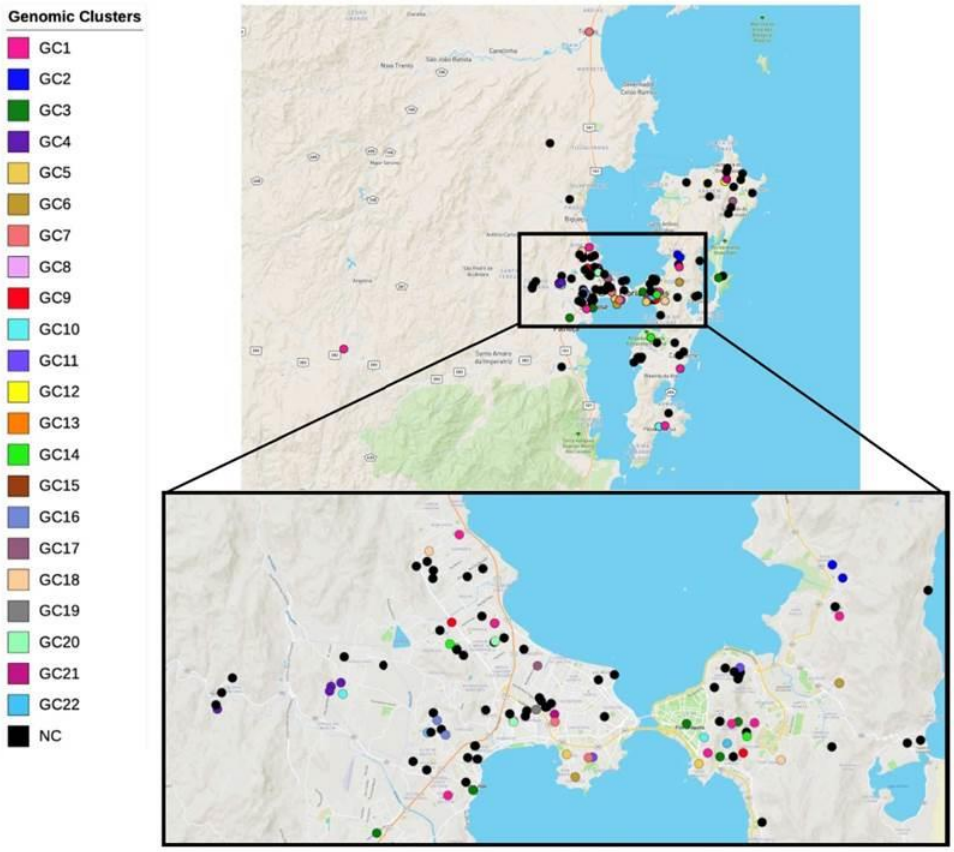


Figure 3. Geographic distribution of 142 out of the 151 TB cases (excluding 9 homeless) included in the study for which address of residence was available. These cases were geographically mapped according to the respective address of residence and are shown in the map colored according to the respective isolate's genomic cluster. The spatial distribution of these cases and correlation with the genomic cluster shows an extensive geographical spread of the larger clusters ($n \geq 3$) except for GC2, GC4 and GC10. The map was created using the online microreact tool, available at https://microreact.org/project/hrdK7GcztX8M2ZXEXjc579.

Discussion

This study evaluates the scenario of *M.tb* strain diversity between May-2014 and May-2016, analyzing individuals from the metropolitan region of Florianópolis diagnosed with pulmonary TB at Florianópolis and São José (the two most populous of the 22 cities that comprise the metropolitan region of Florianópolis). Despite the importance of knowing the molecular profile of circulating *M.tb* strains in a given setting, the existing data for this particular region is limited to the classical genotypic characterization of circulating strains[17,18,20]. The present study comprises the first genome-wide study on *M.tb* in Santa Catarina state. Greater Florianópolis has a higher TB incidence when compared to the overall incidence rate on the state level, an intermediate prevalence setting[21]. In this study sample we found a high rate of TB-HIV co-infection (19.71%) that is consistent with the rate expected for this region (22.6%) and is well above the national rate (9.4%)[22]. Co-infected patients were associated with a low cure rate and excessive consumption of alcohol and drugs. HIV-coinfection is a widely known risk factor for TB development since the risk for TB development is 19 times higher in the population living with HIV when compared with the rest of the population, leading to poorer outcomes and lower relapse-free cure rates[1,23]. Nonetheless, alcohol abuse or unhealthy alcohol usage is an increasingly acknowledged risk factor known to affect the outcome of TB treatment and a risk factor for TB treatment adherence[24,25]. Recent data obtained in Uganda and Kenya further demonstrate that alcohol use among HIV-infected individuals appears

	TB outcome					Treatment dropout				Hemoptysis			
Risk factor	Death	Cure	Treatment failure	Total	*p*	No	Yes	Total	*p*	No	Yes	Total	*p*
Alcohol intake													
No	3 (4.9%)	55 (90.2%)	3 (4.9%)	61		61 (83.6%)	12 (16.4%)	73		44 (63.8%)	25 (36.2%)	69	
Moderate	0 (0.0%)	23 (100.0%)	0 (0.0%)	23	0.230	23 (85.2%)	4 (14.8%)	27	**0.009**	17 (63.0%)	10 (37.0%)	27	0.813
Excessive	3 (13.6%)	18 (81.8%)	1 (4.6%)	22		22 (59.5%)	15 (40.5%)	37		23 (69.7%)	10 (30.3%)	33	
HIV coinfection													
Pos	3 (18.7%)	12 (75%)	1 (6.3%)	16	**0.022**	16 (59.3%)	11 (40.7%)	27	**0.009**	21 (87.5%)	3 (12.5%)	24	**0.011**
Neg	2 (2.2%)	85 (94.5%)	3 (3.3%)	90		90 (82.6%)	19 (17.4%)	109		63 (60.0%)	42 (40.0%)	105	
Illicit drug usage													
Yes	2 (8.3%)	22 (91.7%)	0 (0.0%)	24	0.464	24 (51.1%)	23 (48.9%)	47	**0.000**	32 (71.1%)	13 (28.9%)	45	0.290
No	3 (3.7%)	73 (91.3%)	4 (5.0%)	80		80 (90.9%)	8 (9.1%)	88		50 (61.7%)	31 (38.3%)	81	
Clustering													
Yes	2 (3.9%)	46 (90.2%)	3 (5.9%)	51	0.391	51 (78.5%)	14 (21.5%)	65	0.905	42 (66.7%)	21 (33.3%)	63	0.830
No	4 (6.1%)	61 (92.4%)	1 (1.5%)	66		66 (77.6%)	19 (22.4%)	85		50 (64.9%)	27 (35.1%)	77	

Table 2. Risk factors associated with TB outcome, treatment dropout and hemoptysis. Statistically significant associations are highlighted in bold.

to be associated with decreased viral suppression due to the lower diagnosis rate and lower likelihood of being on an anti-retroviral treatment regimen if already HIV diagnosed[26]. In this regard, an integrated approach to reduce unhealthy alcohol consumption or illicit drug usage may lead to better outcomes[27].

WGS data enabled the screening of drug-resistance conferring mutations on a genome-wide scale which allowed the identification of 10 (6.6%) isolates resistant to at least one anti-TB drug and one (0.7%) MDR-TB isolate. Despite the low MDR-TB rate in this study, two clinical isolates from different patients are predicted to be mono-resistant to the FQs due to two distinct high-confidence *gyrA* mutations for FQs resistance prediction[16,28]. FQ mono-resistance or among non-MDR isolates has been reported across multiple settings with varying ranges[29]. Recently Kim et al.[30] reported a 0.8% FQ resistance rate among non-MDR strains detected across multiple hospitals in South Korea while revealing an increasing trend over the last two decades. TB patients with prior FQ prescription to TB diagnosis, usually to treat community-acquired pneumonia, have a three-fold higher risk of having FQ-resistant TB[31]. Multiple FQ prescriptions, FQ prescription more than 60 days prior for TB diagnosis and for more than 10 days are associated with FQ-resistant TB[31,32]. This study shows a 1.3% FQ-resistance rate among non-MDR-TB patients for this setting driven by two independent mutational events. One limitation to the interpretation of the data lies in the fact that no data regarding previous FQ exposure was obtained for these two patients and, as such, we cannot exclude these from being primary FQ mono-resistant TB cases[33]. Nonetheless, the detection of these two FQ mono-resistant isolates warns against a possible excessive usage of FQs in the community and calls upon additional antimicrobial stewardship measures since to our knowledge these comprise the first two FQ mono-resistance cases to be reported in Brazil.

Regarding *M.tb* diversity, SNP-based typing classified all the 151 isolates as Euro-American lineage 4 strains. The Euro-American lineage predominance in Santa Catarina State and in the Southern Brazil[17,34], occurred due to migratory processes from Europe to South America that increased in the seventeenth century[35]. Therefore, LAM, T and Haarlem, the most common spoligotyping families identified in this study, were also found in other studies in Southern Brazil[17,20,34,36,37]. Our examination of the distribution of the fbpC[103] SNP, considered as a highly specific marker for the LAM family not only confirmed the identification of all spoligotyping-based identified LAM, but increased the frequency of this lineage to 69.5% (105/151). The three SIT823/T1 isolates, assigned to the "ill-defined" T family according to SITVIT2 are in fact phylogenetically positioned as LAM in the phylogenetic tree.

In accordance with previous studies[38,39], using a conservative five-SNP threshold to define genomic transmission clusters enabled the clustering of 43.7% of the isolates included in the study thereby demonstrating that approximately half of the TB cases in metropolitan region between 2014 and 2016 were due to recent transmission. One recent epidemiological transmission study conducted in the Pará State, in North Brazil, across household contacts did find two cases in the same household whose isolates are described as being 9 SNPs apart[40]. In this latter study, the route of transmission is unclear since the temporal distance between these isolates was 7 years and therefore questions if the transmission between these cases was direct and, if missing links are likely to exist. Herein, we opted to use a 5 SNP distance cutoff as to obtain high-confidence genomic clusters that are driven by recent transmission. Also, no association was found between clustering and risk factors, treatment dropout and clinical characteristics when using 12 or 25 SNPs cutoff distances to define clusters (data not shown).

The recent transmission scenario herein obtained lends support to a continued TB transmission, most likely still ongoing. A limitation to the study relies in the fact that only 151 (57.4%) cases were analyzed from a total of 263 reported cases for the same period in Greater Florianópolis and missing links in transmission chains are therefore likely to exist. However, the data conveyed has already enabled the detection of one large transmission cluster (GC1) that is responsible for 7.9% of the cases analyzed. All the isolates grouped into GC1 belonged to *M.tb* sublineage 4.3.3 and to SIT216/LAM5. Interestingly, SIT216/LAM5 has been described in Santa Catarina

state as the most prevalent SIT in prison establishments but only one isolate in GC1 had spent time in prison[18]. These facts highlight the epidemiological importance of SIT216/LAM5 clones, now at the community level, and its association with recent transmission. LAM5 spoligotype is also the most reported in Rio Grande do Sul, another state in Southern Brazil, however belonging to SIT93[34]. A parallel situation has been reported with SIT863 MDR-TB strains in Rio Grande do Sul, that was initially identified in prison establishments but is now responsible for the majority of MDR-TB cases in that state[37,41]. The presence of isolates from inmate individuals in the TB transmission chains supports the dissemination of strains between the general population and prison establishments and although the directionality is unclear at this point, the latter along with the entire prison system may act as reservoir of specific strains and promote a wider dispersion of specific clones[42]. Herein, the GC1 cluster showed a widespread geographical distribution when patient residence is considered (Supplementary Figure S2). Additionally, the low core-genome SNP distances among the isolates in these largest genomic clusters, may indicate that transmission has occurred in a very recent timeframe and is likely ongoing, suggesting that in order to prevent further spread of these strains, closer surveillance of these phylogenetically distinct clades is highly important.

Here we conducted the first WGS-based *M.tb* diversity study in SC state. Our study has some limitation, among them, for several samples phenotypic DST were not available, thus, we used WGS data to drug resistance prediction. Besides that, the present study only included new TB cases that started a first-line anti-TB treatment. Although the sampling in the present study occurred from 2014 to 2016, more recent data show an increase in the incidence TB rate (43.5% in 2016–46.0% in 2018)[2,22] and it would be interesting to evaluate if this is due to strains belonging to the main transmission chains herein identified. This fact reinforces the importance of molecular epidemiology as an instrument for TB surveillance and supporting public health measures.

Conclusions

The present study stresses the importance of WGS-based approaches that enable high-resolution phylogenetic analysis to investigate *M.tb* transmission in Brazil and when compared with classical genotyping techniques that may lead to overestimated clustering rates[43]. In this study we undertook a WGS-based analysis of the genetic diversity and recent *M.tb* transmission in a Southern Brazil region improving the knowledge about TB dynamics in this setting and filling out the lack of genomic data on *M.tb* strains circulating in the state of Santa Catarina. The data shows that uncontrolled TB transmission in the metropolitan region of Florianópolis occurred and provides precise data to support TB control measures in this region.

Materials and methods

Study design and population. A total of 151 M.*tb* complex strains obtained in the Central Laboratory of the State of Santa Catarina (LACEN), the reference laboratory for TB diagnosis in Santa Catarina were included in this study. The samples were from patients with bacteriological confirmation of pulmonary TB, diagnosed in health units at Florianópolis and São José cities (Florianópolis metropolitan region) within a 2-year period (May-2014–May-2016). The study included individuals that started the first-line anti-TB treatment and patients that met the inclusion criteria and agreed to participate were invited to sign the informed consent. Patient enrollment and medical record review were carried out in health units (Hospital Universitário da Universidade Federal de Santa Catarina and primary health units from Florianópolis and a specialized outpatient clinic responsible for TB care in São José city), and an additional epidemiological questionnaire was applied in an interview. Patients with previous anti-TB treatment history were excluded from the study.

The estimated population in the both cities is 700,000 inhabitants[44] and the TB incidence rate approximately 46 cases per 100,000 population[21]. During the study period, a total of 263 new pulmonary TB cases with positive culture were notified, of these, 218 were able to contact and accepted to participate in the study; 151 had DNA available for sequencing, representing 57.4% of the total (Supplementary Table S3). *Mycobacterium tuberculosis* was isolated from sputum samples obtained from patients before starting anti-TB treatment. The treatment outcomes were assessed upon treatment completion and defined as cure, treatment failure (persistence of smear microscopy and/or culture positives after the treatment), treatment dropout or death.

Smear microscopy and culture. The smear microscopy and Ziehl–Neelsen method were performed at the Central Laboratory of the State of Santa Catarina (LACEN), the reference laboratory for TB diagnosis in Santa Catarina according to the recommendations of the Recommendations Manual for TB control in Brazil[45]. Isolation *M.tb* complex was performed in Ogawa–Kudoh solid media according to the same Manual[45]. The identification to the *M.tb* complex level was carried by the presence of the cord factor and by MPT64 protein detection-based immunochromatographic test (SD Bioline Kit, Standard Diagnostics, Inc., Korea).

DNA extraction. The *M.tb* nucleic acids were extracted from mycobacterial cultures grown on Ogawa–Kudoh solid medium using the Cetyltrimethylammonium Bromide (CTAB) method, as described by van Soolingen et al.[46], at the Laboratory of Molecular Biology, Microbiology and Serology (LBMMS)-UFSC.

Spoligotyping. Spoligotyping was carried out as described by Kamerbeek et al.[9] using commercially prepared membrane (Ocimum Biosolutions, Hyderabad, Telangana, India). Hybridizing fragments were detected by chemiluminescence using peroxidase-labelled streptavidin and ECL detection kit (Amersham Biosciences, Amersham, Buckinghamshire, England, UK). The spoligotypes were classified according to the Spoligotyping International Type (SIT) and families, based on SITVIT2 database (https://www.pasteur-guadeloupe.fr:8081/SITVIT2/). The technique was performed by Scheffer during doctoral thesis[47].

Whole genome sequencing (WGS). The *M.tb* genomic DNA of 151 clinical strains was submitted to WGS. Approximately one microgram of DNA was fragmented using a Q800R2 sonicator (QSonica, Newtown, CT, USA) with the following parameters: 3 min sonication with 15 s pulse on, 15 s pulse off and 20% amplitude. The fragmented DNA was selected by size to target 600–700 bp by fragment separation using the Agencourt AMPure XP beads (Beckman Coulter, Code A63882). DNA Library preparation was performed using the NEBNext Ultra II DNA Library Prep Kit for Illumina (New England BioLabs, Code E7645L). The adapters and 8 bp index oligos based on Kozarewa and Turner[48] were purchased from IDT (Integrated DNA Technologies, San Diego, CA, USA) and used in place of those supplied in the NEB preparation kit in a dual-indexing approach[49]. Paired-end sequencing (2×150 bp) was performed on a NextSeq sequencer from Illumina using either a 300 cycle v2 mid output or high output kit (Illumina, Code FC-404-2003 or Code FC-404-2004) using standard Illumina procedure.

Bioinformatics analysis. Bioinformatic analysis of raw sequence reads was carried out initially using an in-house pipeline for genome-wide variant calling. Raw reads were trimmed to remove adapter sequences and low quality reads using *Trimmomatic* (v0.33) (parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36)[50] and the read quality control was performed using *FastQC* (v0.11.7) (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Trimmed reads were mapped to the *M.tb* H37Rv reference genome (GenBank accession number: NC_000962.3) using *BWA-MEM* (v0.7.16). *Samtools* (v1.9)[51] was used to convert from SAM to BAM format and sorting of mapped sequences. The quality of the resulting BAM file was checked using *Qualimap*[52] and *Sambamba* (v0.6.8) was used to mark read duplicates[53]. Variants (SNPs and small INDELs) were called using *Samtools*. Variants were filtered based on the following criteria: mapping quality ≥ 50, base alignment quality ≥ 23 and ≤ 2,000 reads covering each site. Variant functional annotation was performed with *SnpEff* (v4.3)[54]. WGS-based resistance prediction was performed using the command-line version of *TB-Profiler*[55] (v2.4), using the previously produced BAM files as input complementarily, VCF files were also screened for drug resistance associated variants and, if required, visually inspected in the BAM files. The definition of *M.tb* lineages based in SNP-typing method was performed according to the 62 SNPs barcode proposed by Coll et al. 2014[56] and implemented in *TB-Profiler*. The *fbp*C[103] polymorphism (G->A at codon 103) was used to differentiate LAM strains from non-LAM strains[57].

Phylogenetic analysis was performed using Snippy pipeline v4.3.6 (https://github.com/tseemann/snippy) for variant calling and alignment of all core SNP variants. SNP positions within PE/PPE genes or other repetitive regions associated with low mappability scores were removed from the final core-genome alignment, which was composed of 17,027 positions. A maximum-likelihood phylogenetic tree was generated using the PhyML, applying the generalized time reversible (GTR) model and branch support assessed by the approximate Likelihood Ratio Test (aLRT) as implemented in Seaview[58]. The resulting tree was rooted using *M. canettii* (Genbank accession number: NC_019950.1). A minimum spanning tree was generated using Phyloviz (v2.0) and the therein implemented goeBURST algorithm[59]. We used a 5, 12 and 25 SNPs cut-off to delineate genomic clusters[60] among the core SNP alignment using R along with the *ape* package and the *hclust* function. The patient address was plotted in a map using the online tool Microreact (www.microreact.org). Pairwise geographical distances between patients were assessed in R using the *Imap* package.

Statistical analysis. We tested the possible association between TB outcome (favorable: cure or treatment completion versus unfavorable: treatment failure or death), treatment dropout and hemoptysis with the following risk factors: alcohol intake, HIV coinfection, illicit drugs usage and clustering among the 151 individuals with available *M.tb* DNA. Besides, we tested the association among the risk factors. For this purpose, we applied the statistical tests Chi-square and Fisher's exact Test performed in IBM SPSS Statistics v.26.

Ethical approval. This study was approved by the Human Health Research Ethics Committee of Federal University of Santa Catarina (UFSC), protocol number: 2.054.560 CAAE: 66795917.1.0000.0121. Individuals who agreed to participate in the study signed the Informed Consent Term. The study was performed in accordance with relevant guidelines and regulations.

Bio-containment measures. Diagnosis activities including cultures of clinical specimens of the *M. tb* complex and DNA extraction were carried out under Biosafety Level 2 (BSL-2) containment with BSL-3 safety equipment and work practices.

Data availability

Mycobacterium tuberculosis genome data were deposited in the NCBI BioProject ID PRJNA599957 (see Supplementary spreadsheet).

Received: 9 April 2020; Accepted: 22 June 2020
Published online: 30 July 2020

References

1. WHO. *Global Tuberculosis Report 2019* (WHO, Geneva, 2019).
2. Brasil. Ministério da Saúde. Ministério da Saúde. Departamento de Vigilância de Doenças Transmissíveis. Boletim Epidemiológico: Brasil Livre da Tuberculose: Evolução dos cenários epidemiológicos e operacionais da doença. (2019).
3. Couvin, D., David, A., Zozio, T. & Rastogi, N. Macro-geographical specificities of the prevailing tuberculosis epidemic as seen through SITVIT2, an updated version of the *Mycobacterium tuberculosis* genotyping database. *Infect. Genet. Evol.* **72**, 31–43 (2019).

4. van Soolingen, D. *et al.* Molecular epidemiology of tuberculosis in The Netherlands: A nationwide study from 1993 through 1997. *J. Infect. Dis.* **180**, 726–736 (1999).
5. Golub, J. E. *et al.* Transmission of *Mycobacterium tuberculosis* through casual contact with an infectious case. *Arch. Intern. Med.* **161**, 2254–2258 (2001).
6. Small, P. M. *et al.* The epidemiology of tuberculosis in san francisco: A population-based study using conventional and molecular methods. *N. Engl. J. Med.* **330**, 1703–1709 (1994).
7. Alland, D. *et al.* Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. *N. Engl. J. Med.* **330**, 1710–1716 (1994).
8. Gagneux, S. Current topics in microbiology and immunology. *Curr. Top. Microbiol. Immunol.* **374**, 1–25 (2013).
9. Kamerbeek, J. *et al.* Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J. Clin. Microbiol.* **35**, 907–914 (1997).
10. 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).
11. Supply, P. *et al.* Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of *Mycobacterium tuberculosis*. *J. Clin. Microbiol.* **44**, 4498–4510 (2006).
12. Kostoff, D. & Kendall, J. A quantitative evaluation ofMIRU-VNTR typing against whole-genome sequencing for identifying *Mycobacterium tuberculosis* transmission: A prospective observational cohort study. *Genetica* **13**, 17–26 (2018).
13. Meehan, C. J. *et al.* Whole genome sequencing of *Mycobacterium tuberculosis*: Current standards and open issues. *Nat. Rev. Microbiol.* https://doi.org/10.1038/s41579-019-0214-5 (2019).
14. Walker, T. M. *et al.* A cluster of multidrug-resistant *Mycobacterium tuberculosis* among patients arriving in Europe from the Horn of Africa: A molecular epidemiological study. *Lancet. Infect. Dis.* **18**, 431–440 (2018).
15. Cohen, K. A. *et al.* Evolution of extensively drug-resistant tuberculosis over four decades: Whole genome sequencing and dating analysis of *Mycobacterium tuberculosis* isolates from KwaZulu-Natal. *PLoS Med.* **12**, 20 (2015).
16. Perdigão, J. *et al.* Using genomics to understand the origin and dispersion of multidrug and extensively drug resistant tuberculosis in Portugal. *Sci. Rep.* **10**, 2600 (2020).
17. Nogueira, C. L. *et al.* First insight into the molecular epidemiology of *Mycobacterium tuberculosis* in Santa Catarina, southern Brazil. *Tuberculosis* **97**, 57–64 (2016).
18. Medeiros, T. F. *et al.* Molecular epidemiology of *Mycobacterium tuberculosis* strains from prison populations in Santa Catarina, Southern Brazil. *Infect. Genet. Evol.* **58**, 34–39 (2018).
19. Perdigão, J. & Portugal, I. Genetics and roadblocks of drug resistant tuberculosis. *Infect. Genet. Evol.* **72**, 113–130 (2019).
20. Prim, R. I. *et al.* xMolecular profiling of drug resistant isolates of *Mycobacterium tuberculosis* in the state of Santa Catarina, southern Brazil. *Mem. Inst. Oswaldo Cruz* **110**, 618–623 (2015).
21. Brasil. Ministério da Saúde. TabNet Win32 3.0: TUBERCULOSE—Casos confirmados notificados no Sistema de Informação de Agravos de Notificação-Brasil. https://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinannet/cnv/tubercbr.def (2020).
22. Brasil. Ministério da Saúde. Ministério da Saúde. Departamento de Vigilância de Doenças Transmissíveis. Boletim Epidemiológico: Indicadores prioritários para o monitoramento do Plano Nacional pelo Fim da Tuberculose como Problema de Saúde Pública no Brasil (2017).
23. Magis-Escurra, C. *et al.* Treatment outcomes of MDR-TB and HIV co-infection in Europe. *Eur. Respir. J.* **49**, 20 (2017).
24. Fleming, M. F. *et al.* Alcohol and drug use disorders, HIV status and drug resistance in a sample of Russian TB patients. *Int. J. Tuberc. Lung Dis.* **10**, 565–570 (2006).
25. Chaves Torres, N. M., Quijano Rodríguez, J. J., Porras Andrade, P. S., Arriaga, M. B. & Netto, E. M. Factors predictive of the success of tuberculosis treatment: A systematic review with meta-analysis. *PLoS One* **14**, e0226507 (2019).
26. Puryear, S. B. *et al.* Associations between alcohol use and HIV care cascade outcomes among adults undergoing population-based HIV testing in East Africa. *AIDS* **34**, 405–413 (2020).
27. Sandgren, A., Lonnroth, K., Piret, V., Pierpaolo, de C. & Aljona, K. Collaborative action on tuberculosis and alcohol abuse in Estonia First report of a demonstration project. WHO Reg. Off. Eur. (2013).
28. Miotto, P. *et al.* A standardised method for interpreting the association between mutations and phenotypic drug resistance in *Mycobacterium tuberculosis*. *Eur. Respir. J.* **50**, 20 (2017).
29. Zignol, M. *et al.* Population-based resistance of *Mycobacterium tuberculosis* isolates to pyrazinamide and fluoroquinolones: Results from a multicountry surveillance project. *Lancet Infect. Dis.* **16**, 1185–1192 (2016).
30. Kim, H. *et al.* Trend of multidrug and fluoroquinolone resistance in *Mycobacterium tuberculosis* isolates from 2010 to 2014 in Korea: A multicenter study. *Korean J. Intern. Med.* **34**, 344–352 (2019).
31. Devasia, R. A. *et al.* Fluoroquinolone resistance in *Mycobacterium tuberculosis* the effect of duration and timing of fluoroquinolone exposure. *Am. J. Respir. Crit. Care Med.* https://doi.org/10.1164/rccm.200901-0146OC (2009).
32. Long, R. *et al.* Empirical treatment of community-acquired pneumonia and the development of fluoroquinolone-resistant tuberculosis. *Clin. Infect. Dis.* **48**, 1354–1360 (2009).
33. Wang, Z. *et al.* Molecular characteristics of ofloxacin mono-resistant *Mycobacterium tuberculosis* isolates from new and previously treated tuberculosis patients. *J. Clin. Lab. Anal.* **32**, 20 (2018).
34. Salvato, R. S. *et al.* First insights into circulating XDR and pre-XDR *Mycobacterium tuberculosis* in Southern Brazil. *Infect. Genet. Evol.* https://doi.org/10.1016/j.meegid.2019.104127 (2019).
35. Brynildsrud, O. B. *et al.* Global expansion of *Mycobacterium tuberculosis* lineage 4 shaped by colonial migration and local adaptation. *Sci. Adv.* **4**, eaat5869 (2018).
36. Salvato, R. S. *et al.* Molecular characterisation of multidrug-resistant *Mycobacterium tuberculosis* isolates from a high-burden tuberculosis state in Brazil. *Epidemiol. Infect.* **147**, e216 (2019).
37. Perdigão, J. *et al.* Clonal expansion across the seas as seen through CPLP-TB database: A joint effort in cataloguing *Mycobacterium tuberculosis* genetic diversity in Portuguese-speaking countries. *Infect. Genet. Evol.* https://doi.org/10.1016/j.meegid.2018.03.011 (2018).
38. Walker, T. M. *et al.* Assessment of *Mycobacterium tuberculosis* transmission in Oxfordshire, UK, 2007–12, with whole pathogen genome sequences: An observational study. *Lancet Respir. Med.* **2**, 285–292 (2014).
39. 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.* **6**, 1–8 (2016).
40. Conceição, E. C. *et al.* Analysis of potential household transmission events of tuberculosis in the city of Belem, Brazil. *Tuberculosis* https://doi.org/10.1016/j.tube.2018.09.011 (2018).
41. DallaCosta, E. R. *et al.* Multidrug-resistant *Mycobacterium tuberculosis* of the Latin American Mediterranean Lineage, Wrongly Identified as *Mycobacterium pinnipedii* (Spoligotype International Type 863 [SIT863]), causing active tuberculosis in South Brazil. *J. Clin. Microbiol.* **53**, 3805–3811 (2015).
42. Sacchi, F. P. C. *et al.* Prisons as reservoir for community transmission of tuberculosis, Brazil. *Emerg. Infect. Dis.* **21**, 452–455 (2015).
43. Stucki, D. *et al.* Standard genotyping overestimates transmission of *Mycobacterium tuberculosis* among immigrants in a low-incidence country. *J. Clin. Microbiol.* **54**, 1862–1870 (2016).
44. IBGE. IBGE cidades. https://cidades.ibge.gov.br/brasil/panorama (2020).

45. Brasil. Ministério da Saúde. Ministério da Saúde. Departamento de Vigilância de Doenças Transmissíveis. Manual de Recomendações para o Controle da Tuberculose no Brasil. (2019).
46. van Soolingen, D., Hermans, P. W., de Haas, P. E., Soll, D. R. & van Embden, J. D. Occurrence and stability of insertion sequences in *Mycobacterium tuberculosis* complex strains: Evaluation of an insertion sequence-dependent DNA polymorphism as a tool in the epidemiology of tuberculosis. *J. Clin. Microbiol.* **29**, 2578–2586 (1991).
47. Scheffer, M. C. Influência de fatores de risco do paciente e de características de *Mycobacterium tuberculosis* no desfecho dos casos novos de tuberculose pulmonar tratados com esquema básico: Uma coorte prospectiva de dois anos na Grande Florianópolis/SC. PhD Thesis. UFSC (2017).
48. Kozarewa, I. & Turner, D. J. High-throughput next generation sequencing. *Genome Biol.* **733**, 20 (2011).
49. Stone, N. E. *et al.* More than 50% of clostridium difficile isolates from pet dogs in Flagstaff, USA, carry toxigenic genotypes. *PLoS One* **11**, 1–21 (2016).
50. Bolger, A. M., Lohse, M. & Usadel, B. Genome analysis Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
51. Li, H. *et al.* The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
52. Cruz, L. M. *et al.* Qualimap: Evaluating next-generation sequencing alignment data. *Bioinformatics* **28**, 2678–2679 (2012).
53. Tarasov, A., Vilella, A. J., Cuppen, E., Nijman, I. J. & Prins, P. Genome analysis Sambamba: Fast processing of NGS alignment formats. *Bioinformatics* **31**, 2032–2034 (2015).
54. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)* **6**, 80–92 (2012).
55. Phelan, J. E. *et al.* Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. *Genome Med.* https://doi.org/10.1186/s13073-019-0650-x (2019).
56. Coll, F. *et al.* A robust SNP barcode for typing *Mycobacterium tuberculosis* complex strains. *Nat. Commun.* **5**, 4–8 (2014).
57. De Almeida, I. N. *et al.* Frequency of the *Mycobacterium tuberculosis* RDRio genotype and its association with multidrug-resistant tuberculosis. *BMC Infect. Dis.* **19**, 556 (2019).
58. Gouy, M., Guindon, S. & Gascuel, O. Sea view version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2010).
59. Francisco, A. P., Bugalho, M., Ramirez, M. & Carriço, J. A. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinform.* **10**, 152 (2009).
60. Roslanowski, A., Shelah, S. & Spinas, O. Nonproper products. *Bull. Lond. Math. Soc.* **44**, 299–310 (2012).

Acknowledgements
The authors are grateful to TGen, C-Path and ReSeqTB for supporting whole genome sequencing and to Brazilian Network of Tuberculosis Research (REDE-TB) for enabling this partnership. We would like to thank to Laboratory of Bacteriology of Tuberculosis from LACEN/SC for the support and infrastructure and to Centro de Desenvolvimento Cientifico e Tecnológico (CDCT)/CEVS/SES-RS for the infrastructure in bioinformatics analysis. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES), finance code: 001 and supported by Laboratory of Molecular Biology, Microbiology and Serology (LBMMS)-Federal University of Santa Catarina-UFSC.

Author contributions
M.V. and R.S.S. conducted the bioinformatic analysis, interpreted the results and wrote the first draft of the manuscript, M.C.S. designed the study, collected the clinical data, performed sample collection and processing and spoligotyping, M.A.S. assist in bioinformatic analysis, F.H.B. processed the samples, H.M.M. performed statistical analysis, T.F.M. helped with experimental assays, D.B.R. performed sample processing and cultures, J.P. bioinformatic analysis supervision, M.V., I.P., A.K., M.L.B. helped in the interpretation of the data, providing critical intellectual content and A.K., M.L.B. designed and supervised the study. All authors provided key edits, commented and approved the final version of the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-69755-9.

Correspondence and requests for materials should be addressed to M.V. or R.S.S.

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.

3.5 CAPÍTULO V

O manuscrito que constitui este capítulo, intitulado **"Genomic characterization of variants on mycolic acid metabolism genes in *Mycobacterium tuberculosis* isolates from Santa Catarina, Southern Brazil"** apresenta uma análise das mutações em genes relacionados ao metabolismo de ácidos micólicos em isolados de *M. Tuberculosis* de Santa Catarina.

O trabalho foi publicado na revista **Infection, Genetics and Evolution** (https://www.sciencedirect.com/journal/infection-genetics-and-evolution), com fator de impacto **JCR 2021 = 3.342** .

O manuscrito está disponível em: https://doi.org/ 10.1016/j.meegid.2021.105107.

Infection, Genetics and Evolution 96 (2021) 105107

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Genomic characterization of variants on mycolic acid metabolism genes in *Mycobacterium tuberculosis* isolates from Santa Catarina, Southern Brazil

Taiane Freitas Medeiros [a,b], Mara Cristina Scheffer [b], Mirela Verza [c], Richard Steiner Salvato [d,e], Marcos André Schörner [a,b], Fernando Hartmann Barazzetti [b], Darcita Buerger Rovaris [f], Maria Luiza Bazzo [a,b,*]

[a] Programa de Pós-graduação em Farmácia, Centro de Ciências da Saúde, Universidade Federal de Santa Catarina (UFSC), Florianópolis, SC, Brazil
[b] Laboratório de Biologia Molecular, Microbiologia e Sorologia, Centro de Ciências da Saúde, Universidade Federal de Santa Catarina (UFSC), Florianópolis, Santa Catarina, Brazil
[c] Programa de Pós-graduação em Clínica Médica, Faculdade de Medicina, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brazil
[d] Programa de Pós-graduação em Biologia Celular e Molecular, Universidade Federal do Rio Grande do Sul (UFRGS), Porto Alegre, Rio Grande do Sul, Brazil
[e] Centro de Desenvolvimento Científico e Tecnológico (CDCT), Centro Estadual de Vigilância em Saúde, Secretaria Estadual da Saúde do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil
[f] Setor de Bacteriologia da Tuberculose, Laboratório Central do Estado de Santa Catarina (LACEN-SC), Florianópolis, Santa Catarina, Brazil

ARTICLE INFO

Keywords:
Mycolic acid
Mycobacterium tuberculosis
Single nucleotide polymorphism

ABSTRACT

Mycobacterium tuberculosis has a complex cell wall containing mycolic acids (MA), which play an important role in pathogenesis, virulence, and survival by protecting the cell against harsh environments. Studies have shown that genes encoding enzymes involved in MA synthesis are essential to mycobacterial functionality. Here, we used whole-genome sequencing to evaluate mutations in genes related to MA metabolism in *M. tuberculosis* isolates from pulmonary tuberculosis patients of the Florianópolis Metropolitan Area, Santa Catarina, Brazil, and assessed associations with clinical, epidemiological, and genotypic data. The mutations Rv3057c Asp112Ala (104/151), Rv3720 His70Arg (104/151), and Rv3802c Val50Phe (105/151) were identified in about 69% of the isolates and were related to the LAM lineage. SIT 216/LAM5 (13.2%, 20/151) had the highest frequency and presented the mutations *accD2* Lys23Glu, *kasA* Gly269Ser, *mmaA4* Asn165Ser, *otsB1* Asp617Asn, *Rv3057c* Asp112Ala, *Rv3720* His70Arg, *Rv3802c* Val50Phe, and *tgs4* Ala216Glu. All SIT 73/T isolates (6.6%, 10/151) showed a characteristic and exclusive gene mutation pattern: *amiD* Rv3376 3790075G > A, *fbpA-aftB* 4266941G > A, *echA11* Asn220fs, and *otsB2* Ser110Arg. SITs 20/LAM1, 64/LAM6, 50/H3, 137/X2, and 119/X1 were also related to specific mutations. SITs from the LAM lineage differed in mutation profile from those of the T, Haarlem, and X lineages. Isolates from patients who had treatment failure showed mutations that do not seem to have a pattern related to this outcome. It was possible to identify a broad repertoire of single-nucleotide polymorphisms in genes related to MA metabolism in *M. tuberculosis* isolates. This study also described, for the first time, the variability between different SITs/sublineages of Lineage 4 circulating in Florianópolis Metropolitan Area.

1. Introduction

Tuberculosis (TB) is a communicable disease that constitutes a serious and challenging global public health problem. It is one of the top 10 causes of death worldwide and one of the most world's deadliest infectious diseases. In 2019, there were 10 million new cases and 1.4 million deaths from the disease (WHO, 2020); 73,864 new cases occurred in Brazil (35.0/100,000 inhabitants) (Brazil, 2008; Brazil, 2020). Santa Catarina State has a low TB incidence rate (23.7/100,000 inhabitants), but some municipalities exceed the national rate, such as the state capital, Florianópolis, with 44.5 cases per 100,000 inhabitants (Brazil, 2020).

M. tuberculosis, the major causative agent of the disease, has a complex cell wall, which contributes greatly to pathogenicity and cell

* Corresponding author at: Programa de Pós-graduação em Farmácia, Centro de Ciências da Saúde, Universidade Federal de Santa Catarina (UFSC), Florianópolis, SC, Brazil.
E-mail addresses: marialuizabazzo@gmail.com, m.l.bazzo@ufsc.br (M.L. Bazzo).

https://doi.org/10.1016/j.meegid.2021.105107
Received 28 March 2021; Received in revised form 24 September 2021; Accepted 5 October 2021
Available online 8 October 2021

integrity by protecting mycobacterial cells against environmental stress (Druszczynska et al., 2017; Fay et al., 2019). The wall is composed of a peptidoglycan layer covalently bound to arabinogalactans esterified with mycolic acids (MA) in the outer portion (Barrera, 2007; Guenin-Mace et al., 2009). MA are long-chain α-alkyl-β-hydroxy fatty acids. *M. tuberculosis* produces three major classes of MA by enzymatic modification of the meromycolate chain: α-, methoxy-, and keto-mycolic acids (Barkan et al., 2010; Barry 3rd et al., 1998; Kowalski et al., 2014; Pawełczyk and Kremer, 2014). Studies on genes encoding enzymes involved in MA synthesis have highlighted the importance of these molecules to mycobacterial functionality (Glickman et al., 2000). *M. tuberculosis* is part of a group of highly virulent bacteria with extremely low levels of genetic diversity (Achtman, 2008). This mycobacterium produces a wide variety of complex lipids, including several classes of MA, as evidenced by the high number of genes involved in their metabolism. Genes associated with *M. tuberculosis* physiology are believed to undergo strong selection; thus, it is expected that essential genes are conserved, such as those related to MA metabolism (Cole et al., 1998; Pepperell et al., 2013a, 2013b; Portevin et al., 2014). Given that one of the main drugs used in TB treatment, isoniazid, targets MA metabolism, it is possible that associated genes may be under diverse selection pressures (Portevin et al., 2014).

In this study, whole-genome sequencing (WGS) was used to determine, for the first time, mutations in MA metabolism genes in *M. tuberculosis* isolates from pulmonary TB patients in Florianópolis Metropolitan Area, Brazil, and their potential association with clinical outcomes, epidemiological features, and spoligotype international type (SIT) designation.

2. Materials and methods

2.1. Ethics statement

The research was approved by the Human Research Ethics Committee of the Federal University of Santa Catarina (UFSC) (protocol no. 2.054.560, CAAE no. 66795917.1.0000.0121). All individuals who agreed to participate signed an informed consent form before enrollment.

2.2. Patients and samples

This study analyzed 151 clinical isolates of *M. tuberculosis* obtained from patients with pulmonary TB diagnosed in the municipalities of Florianópolis and São José, Florianópolis Metropolitan Area, Santa Catarina, Brazil, between May 2014 and May 2016. Patients with a previous history of anti-TB treatment were excluded from the study. All isolates were collected prior to initiation of anti-TB therapy. Patient outcomes were evaluated after treatment completion and categorized as cure, treatment failure (positive sputum smear and culture after the fourth month of treatment), default, and death.

2.3. DNA extraction

DNA was extracted from isolated *M. tuberculosis* colonies by the cetyltrimethylammonium bromide (CTAB) method, according to van Soolingen et al. (1994).

2.4. Spoligotyping

Spoligotyping was performed as described by Kamerbeek et al. (1997). Spoligotypes were classified according to SIT number and lineage using the SITVIT2 database (http://www.pasteur-guadeloupe.fr:8081/SITVIT2/).

2.5. WGS analysis

Genomic DNA samples were subjected to WGS. For this, about 1 μg of DNA was fragmented on a Q800R2 sonicator (QSonica, Newtown, CT, USA) at 20% amplitude for 3 min with a pulse on/off time of 15/15 s. Fragmented DNA was purified and size selected to 600–700 bp using Agencourt AMPure XP beads (Beckman Coulter, catalog no. A63882). The DNA library was prepared using the NEBNext® Ultra™ II DNA library preparation kit for Illumina® (New England BioLabs, catalog no. E7645L). Adapters and 8 bp index oligos, based on Kozarewa and Turner (2011) and purchased from IDT® (Integrated DNA Technologies, San Diego, CA, USA), were used in place of those supplied with the NEB prep kit. A dual indexing approach was adopted. Paired-end sequencing (2 × 150 bp) was performed on an Illumina NextSeq system using a mid- or high-output v2 300-cycle kit (Illumina) and the standard Illumina® procedure.

2.6. Bioinformatic analysis

The FastQC tool (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/) was used to perform an initial quality assessment of reads. Adapter sequences were removed and low-quality reads were excluded using Trimmomatic version 0.33 with the following parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36 (Bolger et al., 2014). After trimming, quality control was performed again using FastQC, and reads were mapped to the reference genome (*M. tuberculosis* H37Rv, GenBank accession no. NC_000962.3) using BWA-MEM version 0.7.16. In the next step, SAMtools version 1.9 (Li et al., 2009) was used to convert SAM files to the BAM format and sort mapped sequences. The quality of the resulting BAM file was verified using Qualimap (García-Alcalde et al., 2012). Then, Sambamba version 0.6.8 was used to mark duplicate reads (Tarasov et al., 2015). Single-nucleotide polymorphisms (SNPs) and small insertions/deletions were identified using SAMtools. Variants were filtered based on the following criteria: mapping quality ≥50, base alignment quality ≥23, and ≤2000 readings at each site. Finally, raw VCF files were filtered using the following variant calling parameters: minimum read depth of 10 and maximum read depth of 2000. Functional annotation of VCF files was performed using SnpEff version 4.3 (Cingolani et al., 2012). This in-house pipeline was implemented to identify the nonsynonymous variants that were next tabulated for statistical analysis. The analyzed genes were as follows: *accA2*, *accA3*, *accD2*, *accD3*, *accD4*, *accD5*, *accD6*, *acpM*, *adhD*, *adhE1*, *amiD*, *bacA*, *cmaA1*, *cmaA2*, *desA1*, *desA2*, *desA3*, *echA10*, *echA11*, *fabD*, *fabD2*, *fabG1*, *fabG2*, *fabG4*, *fabH*, *fadD13*, *fadD32*, *fgd2*, *faz*, *fbpA*, *fbpB*, *fbpC*, *fbpD*, *fcoT*, *fgd2*, *hadA*, *hadB*, *hadC*, *inhA*, *irtA*, *irtB*, *kasA*, *kasB*, *lipR*, *mmaA1*, *mmaA2*, *mmaA3*, *mmaA4*, *mmpL11*, *mymA*, *mymT*, *otsB1*, *otsB2*, *pcaA*, *pks13*, *rip*, *Rv0161*, *Rv0194*, *Rv0519c*, *Rv0774c*, *Rv1272c*, *Rv1273c*, *Rv1686c*, *Rv1687c*, *Rv1747*, *Rv2509*, *Rv3057c*, *Rv3087*, *Rv3400*, *Rv3720*, *Rv3802c*, *sadH*, *tgs4*, *ufaA1*, *umaA*, and *virS* (Takayama, Wang and Besra, 2005; Lew et al., 2011; Portevin et al., 2014).

Phylogenetic analysis was performed using Snippy version 4.3.6 (https://github.com/tseemann/snippy) for variant calling and alignment of core SNPs based on the WGS of *M. tuberculosis*. Variants identified in *pe* and *ppe* genes or in other repetitive regions associated with low mappability were removed from the final core-genome alignment, resulting in 17,027 sites. A maximum-likelihood phylogenetic tree was constructed using the PhyML tool with the general time-reversible nucleotide substitution model. Branch support was assessed by the approximate likelihood ratio test (aLRT) in Seaview (Gouy, Guindon and Gascuel, 2010). The resulting tree was rooted with the common ancestor *Mycobacterium canetti* (GenBank accession no. NC_019950.1). We applied a 5 SNPs cut-off to identify genomic clusters among the core SNP alignment using R with ape package and hclust function. Finally, the online tool iTOL (https://itol.embl.de) was used to visualize and annotate the phylogenetic tree.

2.7. Statistical analysis

Statistical analysis was performed using IBM SPSS Statistics version 26.

3. Results

3.1. Study sample and epidemiology

Of the 151 patients included in this study, 96 (63.6%) were male. The age range was 15 to 83 years ($M = 37.13 \pm 13.98$ years). Twenty-seven (17.9%) individuals had HIV/TB coinfection. Other health-impacting factors were identified: 13 (8.6%) individuals had diabetes mellitus, 47 (31.1%) had a history of alcohol abuse, 85 (56.3%) were tobacco users, and 47 (31.1%) were illicit drug users. The treatment regimen used by all patients included isoniazid, rifampicin, pyrazinamide, and ethambutol. The first sputum smear test was positive in 92.0% (139/151) of the cases. A total of 89 (58.9%) individuals performed a sputum smear test in the second month, and, of these, 34 (38.2%) had a positive result. The fourth-month sputum smear, performed by 72 (47.7%) patients, was positive in 19.4% (14/72) of the cases. The outcome was cure in 70.9% (107/151), treatment dropout in 21.9% (33/151), death in 4.0% (6/151), and treatment failure (positive sputum smear test after four months) in 2.6% (4/151) of the cases. All strains included in this study were isoniazid-susceptible.

3.2. Genomic population structure

Spoligotyping of the 151 isolates allowed identification of 53 different spoligotypes: 136 isolates belonged to 38 distinct SITs, 15 isolates belonged to 13 new (orphan) profiles, and 4 isolates belonged to three unknown spoligotypes. Among the 53 spoligotypes, 5 families and 19 subfamilies were identified (H1, H2, H3, LAM, LAM1, LAM11-ZWE, LAM2, LAM3, LAM4, LAM5, LAM6, LAM9, S, T, T1, T4-CEU1, T5, X1, X2). The Latin American–Mediterranean (LAM) family was the most frequent, accounting for 60.3% (91/151) of all isolates, followed by T (16.6%, 25/151), Haarlem (H) (7.9%, 12/151), X (3.3%, 5/151), and S (1.3%, 2/151).

The most frequent subfamily was LAM9, representing 13.9% (21/151) of all isolates, followed by LAM5 (13.2%, 20/151), LAM2 (9.3%,14/151), T1 (8.6%, 13/151), LAM1 (7.9%, 12/151), LAM6 (7.3%, 11/151), H3 and T (6.6% each, 10/95), LAM (3.3%, 5/151), LAM4 (2.6%, 4/151), and X2 (2.0%, 3/151). LAM11-ZWE, LAM3, S, and X1 represented 1.3% (2/151) each, and H1, H2, T4-CEU1, and T5 represented 0.7% (1/151) each.

Verza et al. (2020) analyzed the polymorphism distribution of *fbpC*103 (a highly specific marker for the LAM family); interestingly, the marker confirmed the identity of all LAM isolates characterized by spoligotyping but classified three SIT823/T1 isolates (assigned to the ill-defined T family by SITVIT2) as LAM. The frequency of LAM isolates increased to 69.5% (105/151) considering the presence of this SNP as an

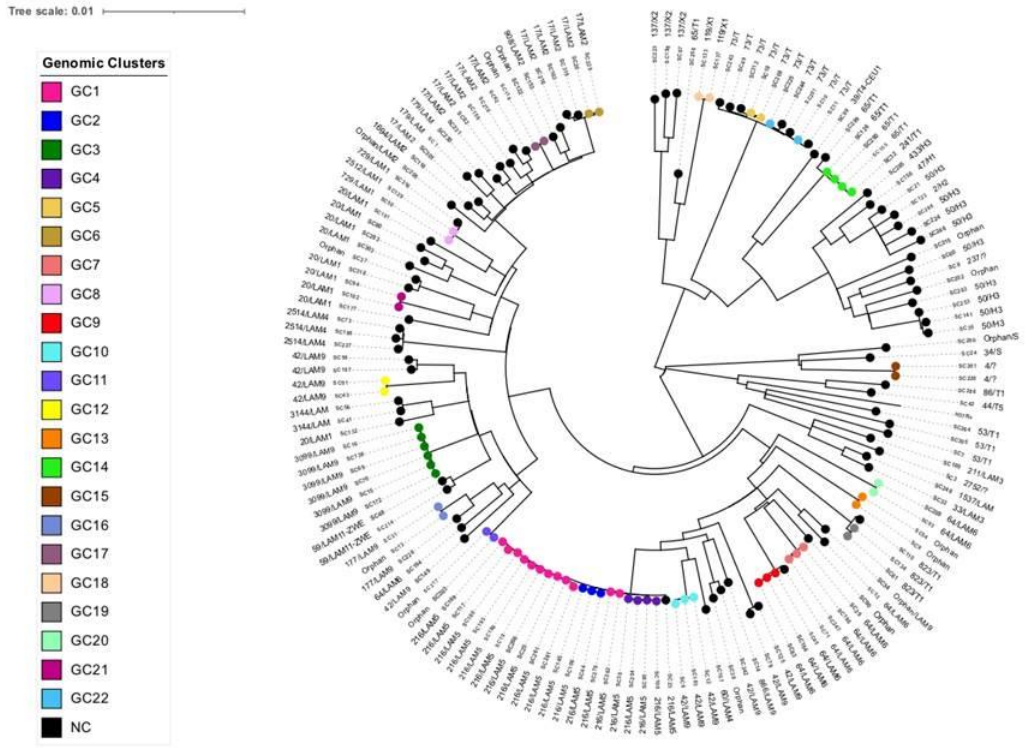


Fig. 1. Maximum-likelihood phylogenetic tree for the 151 *Mycobacterium tuberculosis* isolates included in the study. This tree was constructed based on 17,027 core SNPs and is shown annotated with (from the outside to the inside): SIT and spoligotyping clade; SNP-based sub-family. Tips are shown coloured according to the genomic cluster (see legend). Legend: NC, not clustered; GC, genomic cluster; LAM, Latin American–Mediterranean family; H, Haarlem family.

absolute marker for LAM family.

All clinical isolates were subjected to WGS. Consistent with spoligotyping results, the 151 isolates were classified as *M. tuberculosis* Lineage 4 (Euro-American).

SNP-based clustering was done from WGS data (Fig. 1). The average distance amid the core-genome of the 151 isolates was 420.3 SNPs. To delineate genomic clusters, isolates within a maximum distance of 5 SNPs were assigned to genomic clusters (GC). Using this cut-off criterion, a total of 66 (43.7%) isolates were grouped into 22 clusters. The largest cluster (GC1) involved 12 isolates identified as the SIT216 / LAM5 subfamily, one of these isolates, unlike the others, had no mutation in the tgs4 gene. The second largest cluster (GC3) involved six isolates of the SIT3099 / LAM9 subfamily. In the GC3, one isolate had one more mutation, in the aacD6 gene. The third largest cluster (GC14 and GC4) included four isolates each; GC14 included one isolate from SIT241/T1 subfamily and three isolates from SIT65 / T1 subfamily and GC4 cluster included four isolates from SIT216 / LAM5 subfamily. Mutations in MA genes did not diversify in this clusters (GC4 and GC14). The others 18 genomic clusters represented 42 isolates.

The three isolates characterized as SIT823/T1 by spoligotyping but identified as LAM upon considering the presence of *fbpC*^{103} SNP as an absolute marker for LAM family. These isolates were phylogenetically positioned as LAM in the phylogenetic tree.

3.3. Mutation in genes associated with mycolic acid metabolism

We identified mutations in 76 genes associated with mycolic acid metabolism. Analysis of the 151 isolates revealed 163 variations non-

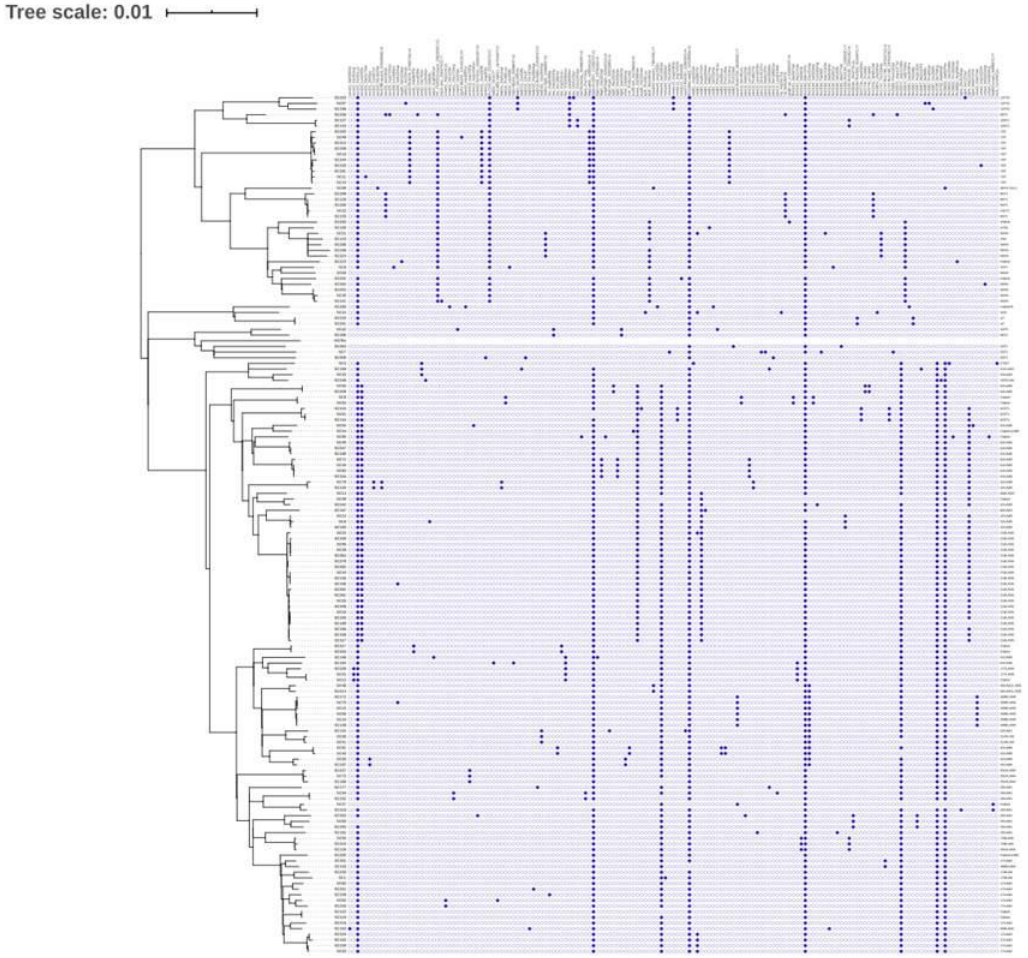


Fig. 2. Maximum-likelihood phylogenetic tree for the 151 isolates of *Mycobacterium tuberculosis* included in the study. This tree was constructed based on 17,027 core SNPs in genes related to mycolic acid metabolism and is shown annotated with (from the outside to the inside) SIT, sublineage, and gene mutations. Legend: LAM, Latin American–Mediterranean family; H, Haarlem family.

synonym, 134 of which were genic (Table S1, Supplementary Material) and 29 intergenic (Table S2, Supplementary Material). Fifteen genes were found to not carry mutations (*adhD*, *desA1*, *fabG2*, *fcoT*, *hadA*, *hadB*, *hadC*, *irtB*, *mmaA2*, *mymA*, *Rv1273c*, *Rv1686c*, *Rv3400*, *sadH*, and *ufaA1*). Phylogenetic analysis was carried out for all isolates, and trees were annotated with variants in genes associated with mycolic acid metabolism as well as SIT/sublineage data obtained by spoligotyping, as shown in Fig. 2.

The most frequent intergenic mutations were *Rv2005c-otsB1* 2251999A > G (147/151), and *fbpB-Rv1887* 2135870 T > C (142/151), and the most frequent genic mutation were *Rv0194* Met74Thr (147/151), and *accD2* Asn51Lys (143/151). These mutations were observed in all spoligotype families.

The three mutations found in about 69% of isolates, *Rv3057c* Asp112Ala (104/151), *Rv3720* His70Arg (104/151), and *Rv3802c* Val50Phe (105/151). In 104 isolates, these mutations occurred concomitantly, and in isolates exclusively on the family LAM.

When analyzing mutations in genes related to mycolic acid metabolism in specific SITs some data stood out.

The most common SIT was 216/LAM5 (13.2%, 20/151). All 20 isolates showed the same mutation pattern—12 mutations in 10 genes—beyond the mutations that to be observed in the most isolates *accD2* Ans51Lys (20/143), *fbpB-Rv1887* 2,135,870 T > C (20/142), *Rv2005c-otsB1* 2251999A > G (20/147), and *Rv0194* Met74Thr (20/147), as well were found the mutations that occurred exclusively in the LAM family: *accD2* Lys23Glu (20/47), *kasA* Gly269Ser (20/47), *mmaA4* Asn165Ser (20/97), *otsB1* Asp617Asn (20/27), *Rv3057c* Asp112Ala (20/101), *Rv3720* His70Arg (20/104), *Rv3802c* Val50Phe (20/105), and *tgs4* Ala216Glu (19/40). The mutations *fgd2* Ala122Val (2/2), *fgd2* Gly323Ser (4/4), *pks13* Lys725Glu (4/4), and *Rv0774c* Pro64fs (2/2) were observed exclusively in SIT 64/LAM6 isolates, *echA10* Asp79Gly (3/3) in isolates of LAM4 subfamily, *amiD* Ala74Ser (2/2) only in LAM3 subfamily, and the mutations *cmaA1* His173Arg (2/2), *fbpA* Pro77Arg (2/2), and *Rv0519c* Gln271Leu (3/3) exclusively in LAM1 subfamily.

Isolates belonging to SIT 73/T (6.6% - 10/151) showed 12 genic and intergenic mutations, with *amiD-Rv3376* 3790075G > A (10/10), *fbpA-aftB* 4266941G > A (10/10), *echA11* Asn220fs (10/10), and *otsB2* Ser110Arg (10/10) occurring exclusively in these 10 isolates. Furthermore, SIT 73/T isolates presented mutations as *PE_PGRS33-bacA* 2062805C > G (10/32) and *Rv2242-fabD* 2516567G > C (10/36), which were also detected in Haarlem, X, unknown, and orphan/new profiles, but not in LAM. Only one of the nine isolates classified as SIT 50/H3 did not show any of these mutations.

In the SIT 50/H3 isolates were observed the mutations *lipR* Ala156Val (7/13) and *Rv3057c* Val275Ala (8/14); both occurred exclusively in Haarlem and unknown lineages classified as Haarlem to by WGS.

SIT 137/X2 (2.0%; 3/151) showed *fabH* Gln182His (3/3), *fas* Ala2699Ser (3/6), *mmpL11* Arg449Gly (3/3), and Rv2242-fabD 2516567G > C (3/36). The mutation *fabH* Gln182His was found exclusively in SIT 137/X2. *fas* Ala2699Ser was also associated with SIT 119/X1 isolates (2/6); this mutation was exclusive to the X lineage and one isolate classified how SIT65/T1 by spoligotype method but classified by WGS how 'X-type (coll lineage 4.1.1.3).

When analyzing the families in the general, evaluating mutations in genes related to mycolic acid metabolism, it was possible to observe differences in the mutations pattern according to the family. The mutations that occurred in the LAM family were different than those that were observed in the T, Haarlem, and X families. The LAM family isolates were related with mutations in the genes *accD2*, *amiD*, *cmaA1*, *echA10*, *fbpA*, *fgd2*, *kasA*, *mmaA4*, *otsB1*, *pks13*, *tgs4*, *Rv0519c*, *Rv0774c*, *Rv3057c* (specifically Asp112Ala), *Rv3720*, and *Rv3802c*. While mutations in isolates of the T, Haarlem, and X families are related to genes *echA11*, *fabH*, *fas*, *lipR*, *mmpL11*, *otsB2*, and *Rv3057c* (specifically Val275Ala); and intergenic region *amiD-Rv3376*, *fbpA-aftB*, *PE_PGRS33-bacA*, *Rv2242-fabD*.

When evaluating isolates from patients who had treatment failure (2.6%, 4/151), the mutations *echA11* Asn220fs (2/10), *lipR* Pro43Leu (1/1), *otsB1* Ala386Thr (2/7), *otsB2* Ser110Arg (2/10), *rip* Arg351Trp (1/1), *Rv0194* Asp714Glu (1/1), *Rv1687c* Gly55Ser (1/1), *PE_PGRS33-bacA* 2062805C > G (3/31), *glyA1-desA2* 1221912C > A (1/1), *Rv2242-fabD* 2516567G > C (3/36), and *fbpA-aftB* 4266941G > A (2/10) were found. These mutations do not seem to have a pattern related to this outcome, the mutations that were found in more than one isolate may be related to the family to which the isolates were classified: T, Haarlem and S families.

The isolates SC37, SC60 and, SC80 showed a divergence of mutations in mycolic acid genes with their closest relatives. A very small number of mutations was observed in these isolates. Unfortunately, due to the lack of epidemiological data related to these patients, it was not possible to verify a possible relationship between mutations and this data.

No pattern of mutations or point mutations related to any epidemiological or laboratory data was observed.

4. Discussion

Using a genomic approach, in this study we investigated the mutation profile of genes involved with mycolic acid metabolism in circulating isolates of *M. tuberculosis* in Florianópolis and São José, Santa Catarina, Brazil. Together, the municipalities have about 700,000 inhabitants (IBGE, 2019). Driven by the importance of monitoring the molecular profile of *M. tuberculosis* circulating strains, this study sought to complement previous reports of *M. tuberculosis* genotypic characterization and, more recently, genomic epidemiology analysis in Santa Catarina State (Verza et al., 2020). This is the first study to investigate mutations in MA metabolism genes among *M. tuberculosis* isolates from Southern Brazil.

Although *M. tuberculosis* exhibits extremely low levels of genetic diversity (Achtman, 2008), we found that genes that are essential for mycobacterial physiology may be under diverse selection pressures. This may be expected, as MA metabolism is the target of most drugs used in TB treatment (Pepperell et al., 2013a, 2013b; Portevin et al., 2014). With WGS-based approaches, it is possible to detect microevolutions in genomic regions other than those known to be highly variable and commonly evaluated in traditional methods (Niemann and supply, 2014; Nikolayevskyy et al., 2019).

Portevin et al. (2014) performed a relative quantification of mycolic acid subtypes and analyzed genomic sequences of clinical isolates of the *M. tuberculosis* complex, including four major phylogenetic lineages (Lineages 1, 2, 4, and 6). Significant variations in mycolic acid patterns were found between lineages. The authors found that the patterns of ancient lineages contrasted with those of modern lineages and identified relevant SNPs possibly associated with specific mycolic acid patterns. In our study, the 151 isolates were classified as Euro-American (Lineage 4). Of the major lineages, Lineage 4 is the most frequent worldwide (Demay et al., 2012). Previous reports have shown that Lineage 4 shows genetic and phenotypic diversity, which may determine the epidemiology of its subtypes (Coscolla and Gagneux, 2014; Stucki et al., 2016). LAM, T, and Haarlem were the most frequent families in the current study, in agreement with the results of other studies conducted in southern Brazil (Nogueira et al., 2016; Prim et al., 2015).

The eight mutations identified in Lineage 4 by Portevin et al. (2014) led to amino acid changes, which were also detected in the clinical isolates evaluated in the present study. Six mutations—*accD2* Lys23Glu (46/151), *fas* Thr2595Asn (5/151), *kasA* Gly269Ser (46/151), *mmaA4* Asn165Ser (97/151), *Rv3057c* Asp112Ala (104/151), and *Rv3802c* Val50Phe (105/151)—were found more frequently in LAM isolates but were also observed in orphan isolates. In the current study, the mutation *fabH* Gln182His occurred exclusively in SIT 137/X2, and *Rv3057c* Val275Ala was identified only in isolates of the Haarlem lineage. The results observed in our study demonstrate an intralineage diversity as well as Portevin et al. (2014) report in their study, a substantial

intralinear diversity without in metabolism reported, who detected a high number of SNPs in two or more strains but not in all.

Our study highlighted trends in genomic profile related to mycolic acid metabolism. Similar profiles were shared by strains of the same SIT and/or family. Phylogenetic analysis showed that the mutation profile of LAM strains differed from those of T, Haarlem, and X strains. LAM isolates were mainly associated with mutations in genes related to MA structure. The gene *mmaA4* is involved in the introduction of the methyl branch and the hydroxyl group adjacent to the distal *cis* unsaturation; deletion of *mmaA4* inhibits the synthesis of methoxy- and keto-mycolic acids (Alahari et al., 2009; Coxon et al., 2013). In the present study, *mmaA4* was mutated in virtually all isolates of the LAM family. The gene has been reported as a determinant of intrinsic antibiotic resistance shared by the isoniazid, ethambutol, vancomycin and meropenem (Xu et al., 2017), representing an evolutionary advantage for the LAM family. The *kasA* gene encodes an enzyme involved in the initial elongation of MA *via* condensation (Bhatt et al., 2007; Kremer et al., 2002; Slayden and Barry 3rd., 2002). *kasA* Gly269Ser mutation resulted in loss of conformational stability and altered dynamic behavior in KasA (Jayaraman; Rajendra; Ramadas, 2019). Gande et al. (2004), demonstrated that the *accD2* gene is fundamental for the biosynthesis of AM, since it provides a carboxylated intermediate for the condensation reaction that is catalyzed by the protein produced from the *pks13* gene. This condensation will produce α-alkyl β-keto acids, precursors of MA (Gavalda et al., 2009). *fbpA* gene encodes a protein that catalyzes the transfer of mycolic acids to produce trehalose dimycolate (Goins et al., 2018) and acts as a fibronectin-binding protein, capable of binding to human macrophages during the initial stages of infection (Wilkinson et al., 2001; Romero et al., 2010). As well as *mmaA4*, it has also been reported as a determinant of intrinsic antibiotic resistance shared by the isoniazid, ethambutol and meropenem (Xu et al., 2017). Mutations related to the LAM subfamily have also been found in *fgd2* gene that encodes a dehydrogenase responsible for the oxidation of hydroxymycolic acid to keto-mycolic acid (Purwantini and Mukhopadhyay, 2013) and *cmaA1* that is associated with *cis*-cyclopropanation at the distal site of α-mycolates (Choudhury et al., 2015).

We identified mutations occurring exclusively in the 10 isolates classified as SIT 73/T. Most patients infected with SIT 73/T isolates had treatment failure. One of the mutations occurred in *otsB2*, a gene considered essential for *M. tuberculosis* growth, as it is involved in trehalose formation, an important structural constituent of the cell wall that acts as a mycolic acid carrier during cell wall synthesis (Murphy et al., 2005; Thanna and Suchecka, 2016). The *lipR* gene showed a mutation unique to Haarlem isolates and was associated with SIT 53/H3. This gene encodes *M. tuberculosis* lipase and is part of the *mymA* operon (Rv3083–Rv3089). The operon encodes proteins that can modify the cell envelope for intracellular survival (Cole et al., 1998; Kong et al., 2007). SIT 137/X2 isolates were associated with mutations in *fabH*, which encodes a protein that catalyzes the initial condensation reaction between acyl-CoA and malonyl-acyl carrier protein, leading to the formation of long-chain mycolic acids (Sachdeva and Reynolds, 2008), and *fas*, which encodes an enzyme that catalyzes the formation of fatty acids from acetyl-CoA, malonyl-CoA, and NADPH (Cole et al., 1998).

Unsuccessful TB treatments may be due to risk factors associated with the host (Costa-Veiga et al., 2017) as well as to factors related to mycobacteria. Sambandan et al. (2013) demonstrated that changes in the composition of cell wall mycolic acids are of particular importance in the context of TB treatment. For instance, changes in the methoxy−/keto-mycolic acid ratio can alter basal susceptibility to antimycobacterial drugs in the absence of drug resistance mechanisms. Mutations in genes related to mycolic acid metabolism have been described and, although none were directly related to treatment failure, one mutation occurred in a gene considered essential, the gene *rip*. The gene *rip* regulates cell envelope composition, growth, and persistence *in vivo* (Makinoshima; Glickman, 2005) and has been reported as a determinant of intrinsic antibiotic resistance shared by the rifampicin, vancomycin and meropenem (Xu et al., 2017).

A large number of intergenic mutations were found in this study and were shown to be related to specific SITs/sublineages. These mutations are little studied. In fact, intergenic SNPs are often called neutral SNPs, because, in general, their effect on phenotype is small. Nevertheless, SNPs in noncoding regions can have important functional consequences (Coscolla and Gagneux, 2014).

The variable outcome of the infection has been attributed mainly to the host and environmental factors because the genetic diversity of *M. tuberculosis* is less pronounced than in other pathogens. However, it is known that the diversity of the strain can play an important role during an infection. It has been determined that there are widely different metabolic phenotypes among strains of *M. tuberculosis* complex. Correlation studies of the genetic variation of *M. tuberculosis* complex with the phenotype have been successful in identifying clinically prevalent mutations that confer strong phenotypes, such as drug resistance (Hicks et al., 2018). The importance of relating genetic variation to more subtle phenotypes is highlighted, as in the study by Øyås et al. (2020) who managed to predict the effects of SNPs on genes that encode enzymes in metabolic phenotypes and identified desired SNPs that are associated with metabolic vulnerabilities. This increase the importance of studying the profile of mutations related to AM metabolism, which are correlated with more subtle phenotypes such as those related to adaptation to the host, transmissibility and basal resistance, which can be used as a possibility for the identification of more selective treatment strategy.

5. Conclusion

WGS allowed us to identify a broad repertoire of SNPs in genes related to mycolic acid metabolism in *M. tuberculosis* isolates and describe, for the first time, the variability between different SITs/sublineages of Lineage 4 circulating in Greater Florianópolis, Santa Catarina State, Brazil. Although it has been suggested that genetic background influences MA metabolism, further studies are needed to elucidate the extent to which genomic differences affect the phenotypic properties of *M. tuberculosis*, particularly with regard to antimicrobial resistance.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2021.105107.

Declaration of Competing Interest

None.

Acknowledgment

The authors are grateful to TGen, C-Path and ReSeqTB for supporting whole genome sequencing.

References

Achtman, M., 2008. Evolution, population structure, and phylogeography of genetically monomorphic bacterial pathogens. Annu. Rev. Microbiol. https://doi.org/10.1146/annurev.micro.62.081307.162832.

Alahari, A., Alibaud, L., Trivelli, X., Gupta, R., Lamichhane, G., Reynolds, R.C., Bishai, W. R., Guerardel, Y., Kremer, L., 2009. Mycolic acid methyltransferase, MmaA4, is necessary for thiacetazone susceptibility in *Mycobacterium tuberculosis*. Mol. Microbiol. https://doi.org/10.1111/j.1365-2958.2009.06604.x.

Barkan, D., Rao, V., Sukenick, G.D., Glickman, M.S., 2010. Redundant function of *cmaA2* and *mmaA2* in *Mycobacterium tuberculosis* cis Cyclopropanation of oxygenated Mycolates. J. Bacteriol. https://doi.org/10.1128/JB.00312-10.

Barrera, L., 2007. Chapter 3: The basics of clinical bacteriology. In: Palomino, J.C., Leao, S.C., Ritacco, V. (Eds.), Tuberculosis. From basic science to patient care.

Barry 3rd, C.E., Lee, R.E., Mdluli, K., Sampson, A.E., Schroeder, B.G., Slayden, R.A., Yuan, Y., 1998. Mycolic acids: structure, biosynthesis and physiological functions. Prog. Lipid Res. https://doi.org/10.1016/S0163-7827(98)00008-3.

Bhatt, A., Molle, V., Besr, G.S., Jacobs, W.R., Kremer, L., 2007. The *Mycobacterium tuberculosis* FAS-II condensing enzymes: their role in mycolic acid biosynthesis, acid-fastness, pathogenesis and in future drug development. Mol. Microbiol. https://doi.org/10.1111/j.1365-2958.2007.05761.x.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. https://doi.org/10.1093/bioinformatics/btu170.

Brazil, 2020. Ministério da Saúde. Secretaria de Vigilância em Saúde. Boletim Epidemiológico: Tuberculose 2020. http://www.aids.gov.br/pt-br/pub/2020/boletim-epidemiologico-de-turbeculose-2020. Accessed date: May 2020.
Choudhury, C., Priyakumar, U.D., Sastry, G.N., 2015. Dynamics based pharmacophore models for screening potential inhibitors of mycobacterial cyclopropane synthase. J. Chem. Inf. Model. https://doi.org/10.1021/ci500737b.
Cingolani, P., Platts, A., Wang Le, L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X., Ruden, D.M., 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly. https://doi.org/10.4161/fly.19695.
Cole, S.T., Brosch, R., Parkill, J., Garnier, T., Churcher, C., Harris, D., Gordon, S.V., Eiglmeier, K., Gas, S., Barry 3rd, C.E., Tekaia, F., Badcock, K., Basham, D., Brown, D., Chillingworth, T., Connor, R., Davies, R., Devlin, K., Feltwell, T., Gentles, S., Hamlin, N., Holroyd, S., Hornsby, T., Jagels, K., Krogh, A., Mc Lean, J., Moule, S., Murphy, L., Oliver, K., Osborne, J., Quail, M.A., Rajandream, M.A., Rogers, J., Rutter, S., Seeger, K., Skelton, J., Squares, R., Squares, S., Sulston, J.E., Taylor, K., Whitehead, S., Barrell, B.G., 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature. https://doi.org/10.1038/31159.
Coscolla, M., Gagneux, S., 2014. Consequences of genomic diversity in Mycobacterium tuberculosis. Semin. Immunol. https://doi.org/10.1016/j.smim.2014.09.012.
Costa-Veiga, A., Briz, T., Nunes, C., 2017. Unsuccessful treatment in pulmonary tuberculosis: factors and a consequent predictive model. Eur. J. Public Health. https://doi.org/10.1093/eurpub/ckx136.
Coxon, G.D., Craig, D., Corrales, R.M., Vialla, E., Gannoun-Zaki, L., Kremer, L., 2013. Synthesis, antitubercular activity and mechanism of resistance of highly effective thiacetazone analogues. PLoS One. https://doi.org/10.1371/journal.pone.0053162.
Demay, C., Liens, B., Burguière, T., Hill, V., Couvin, D., Millet, J., Mokrousov, I., Sola, C., Zozio, T., Rastogi, N., 2012. SITVITWEB – a publicly available international multimarker database for studying *Mycobacterium tuberculosis* genetic diversity and molecular epidemiology. Infect. Genet. Evol. https://doi.org/10.1016/j.meegid.2012.02.004.
Druszczynska, M., Wawrocki, S., Szewczyk, R., Rudnicka, W., 2017. Mycobacteria-derived biomarkers for tuberculosis diagnosis. Indian J. Med. Res. https://doi.org/10.4103/ijmr.IJMR_1441_16.
Fay, A., Czudnochowski, N., Rock, J.M., Johnson, J.R., Krogan, N.J., Rosenberg, O., Glickmana, M.S., 2019. Two accessory proteins govern MmpL3 mycolic acid transport in mycobacteria. mBio. https://doi.org/10.1128/mBio.00850-19.
Gande, R., Gibso, K.J.C., Brown, A.K., Krumbach, K., Dover, L.G., Sahm, H., Shioyama, S., Oikawa, T., Besra, G.S., Eggeling, L., 2004. Acyl-CoA Carboxylases (accD2 and accD3), Together with a Unique Polyketide Synthase (Cg-pks), Are Key to Mycolic Acid Biosynthesis in Corynebacterianeae Such as Corynebacterium glutamicum and Mycobacterium tuberculosis. J. Biol. Chem. https://doi.org/10.1074/jbc.M408648200.
García-Alcalde, F., Okonechnikov, K., Carbonell, J., Cruz, L.M., Götz, S., Tarazona, S., Dopazo, J., Meyer, T.F., Conesa, A., 2012. Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics. https://doi.org/10.1093/bioinformatics/bts503.
Gavalda, S., Léger, M., van der Rest, B., Stella, A., Bardou, F., Montrozier, H., Chalut, C., Burlet-Schiltz, O., Marrakchi, H., Daffé, M., Quemard, A., 2009. The Pks13/FadD32 crosstalk for the biosynthesis of mycolic acids in mycobacterium tuberculosis. J. Biol. Chem. https://doi.org/10.1074/jbc.M109.006940.
Glickman, M.S., Cox, J.S., Jacobs, W.R., 2000. A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of Mycobacterium tuberculosis. Mol. Cell. https://doi.org/10.1016/s1097-2765(00)80250-6.
Goins, C.M., Dajnowicz, S., Smith, M.D., Parks, J.M., Ronning, D.R., 2018. Mycolyltransferase from *Mycobacterium tuberculosis* in covalent complex with tetrahydrolipstatin provides insights into antigen 85 catalysis. J. Biol. Chem. https://doi.org/10.1074/jbc.RA117.001681.
Gouy, M., Guindon, S., Gascuel, O., 2021. Sea view version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol. Biol. Evol. https://doi.org/10.1093/molbev/msp259.
Guenin-Mace, L., Simeone, R., Demangel, C., 2009. Lipids of pathogenic mycobacteria: contributions to virulence and host immune suppression. Transbound. Emerg. Dis. https://doi.org/10.1111/j.1865-1682.2009.01072.x.
Hicks, N.D., Yang, J., Zhang, X., Zhao, B., Grad, Y.H., Liu, L., Ou, X., Chang, Z., Xia, H., Zhou, Y., Wang, S., Dong, J., Sun, L., Zhu, Y., Zhao, Y., Jin, Q., Fortune, S.M., 2018. Clinically prevalent mutations *in Mycobacterium tuberculosis* alter propionate metabolism and mediate multidrug tolerance. Nat. Microbiol. https://doi.org/10.1038/s41564-018-0218-3.
IBGE, 2019. Cidades. Panorama. https://cidades.ibge.gov.br/brasil/panorama.
Rajendra, K.S., Jayaraman, M., Ramadas, K., 2019. Structural insight into conformational dynamics of non-active site mutations in KasA: A Mycobacterium tuberculosis target protein. Gene. https://doi.org/10.1016/j.gene.2019.144082.
Kamerbeek, J., Schouls, L., Kolk, A., van Agterveld, M., van Soolingen, D., Kuijper, S., Bunschoten, A., Molhuizen, H., Shaw, R., Goyal, M., van Embden, J.D.A., 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. J. Clin. Microbiol. https://doi.org/10.1128/JCM.35.4.907-914.1997.
Kong, Y., Cave, M.D., Zhang, L., Foxman, B., Marrs, C.F., Bates, J.H., Yang, Z.H., 2007. Association between *Mycobacterium tuberculosis* Beijing/W lineage strain infection and extrathoracic tuberculosis: insights from epidemiologic and clinical characterization of the three principal genetic groups of *M. tuberculosis* clinical isolates. J. Clin. Microbiol. https://doi.org/10.1128/JCM.01459-06.
Kowalski, K., Trzepiński, P., Druszczyńska, M., Boratyński, J., 2014. Mycolic acids – biological role and potential application in Mycobacterium detection and differentiation. Postepy Hig. Med. Dosw. https://doi.org/10.5604/17322693.1097425.
Kozarewa, I., Turner, D.J., 2011. High-throughput next generation sequencing. Methods Mol. Biol. https://doi.org/10.1038/nmeth.1270.
Kremer, L., Maughan, W.N., Wilson, R.A., Dover, L.G., Besra, G.S., 2002. The *M. tuberculosis* antigen 85 complex and mycolyltransferase activity. Lett. Appl. Microbiol. https://doi.org/10.1046/j.1472-765x.2002.01091.x.
Lew, J.M., Kapopoulou, A., Jones, L.M., Cole, S.T., 2011. TubercuList e 10 years after. Tuberculosis. https://doi.org/10.1016/j.tube.2010.09.008.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup, 2009. The sequence alignment/map format and SAMtools. Bioinformatics. https://doi.org/10.1093/bioinformatics/btp352.
Makinoshima, H., Glickman, M.S., 2005. Regulation of M. tuberculosis cell envelope composition and virulence by Regulated Intramembrane Proteolysis. Nature. https://doi.org/10.1038/nature03713.
Murphy, H.N., Stewart, G.R., Mischenko, V.V., Apt, A.S., Harris, R., Mcalister, M.S.B., Driscoll, P.C., Young, D.B., Robertson, B.D., 2005. The OtsAB pathway is essential for Trehalose biosynthesis in Mycobacterium tuberculosis. J. Biol. Chem. https://doi.org/10.1074/jbc.M414232200.
Niemann, S., Supply, P., 2014. Diversity and evolution of *Mycobacterium tuberculosis* - moving to whole-genome based approaches. Cold Spring Harbor Persp. Med. https://doi.org/10.1101/cshperspect.a021188.
Nikolayevskyy, V., Niemann, S., Anthony, R., van Soolingen, D., Tagliani, E., Kodmon, C., van der Werf, M.J., Cirillo, D.M., 2019. Role and value of whole genome sequencing in studying tuberculosis transmission. Clin. Microbiol. Infect. https://doi.org/10.1016/j.cmi.2019.03.022.
Nogueira, C.L., Prim, R.I., Senna, S.G., Rovaris, D.B., Maurici, R., Rossetti, M.L., Couvin, D., Rastogi, N., Bazzo, M.L., 2016. First insight into the molecular epidemiology of *Mycobacterium tuberculosis* in Santa Catarina, southern Brazil. Tuberculosis. https://doi.org/10.1016/j.tube.2015.12.005.
Øyås, O., Borrelle, S., Traunerc, A., Zimmermanne, M., Feldmannc, J., Liphardta, T., Gagneux, S., Stellinga, J., Sauere, U., Zampierie, M., 2020. Model-based integration of genomics and metabolomics reveals SNP functionality in *Mycobacterium tuberculosis*. In: Proceedings of the National Academy of Sciences - PNAS. https://doi.org/10.1073/pnas.1915551117.
Pawelczyk, J., Kremer, L., 2014. The molecular genetics of mycolic acid biosynthesis. Microbiol. Spect. https://doi.org/10.1128/microbiolspec.MGM2-0003-2013.
Pepperell, C.S., Casto, A.M., Kitchen, A., Granka, J.M., Cornejo, O.E., Holmes, E.C., Birren, B., Galagan, J., Feldman, M.W., 2013a. The role of selection in shaping diversity of natural *M. tuberculosis* populations. PLoS Pathog. https://doi.org/10.1371/journal.ppat.1003543.
Pepperell, C.S., Casto, A.M., Kitchen, A., Julie, M., Granka, J.M., Cornejo, O.E., Holmes, E.C., Birren, B., Galagan, J., Feldman, M.W., 2013b. The role of selection in shaping diversity of natural *M. tuberculosis* populations. PLoS Pathog. https://doi.org/10.1371/journal.ppat.1003543.
Portevin, D., Sukumar, S., Coscolla, M., Shui, G., Li, B., Guan, X.L., Bendt, A.K., Young, D., Gagneux, S., Wenk, M.R., 2014. Lipidomics and genomics of *Mycobacterium tuberculosis* reveal lineage-specific trends in mycolic acid biosynthesis. MicrobiologyOpen. https://doi.org/10.1002/mbo3.193.
Prim, R.I., Schörner, M.A., Senna, S.G., Nogueira, C.L., Figueiredo, A.C., Oliveira, J.G., Rovaris, D.B., Bazzo, M.L., 2015. Molecular profiling of drug resistant isolates of *Mycobacterium tuberculosis* in the state of Santa Catarina, southern Brazil. Mem. Inst. Oswaldo Cruz. https://doi.org/10.1590/0074-02760150100.
Purwantini, E., Mukhopadhyay, B., 2013. *Rv0132c* of *Mycobacterium tuberculosis* encodes a coenzyme F420-dependent hydroxymycolic acid dehydrogenase. PLoS One. https://doi.org/10.1371/journal.pone.0081985.
Romero, I.C., Mehaffy, C., Burchmore, R.J., Dobos-Elder, K., Brennan, P., Walker, J., 2010. Identification of promoter-binding proteins of the *fbp A* and *C* genes in *Mycobacterium tuberculosis*. Tuberculosis. https://doi.org/10.1016/j.tube.2009.10.001.
Sachdeva, S., Reynolds, K.A., 2008. *Mycobacterium tuberculosis* β-Ketoacyl Acyl carrier protein synthase III (mtFabH) assay: principles and method. Methods Mol. Med. https://doi.org/10.1007/978-1-59745-246-5_16.
Sambandan, D., Dao, D.N., Weinrick, B.C., Vilchèze, C., Gurcha, S.S., Ojha, A., Kremer, L., Besra, G.S., Hatfull, F.G., Jacobs Jr., W.R., 2013. Keto-mycolic acid-dependent pellicle formation confers tolerance to drug-sensitive *Mycobacterium tuberculosis*. MBio. https://doi.org/10.1128/mBio.00222-13.
Slayden, R.A., Barry 3rd., C.E., 2002. The role of KasA and KasB in the biosynthesis of meromycolic acids and isoniazid resistance in *Mycobacterium tuberculosis*. Tuberculosis. https://doi.org/10.1054/tube.2002.0333.
Stucki, D., Brites, D., Jeljeli, L., Mireia, Coscolla M., Liu, Q., Trauner, A., Fenner, L., Rutaihwa, L., Borrell, S., Luo, T., Gao, Q., Kato-Maeda, M., Ballif, M., Egger, M., Macedo, R., Mardassi, H., Moreno, M., Vilanova, G.T., Fyfe, J., Globan, M., Thomas, J., Jamieson, F., Guthrie, J.L., Asante-Poku, A., Yeboah-Manu, D., Wampande, E., Ssengooba, W., Joloba, M., Boom, W.H., Basu, I., Bower, J., Saraiva, M., Vasconcellos, S.E.G., Suffys, P., Koch, A., Wilkinson, R., Gail-Bekker, L., Malla, B., Ley, S.D., Beck, H.-P., De Jong, B.C., Toit, K., Sanchez-Padilla, E., Bonnet, M., Gil-Brusola, A., Frank, M., Beng, V.N.P., Eisenach, K., Alani, I., Ndung'u, P.W., Revathi, G., Gehre, F., Akter, S., Ntoumi, F., Stewart-Isherwood, L., Ntinginya, N.E., Rachow, A., Hoelscher, M., Cirillo, D.M., Skenders, G., Hoffner, S., Bakonyte, D., Stakenas, P., Diel, R., Crudu, V., Moldovan, O., Al-Hajoj, S., Otero, L., Barletta, F., Carter, E.J., Diero, L., Supply, P., Comas, I., Niemann, S., Gagneux, S., 2016. *Mycobacterium tuberculosis* lineage 4 comprises globally distributed and geographically restricted sublineages. Nat. Genet. https://doi.org/10.1038/ng.3704.

Tarasov, A., Vilella, A.J., Cuppen, E., Nijman, I.J., Prins, P., 2015. Sambamba: fast processing of NGS alignment formats. Bioinformatics. https://doi.org/10.1093/bioinformatics/btv098.

Thanna, S., Suchecka, S.J., 2016. Targeting the trehalose utilization pathways of Mycobacterium. MedChemComm. https://doi.org/10.1039/C5MD00376H.

Takayama, K., Wang, C., Besra, G.S., 2005. Pathway to synthesis and processing of mycolic acids in Mycobacterium tuberculosis. Clin. Microbiol. Rev. https://doi.org/10.1128/CMR.18.1.81-101.2005.

van Soolingen, D., De Haas, P.E.W., Hermans, P.W.M., van Embden, J.D.A., 1994. DNA fingerprinting of *Mycobacterium tuberculosis*. Methods Enzymol. https://doi.org/10.1016/0076-6879(94)35141-4.

Verza, M., Scheffer, M.C., Salvato, R.S., Schorner, M.A., Barazzetti, F.H., Machado, H.M., Medeiros, T.F., Rovaris, D.B., Portugal, I., Viveiros, M., Perdigão, J., Kritski, A., Bazzo, M.L., 2020. Genomic epidemiology of Mycobacterium tuberculosis in Santa Catarina, Southern Brazil. Sci. Rep. https://doi.org/10.1038/s41598-020-69755-9.

WHO, 2020. Global Tuberculosis Report 2020. https://apps.who.int/iris/bitstream/handle/10665/336069/9789240013131-eng.pdf.

Wilkinson, R.J., Desjardin, L.E., Islam, N., Gibson, B.M., Kanost, R.A., Wilkinson, K.A., Poelman, D., Eisenach, D., Toossi, Z., 2001. An increase in expression of a *Mycobacterium tuberculosis* mycolyl transferase gene (fbpB) occurs early after infection of human monocytes. Mol. Microbiol. https://doi.org/10.1046/j.1365-2958.2001.02280.x.

Xu, W., Dejesus, M.A., Rücker, N., Engelhart, C.A., Wright, M.G., Healy, C., Lin, K., Wang, R., Park, S.W., Ioerger, T.R., Schnappinger, D., Sabine Ehrta, S., 2017. Chemical Genetic Interaction Profiling Reveals Determinants of Intrinsic Antibiotic Resistance in Mycobacterium tuberculosis. Antimicrob. Agents Chemother. https://doi.org/10.1128/AAC.01334-17.

4 DISCUSSÃO

Capítulo I - Epidemiologia molecular de *Mycobacterium tuberculosis* no Brasil antes da era do sequenciamento do genoma completo

O aumento da resistência às drogas anti-TB e o aumento da transmissão descontrolada de cepas com resistência a múltiplas drogas são os principais, e que está na origem dos casos de resistência primária, constitui um dos fatores para manutenção das taxas de mortalidade por TB em diferentes regiões do mundo. A investigação das fontes de transmissão e o monitoramento das cepas de *M. tuberculosis* por meio de estudos de epidemiologia molecular complementados por ferramentas de tipagem molecular é, portanto, essencial para o controle da doença. As ações de vigilância da TB no Brasil consistem basicamente na visita domiciliar de parte dos novos casos da doença, na busca por indivíduos que abandonam o tratamento, busca ativa por casos em hospitais e outras instituições, bem como acompanhamento e encerramento de casos (BRASIL, 2011). Atualmente não foi implementado em nível nacional qualquer método de tipagem que poderia ajudar na detecção de cadeias de transmissão por meio da identificação de clusters genômicos bem como informar acerca da implementação de novas estratégias direcionadas à contenção de cepas específicas porventura associados a fatores de risco.

Antes da ampla expansão do NGS observada na última década, as técnicas de MIRU-VNTR, RFLP-IS6110 e *spoligotyping* eram as técnicas empregadas para tipagem e inferências de cadeias de transmissão de *M. tuberculosis* ou exclusão de relação epidemiológica entre possíveis contatos ou casos potencialmente relacionados. Tanto o *spoligotyping* quanto tipagens baseadas em MIRU-VNTR são armazenados em grandes bancos de dados internacionais que permitem a comparação interlaboratorial de padrões, enquanto RFLP-IS6110, estão principalmente compondo bancos de dados locais (COSCOLLA, 2017; COUVIN & RASTOGI, 2014). Assim, no capítulo I desta tese, nós conduzimos uma revisão sistemática da literatura sobre o uso do RFLP-IS6110 e MIRU-VNTR no Brasil, com o objetivo de (i) descrever a estrutura dos laboratórios da rede brasileira, (ii) caracterizar os estudos de epidemiologia molecular aplicados à TB no Brasil; (iii) estudar a diversidade genética de *M. tuberculosis* no país e (iv) relacionar esses dados ao cenário epidemiológico nacional da TB.

No estudo observamos que, nas últimas duas décadas, a técnica de MIRU-VNTR tem sido menos usada para tipagem do que a técnica padrão-ouro anterior (RFLP IS6110), não apenas devido à implementação anterior desta última (1993, enquanto MIRU-VNTR foi implementada em 2008), mas também devido ao seu custo mais elevado. Isso é de particular importância para um país com considerável incidência de TB como o Brasil. No entanto, a implementação da tipagem para rastrear apenas casos de TB-MDR já teria um impacto significativo no controle da TB, uma vez que os casos MDR são aqueles que terão uma menor chance de cura, demandarão um período maior de tratamento, terão uma maior chance de um mau prognóstico e de transmissão na comunidade (WHO, 2021).

O rastreamento dos casos de TB e seus contatos é de vital importância para o controle da TB em países de alta carga da doença como o Brasil. Estudos sobre a diversidade genética e epidemiologia molecular da TB, por meio destas técnicas, no território brasileiro, foram mais frequentes nas regiões Sul e Sudeste e é imperativo reforçar a necessidade de vigilância da epidemiologia molecular nos estados do centro, nordeste e norte. Isso poderia ser alcançado pela capacitação intensiva de mais profissionais de laboratório e fornecimento dos materiais necessários para a execução da técnica em nível nacional, similar ao que tem sido empregado, desde 2020, para vigilância genômica do coronavirus nos laboratórios centrais de saúde pública do país. Uma alta, mas heterogênea proporção de transmissão de TB foi observada nas regiões brasileiras analisadas no capítulo I. Além disso, no estudo destaca-se a importância de incluir a análise genotípica na vigilância da TB, pelo menos de casos droga resistentes, e de manter um fluxo de dados entre os laboratórios da rede nacional. Assim, se propõe a implementação de técnicas de tipagem molecular para detecção de transmissão de TB, bem como a criação de um banco de dados nacional, visando um melhor controle da incidência da doença no país. Adicionalmente se faz necessário a construção de bancos de dados capazes de possibilitar alguma retrocompatibilidade entre os dados de tipagem por metodologias clásicas (RFLP, *spoligotyping* e MIRU-VNTR) e dados de tipagem por WGS, de modo que seja possível relacionar cepas caracterizadas pelos diferentes métodos.

Capítulo II - Vigilância genômica revela alta transmissão contínua de *Mycobacterium tuberculosis* multi drogaresistente no sul do Brasil

No Capítulo II estão descritos os resultados de uma análise baseada em WGS da diversidade genômica de *M. tuberculosis* circulando no Rio Grande do Sul que incluiu 75,7% dos 403 casos de TB-DR identificados no Laboratório Central de Saúde Pública do Estado de 2011 a 2014. O Rio Grande Sul tem a quarta maior taxa de mortalidade por TB entre os estados brasileiros e em estudos recentes conduzidos na região, observou-se importantes cadeias de transmissão de cepas resistentes de *M. tuberculosis* que dificultam o controle da doença (BRASIL, 2020; ESTEVES *et al*., 2018; SALVATO, *et al*., 2019; SALVATO, *et al*., 2020).

De modo geral, na predição da resistência baseada na identificação de mutações associadas a resistência observou-se uma boa concordância com testes fenotípico, com valores de concordância que variaram de 99,3% para isoniazida e 98,7% para rifampicina a 82,3% para etambutol e 86,9% para estreptomicina. Um achado relevante observado a partir da comparação entre a análise molecular e o teste fenotípico foi a presença de *disputed mutations* (MIOTTO *et al*., 2018), bem caracterizadas previamente, no gene *rpoB* em 22 cepas suscetíveis a rifampicina nos testes de sensibilidade realizados em meio líquido no BD BACTEC MGIT e em duas cepas resistentes. Entre as 96 amostras caracterizadas inicialmente como mono-resistentes a isoniazida no ensaio fenotípico, foram encontradas *disputed mutations* em 21,9% (N=21), evidenciando uma possível subestimação da proporção de casos MDR e podendo levar a prescrição de tratamentos anti-TB incorretos para esses casos. Além disso, o fato de 17 pacientes infectados com essas cepas terem sido submetidos a regimes de tratamento para TB-MDR após falhas do regime de 1ª linha, e outros dois evoluírem para óbito durante o tratamento, mostra a relevância clínica dessas mutações. Ademais, o potencial de disseminação e aparente ausência de perda de transmissibilidade ou fitness relativo dessas cepas é evidenciado pela detecção dessas cepas em importantes clusters genômicos de transmissão, dos quais destacamos o cluster genômico (GC) 4, predominantemente formado por essas cepas. Esses novos achados no Brasil revelam a importância do rastreamento de cepas que abrigam *disputed mutations* para uma melhor vigilância e controle da TB.

A ocorrência de *disputed mutations* tem se mostrado com uma ocorrência importante em diferentes regiões do mundo, em estudos anteriores foi descrito a presença de *disputed mutations* no gene *rpoB* em proporções que variaram de pelo menos 4,6% na África do Sul, 10% em Kinshasa, 13,1% em Bangladesh, 2,5% na China, 6,9% na Coréia do Sul e 9,2% em Taiwan (JO *et al*., 2017; LIN *et al*., 2021; MVELASE *et al*., 2019; VAN DEUN *et al*., 2013; WANG *et al*., 2013).

A boa concordância observada entre a predição da resistência pelo WGS e a detecção de resistência fenotípica, e a falta de capacidade do teste fenotípico automatizado para identificar resistência causada por *disputed mutations*, ressalta a importância da implementação de testes moleculares, como por exemplo, *MTBDRplus/MTBDRsl* ou *Target-NGS*, além dos já amplamente implementados testes rápidos moleculares como *GeneXpert*, que são capazes de detectar essas *disputed mutations*, e no caso do *GeneXpert*, já está disponível em alguns serviços de saúde do Rio Grande do Sul, auxiliando na detecção de algumas dessas mutações e fornecendo uma identificação de resistência oportuna (TORREA *et al*., 2019).

Como já descrito em estudos anteriores realizados no Rio Grande do Sul (ESTEVES *et al*., 2018; SALVATO, *et al*., 2019; SALVATO, *et al*., 2020), os principais genótipos associados à resistência as drogas de primeira linha incluem mutações de alta confiança, entre elas: *katG* S315T (em 50% de cepas resistentes), *katG* S315T + *inhA* -15 C>T (em 17,6%) e *inhA* -15 C>T (em 10,2%) causando resistência ao isoniazida. No entanto, 3,9% (11/284) de cepas resistentes a isoniazida só carregavam mutações relacionadas à resistência na região promotora do gene *ahpC* (-48 G>A, -52 C>T, -54 C>T e -57 C>T), sendo oito dessas cepas mono-resistentes a isoniazida. Apesar de evidências limitadas sobre o papel das mutações na região intergênica *ahpC-oxyR* conferindo resistência a isoniazida, em vários estudos (ALLIX-BÉGUEC *et al*., 2018; LIU *et al*., 2018; SEIFERT *et al*., 2015; WALKER *et al*., 2015) têm-se descrito uma possível associação entre essas variantes e resistência fenotípica a isoniazida, mas geralmente como um mecanismo compensatório para mutações no gene *katG*, principalmente mutações diferentes daquelas ocorrendo no códon 315. Em relação às onze cepas que carregam apenas mutações em *ahpC*, não encontramos outros marcadores moleculares que pudessem elucidar o mecanismo causando tal resistência. Visando uma melhor compreensão da

ocorrência de resistência nessas cepas, outros mecanismos moleculares, como a expressão das bombas de efluxo devem ser analisados (MACHADO *et al*., 2012, 2018).

Em nosso conjunto de amostras, encontramos a resistência a rifampicina sendo causada principalmente por mutações no gene *rpoB*: S450L (em 54,6%), S450L + A286V (em 10,9%) e também pela inserção de 12 nucleotídeos (12nt) no códon 435 do gene *rpoB* (em 9,2%), que evidencia a disseminação significativa dessas cepas altamente resistentes a rifampicina (MIC ≥ 32 mg/L) na população (ROSSETTI *et al*., 2020). Da mesma forma, as mutações mais comuns associadas à resistência foram encontradas entre as cepas clínicas fenotipicamente resistentes a etambutol (*embB* M306V em 83,3% das cepas resistentes) e estreptomicina (*rrs* 492 C>T em 14,9% e *rpsL* K43R em 14,9%), semelhante ao encontrado em conjuntos de amostras em outros países (COLL *et al*., 2018; PHELAN *et al*., 2019).

Mutações que conferem resistência a drogas de segunda linha foram encontradas em nove cepas: seis cepas MDR tinham mutações conferindo resistência a fluoroquinolonas no gene *gyrA*, uma cepa poli-resistente e duas cepas mono-resistentes a isoniazida também apresentaram mutações associadas a resistência no gene *gyrA*, e uma delas também carregava uma mudança de nucleotídeo (1401a>g) no gene *rrs*, que está relacionada à resistência a drogas injetáveis de segunda linha , caracterizando um caso de TB XDR. De acordo com dados clínicos, apenas um desses nove indivíduos pré-XDR/XDR-TB teve um desfecho favorável no tratamento, e três deles abandonaram o regime de tratamento, tal fato contribui para a disseminação dessa resistência. Neste estudo do Capítulo II, não foi realizado teste fenotípico para detecção as drogas de segunda linha, impedindo a comparação com a predição baseada em WGS. No entanto, trata-se de mutações clássicas associadas à resistência a ambos os grupos de drogas aqui evidenciados constituindo estas mutações variantes associadas à resistência de elevado grau de confiança. Adicionalmente, um trabalho anterior revelou uma boa concordância entre mutações associadas a resistência e a resistência fenotípica às fluoroquinolonas em um subconjunto de cepas clínicas da mesma região (SALVATO *et al*., 2020).

Como anteriormente caracterizado por estudos recentes na região sul do Brasil (SALVATO, *et al*., 2019; SALVATO, *et al*., 2020), a estrutura populacional de *M.*

tuberculosis tem sua composição basicamente por cepas das famílias LAM e Haarlem. A mesma predominância de LAM ocorre no estado fronteiriço de Santa Catarina (VERZA *et al*., 2020), apesar da predominância de sublinhagens 4.3.3 no estado vizinho, ao contrário da predominância da sublinhagem 4.3.4.2, seguida da 4.3.3 observada no presente estudo. Além da classificação baseada em SNPs ser capaz de determinar a sublinhagem de cepas que não foram passíveis de classificação pela técnica de *spoligotyping*, a classificação por SNPs também foi capaz de reclassificar em sublinhagens mais plausíveis um grande conjunto de cepas previamente definidas como sublineagem T de acordo com o padrão de *spoligotyping*. Das 93 cepas atribuídas ao clado T pelo SITVIT2, apenas 16 (17,2%) foram classificadas no clado T na tipagem baseada em SNPs e as demais foram atribuídas principalmente a sublinhagem 4.1.2.1/Haarlem (39.8%). Este fato foi evidenciado por 12 cepas que carregavam a variante fbpC103, um reconhecido marcador da família LAM, e pela posição dessas cepas na árvore filogenética.

No Rio Grande do Sul, a proporção de transmissão recente estimada de TB-DR foi de 55,1% entre 305 isolados de *M. tuberculosis* estudados, superior à registrada no estado fronteiriço Santa Catarina (VERZA *et al*., 2020) (conjunto de amostras majoritariamente suscetíveis), e outras regiões de alta carga de TB no mundo (MIYAHARA *et al*., 2020), sinalizando que importantes cadeias de transmissão estão alimentando a transmissão recente da TB na região. No entanto, o fato mais preocupante revelado neste estudo diz respeito às cepas MDR, as quais 73,4% foram agrupadas em clusters genômicos. Os três maiores GCs encontrados em nossa análise incluíam quase que exclusivamente cepas MDR. O GC1 foi constituído principalmente (24/25) por cepas provenientes de indivíduos sem histórico de encarceramento e destes, em 15 indivíduos observou-se links epidemiológicos comunitários identificados, representando uma grande cadeia de transmissão em curso na comunidade. Diferentemente, no GC2, das 21 cepas, 10 delas foram isoladas de indivíduos encarcerados em instituições prisionais, semelhante ao visualizado no GC3, onde 8/17 das cepas agrupadas eram de indivíduos presos ou com histórico recente de encarceramento, juntamente com a amostra de um profissional agente penitenciário. Assim, o GC3 representa uma importante cadeia de transmissão ativa envolvendo a população carcerária e a comunidade, semelhante ao notado em outras regiões do Brasil e do mundo (ROSSETTI *et al*., 2020; SACCHI *et al*., 2015; UMPELEVA *et al*., 2020).

Explorando a distribuição geográfica dos casos, foi possível observar que as maiores cadeias de transmissão em curso estão ocorrendo em Porto Alegre, capital do estado, bem como na sua região metropolitana. Em Porto Alegre, os quatro bairros responsáveis por 40,3% dos casos em clusters genômicos na cidade apresentam o menor Índice de Desenvolvimento Humano (IBGE, 2022), reforçando a necessidade de medidas mais robustas de controle da TB e de proteção social nessas regiões. Além disso, a presença de cepas MDR em clusters genômicos compostos principalmente por perfis mono/poli-resistentes, juntamente com a aquisição intra-cluster de mutações relacionadas à resistência observadas em múltiplos, mas principalmente nos maiores clusters, indiciam a emergência de TB-MDR adquirida/secundária nesses clusters, levando à amplificação da resistência e sendo um importante ponto de preocupação adicional para o controle da TB.

Assim, no capítulo II, utilizando uma abordagem genômica, elucidamos o cenário da TB-DR no maior estado do sul do Brasil. Os achados apontaram para a necessidade urgente de medidas de controle da TB que promovam a desaceleração da disseminação da TB-MDR, a fim de controlar a TB nessa região. A situação epidemiológica aqui observada pode refletir importantes indicadores operacionais dos programas de controle da TB, como a baixa cobertura de teste de suscetibilidade as drogas e tratamento diretamente observado, juntamente com as altas taxas de abandono de tratamento. Outro importante indicador operacional que requer atenção especial, essencialmente no controle da TB-MDR, é a investigação de contatos dos casos. Em 2019, em nível nacional, apenas 55,5% dos contatos identificados de TB foram examinados, essa proporção foi ainda menor no Rio Grande do Sul (37,8%) (BRASIL, 2020) e pode ser um aspecto crucial para sustentação da transmissão contínua de *M. tuberculosis* encontrada em nossa análise.

Neste estudo houve algumas limitações importantes: (i) obtivemos informações limitadas sobre vínculos epidemiológicos entre pacientes em clusters genômicos. Devido à natureza retrospectiva do estudo, só foram acessados bancos de dados secundários, e alguns possíveis links epidemiológicos entre pacientes que compartilhavam o mesmo ambiente no trabalho, escola, unidades de saúde ou a mesma rota utilizando transporte público podem ter sido perdidos; (ii) nós avaliamos a resistência fenotípica apenas aos medicamentos de primeira linha e (iii) a amostragem do estudo ocorreu de 2010 a 2014. Estudos prospectivos futuros com testes fenotípicos mais amplos e rastreamento de

contatos podem superar essas limitações e melhorar a vigilância genômica da TB-DR neste cenário.

Por fim, por meio dos achados descritos no Capítulo II foi possível revelar aspectos importantes sobre as bases moleculares da resistência as drogas que ocorrem em cepas *de M. tuberculosis* circulando no Rio Grande do Sul, mostrando a capacidade dos ensaios moleculares para a detecção da resistência as drogas e sua importância para detectar a resistência a rifampicina causada por *disputed mutations*, evitando a subnotificação de casos de TB-MDR. Além disso, foram identificados múltiplos eventos de transmissão contínua de cepas de *M. tuberculosis* droga resistente, principalmente entre as cepas MDR, ressaltando a necessidade de medidas para interromper a transmissão de TB na região e visando melhorar o controle da TB nas instituições prisionais. Os resultados aqui descritos evidenciam que diferente do que ocorre na maior parte do mundo, a TB-MDR no Rio Grande do Sul ocorre devido a dois fatores simultâneos: alta transmissão e por emergência de novo durante o tratamento. Importante que no Rio Grande do Sul há a transmissão de diversas cepas MDR, o que requer maior aprofundamento para identificar eventuais fatores de risco e medidas de contenção direcionadas a essas diferentes cadeias de transmissão.

Capítulo III - Uma cepa de *Mycobacterium tuberculosis* altamente resistente à rifampicina emergente no sul do Brasil

Em 2012, Perizzolo *et al.*, (2012) descreveu pela primeira vez a presença de uma inserção de 12 nucleotídeos (12nt) no gene *rpoB* entre cepas MDR coletadas de pacientes entre 2004 e 2006 no Rio Grande do Sul. Em 2015, Dalla Costa *et al.*, (2015) descreveram seis cepas apresentando a mesma inserção, todas pertencentes ao SIT 863 e identificadas entre 2006 e 2010. Em estudos mais recentes realizados na região observou-se um aumento da presença dessas cepas que carregam esta inserção, sinalizando sua propagação e transmissão contínua (ESTEVES *et al*., 2018; SALVATO, *et al*., 2019; SALVATO, *et al*., 2020). No capítulo II, onde realizamos um estudo das cepas de *M. tuberculosis* resistentes a medicamentos do Rio Grande do Sul, no período de 2011 e 2014, identificamos 16 cepas MDR carregando a inserção de 12nt no gene *rpoB*.

Para uma melhor gestão do controle da TB no Rio Grande do Sul, é necessária a elucidação das consequências moleculares e fenotípicas dessa inserção de 12nt, bem como

a compreensão sobre os padrões de disseminação dessa cepa na população. Para isso, no capítulo III, nós examinamos a relação filogenética entre essas cepas, as consequências fenotípicas da inserção em relação ao nível de resistência e custo biológico, e a predição *in silico* do efeito dessa inserção para a interação do gene *rpoB* com a rifampicina.

A tipagem baseada em WGS em conjunto à análise filogenética corrobora com dados anteriores a ocorrência da inserção de 12nt com cepas pertencentes ao SIT 863 que apresenta uma distribuição importante e restrita no Rio Grande do Sul (DALLA COSTA *et al*., 2015; SALVATO, *et al*., 2019; SALVATO, *et al*., 2020). De acordo com Dalla Costa *et al*., (2015) as cepas que apresentam o padrão SIT 863, e que foram previamente identificadas como *Mycobacterium pinnipedii*, são na verdade pertencentes a linhagem 4 do CMTB e pertencem a subfamília LAM. A errônea associação do SIT 863 a uma linhagem PINI deve-se pela falta de caracterização genômica existente anteriormente, no entanto, essa classificação permanece por corrigir na base SITVIT2. Nossos resultados estão de acordo com tal assertiva; pois todas as cepas foram classificadas como sublinhagem 4.3.3, subfamília LAM e apresentaram a presença da RD115. Esses dados também são confirmados pela inferência filogenética realizada que mostrou a proximidade genômica entre as cepas com a inserção de 12nt e outras cepas do SIT 863 sem a presença da inserção, bem como, com cepas da sublineagem 4.3.3. Embora em estudos anteriores terem relatado a presença da inserção de 12nt em cepas classificadas como outras sublinhagens, as técnicas de genotipagem anteriormente utilizadas, como RFLP, MIRU-VNTR e *spoligotyping*, poderiam produzir alguma classificação errônea (PERIZZOLO *et al*., 2012; SALVATO, *et al*., 2019).

Com o objetivo de identificar o padrão de transmissão dessas cepas carregando a inserção de 12nt, utilizamos um limite de cinco SNPs entre genomas para definir como um episódio de transmissão recente, e inferir que as cepas dentro dessa distância genômica máxima estão provavelmente ligadas epidemiológicamente (WALKER *et al*., 2013). As 16 cepas que apresentaram a inserção de 12nt no gene *rpoB* constituíram um cluster genômico único, indicativo da transmissão recente e contínua dessas cepas nos últimos anos. Além disso, as cepas pertencentes ao SIT 863 que não tinham a inserção de 12nt tiveram uma distância genômica de 12 SNPs ou menos de cepas com a inserção de 12nt, corroborando com as relações anteriormente citadas entre esses isolados.

As cepas que apresentaram a inserção de 12nt foram provenientes principalmente de indivíduos privados de liberdade (56,25%) de diferentes estabelecimentos prisionais, incluindo os três principais presídios estaduais que sabiamente apresentam superlotação e uma estrutura física limitada. Observou-se ainda entre os indivíduos privados de liberdade incluídos no estudo do capítulo III, a passagem do mesmo indivíduo por diferentes estabelecimentos prisionais em um curto período, podendo esse fato ter desempenhado um papel importante na disseminação da cepa com a inserção de 12nt. Além disso, o elevado número de indivíduos que abandonaram o tratamento anti-TB (43,8%) e outros fatores de risco ou vulnerabilidade social, como indivíduos em situação de rua, podem ser listados como possíveis responsáveis pela substancial transmissão contínua dessa cepa. No entanto, as cepas de indivíduos sem histórico de encarceramento dentro deste mesmo cluster genômico observado em cepas de indivíduos privados de liberdade, apontam para uma cadeia de transmissão envolvendo a população carcerária e a comunidade. Junto ao fato dessa cepa ter sido descrita inicialmente em indivíduos privados de liberdade, trata-se possivelmente de uma cepa cuja disseminação foi inicialmente amplificada em ambiente prisional, entre reclusos, tendo posteriormente ocorrido disseminado na comunidade. O fato de a população reclusa poder atuar como reservatório que alimenta a transmissão da TB para a comunidade é bem estabelecido (SACCHI *et al*., 2015), e estudo recente no Brasil demonstrou diferentes cepas sendo transmitidas da população carcerária para comunidade (WALTER *et al*., 2022) revelando que instituições prisionais também podem amplificar e propagar a TB em comunidades vizinhas. Tal fato requer atenção nas medidas de saúde pública que visam controlar a transmissão da TB em diferentes ambientes.

Para obter uma melhor compreensão quanto à ocorrência de inserção de 12nt, analisamos a sequência de nucleotídeos do gene *rpoB* e observamos que a inserção de 12nt ocorre com uma variação discreta entre as cepas. No geral, a inserção de 12nt ocorreu na posição genômica 761111 correspondente ao códon 435 do gene *rpoB*. No entanto, 12 das 16 cepas com a inserção apresentaram uma mudança adicional de adenina para guanina na posição 761110, resultando na substituição de aminoácidos no códon 435 (GAC > GGC, Asp435Gly). Para facilitar a compreensão, denominamos a inserção de 12nt com polimorfismo adicional na posição genômica 761110 de "*rpoB* INS1", e aquela com apenas a inserção de 12nt de "*rpoB* INS2". Em seguida, nós modelamos por homologia as estruturas da RNA polimerase (RNAp) com as mutações *rpoB* INS1 e *rpoB* INS2, bem

como a estrutura da RNAp wt e ancoramos essas estruturas com a molécula de rifampicina. Nas análises observamos que a inserção de 12nt diminuiu a afinidade da rifampicina com as formas da RNAp_INS1 e RNAp_INS2 mutantes, o que pode explicar a resistência das cepas que carregam essas mutações. Também é possível observar que as inserções promovem mudanças no sítio de ligação com a rifampicina, que provavelmente está ligado à menor afinidade da interação da rifampicina com as formas mutantes.

No capítulo III nós exploramos ainda o impacto fenotípico da inserção de 12nt no que diz respeito ao nível de resistência a rifampicina e à proporção de crescimento dessas cepas. Utilizando a classificação do nível de resistência definida no estudo clássico de Huitric *et al* (HUITRIC *et al*., 2006), os nove isolados clínicos testados mostraram um alto nível de resistência (MIC> 32 μg/mL) a rifampicina. Como observado na análise *in silico*, a inserção dos 12 nucleotídeos provavelmente causa uma diminuição da afinidade na ligação da rifampicina ao seu alvo na β-subunidade da RNA polimerase codificada pelo gene *rpoB* e reflete no seu nível de resistência. Em relação ao custo biológico, observou-se que tanto as cepas com a *rpoB* INS1 como as com a *rpoB* INS2 apresentaram uma proporção de crescimento mais lento quando comparados aos isolados clínicos do tipo selvagem e ao padrão H37Rv. Considerando então que o gene *rpoB* codifica uma subunidade da enzima RNA polimerase responsável pela transcrição do DNA, que é um processo essencial na replicação de bactérias, essa mudança em sua estrutura pode ter influenciado nesse processo e consequentemente no seu tempo de crescimento. Em estudos anteriores foi descrito também um custo biológico associado a mutações no gene *rpoB*, no entanto, poderia ser mitigado com mutações compensatórias (GAGNEUX, LONG, *et al*., 2006; MARIAM *et al*., 2004). É importante ressaltar que, apesar da inserção de 12nt causar um custo biológico, é difícil extrapolar essa característica fenotípica para um cenário epidemiológico, uma vez que se trata de uma interação de fatores ambientais, intrínsecos e do hospedeiro.

Em resumo, no capítulo III descrevemos os principais aspectos moleculares e fenotípicos de uma cepa *de M. tuberculosis* emergindo no sul do Brasil nos últimos anos que carrega uma inserção incomum de 12nt no gene *rpoB* conferindo resistência de alto nível a rifampicina, ainda não descrita em nenhuma outra região do Brasil ou do mundo. A abordagem filogenética demonstrou uma possível transmissão dessa cepa da população

carcerária para a comunidade, demonstrando uma necessidade urgente de maior vigilância na transmissão da TB em meio à população privada de liberdade e ao abandono do tratamento anti-TB, especialmente para os casos MDR nesta região. Devido ao alto nível de resistência a rifampicina - a droga mais importante no tratamento anti-TB - e sua transmissão contínua estabelecida, é essencial monitorar essa cepa que se transmite entre a população.

Capítulo IV - Epidemiologia genômica de *Mycobacterium tuberculosis* em Santa Catarina

No estado de Santa Catarina, sabe-se que a estrutura populacional de *M. tuberculosis* é predominantemente formada por cepas da sublinhagem LAM, seguidas pela sublinhagem T segundo a classificação por *spoligotyping* (NOGUEIRA *et al*., 2016). No entanto, não havia informações sobre a diversidade genômica associada a essa região e os estudos usando métodos de tipagem de maior resolução empregados eram limitados a técnica de MIRU-VNTR 12 *loci* (MEDEIROS *et al*., 2018). Para preencher essa lacuna, no capítulo IV relatamos os resultados de um estudo genômico e epidemiológico na região metropolitana de Florianópolis, com o objetivo de identificar os genótipos de *M. tuberculosis* circulantes nesta região, inferir sobre os clusters de transmissão existentes e ativos na comunidade, examinar a distribuição de resistência a drogas e ainda exploramos possíveis associações entre resultados de tratamento desfavoráveis e linhagens baseadas em SNPs.

Neste estudo avaliou-se o cenário da diversidade genômica de cepas de *M. tuberculosis* isoladas entre maio de 2014 e maio de 2016, analisando amostras de indivíduos diagnosticados com TB pulmonar em Florianópolis e São José (as duas mais populosas das 22 cidades que compõem a região metropolitana de Florianópolis). Apesar da importância de se conhecer o perfil molecular das cepas de *M. tuberculosis* circulantes em um determinado cenário, os dados existentes para essa região específica limitavam-se à caracterização genotípica clássica das cepas circulantes (MEDEIROS *et al*., 2018; NOGUEIRA *et al*., 2016; PRIM *et al*., 2015). O estudo do capítulo IV compreendeu o primeiro estudo genômico de *M. tuberculosis* no estado de Santa Catarina. Na amostra deste estudo encontramos uma alta proporção de co-infecção TB-HIV (19,71%)

compatível com a esperada para esta região (22,6%) e bem acima da nacional (9,4%) (BRASIL, 2017). Pacientes coinfectados foram associados a uma baixa proporção de cura e consumo excessivo de álcool e drogas. A coinfecção por HIV é um fator de risco amplamente conhecido para o desenvolvimento de TB, uma vez que o risco de desenvolvimento de TB é 19 vezes maior na população vivendo com HIV quando comparado ao restante da população, levando a piores desfechos e menores taxas de cura (MAGIS-ESCURRA *et al*., 2017; WHO, 2021). No entanto, o abuso de álcool ou o uso não saudável de álcool é um fator de risco cada vez mais reconhecido por afetar o resultado do tratamento da TB e um fator de risco para a adesão ao tratamento da TB (CHAVES TORRES *et al*., 2019; FLEMING *et al*., 2006), além de ter efeito negativo no sistema imune e alterar a farmacocinética e absorção de diferentes drogas (MYERS *et al.*, 2018). Em estudos recentes realizados em Uganda e Quênia, observou-se ainda que o uso de álcool entre indivíduos infectados por HIV parece estar associado à diminuição da supressão viral devido à menor proporção de diagnóstico e menor probabilidade de estar em regime de tratamento anti-retroviral se já diagnosticado com HIV (PURYEAR *et al*., 2020). Nesse sentido, uma abordagem integrada para reduzir o consumo não saudável de álcool ou o uso de drogas ilícitas pode levar a melhores resultados quanto ao desfecho dos tratamentos anti-TB (SANDGREN *et al*., 2013).

Em Santa Catarina, na análise dos dados de WGS foram identificadas mutações que conferem resistência as drogas ao longo de todo genoma, o que permitiu a identificação de 10 (6,6%) isolados resistentes a pelo menos uma droga anti-TB e um (0,7%) isolado MDR-TB. Apesar da baixa proporção de TB-MDR neste estudo, prevê-se que dois isolados clínicos de pacientes diferentes sejam monorresistentes as fluoroquinolonas pela presença de duas mutações de alta confiança para predição de resistência a fluoroquinolonas no gene *gyrA* (MIOTTO *et al*., 2017; PERDIGÃO *et al*., 2020). A monorresistência a fluoroquinolonas ou entre isolados não MDR foi relatada em várias regiões do mundo, com frequências variadas (ZIGNOL *et al*., 2016). Recentemente, Kim *et al*., (2019), relataram uma proporção de resistência a fluoroquinolonas de 0,8% entre cepas não MDR detectadas em vários hospitais na Coréia do Sul, revelando uma tendência crescente nas últimas duas décadas. Pacientes com TB com prescrição de fluoroquinolonas anterior ao diagnóstico de TB, geralmente para tratar pneumonia adquirida na comunidade, têm risco três vezes maior de ter TB resistente a fluoroquinolonas (DEVASIA *et al*., 2009). Múltiplas prescrições de

fluoroquinolonas, e o tratamento com fluoroquinolonas por mais de 10 dias, principalmente há mais de 60 dias antes do diagnóstico de TB, estão associados à TB resistente a fluoroquinolonas (DEVASIA *et al*., 2009; LONG *et al*., 2009). Em nosso estudo observamos uma proporção de resistência às fluoroquinolonas de 1,3% entre pacientes com TB não MDR para esse cenário, impulsionado por dois eventos mutacionais independentes. Uma limitação para a interpretação dos dados refere-se no fato de que não foram obtidos dados referentes à exposição prévia a fluoroquinolonas para esses dois pacientes e, como tal, não podemos excluí-los de serem casos primários de TB monorresistente as fluoroquinolonas (WANG *et al*., 2018). No entanto, a detecção desses dois isolados monorresistentes a fluoroquinolonas alerta para um possível uso excessivo de fluoroquinolonas na comunidade e exige medidas adicionais de manejo antimicrobiano, pois, até onde sabemos, esses são os dois primeiros casos de monorresistência a fluoroquinolonas a serem relatados no Brasil.

Em relação à diversidade genética de *M. tuberculosis*, a tipagem baseada em SNP classificou todos os 151 isolados de *M. tuberculosis* como linhagem 4 (Euro-americana). A predominância da linhagem Euro-Americana no Estado de Santa Catarina e no sul do Brasil (NOGUEIRA *et al*., 2016; SALVATO *et al*., 2020) ocorreu devido a processos migratórios da Europa para a América do Sul que se intensificaram no século XVII (BRYNILDSRUD *et al*., 2018). Portanto, LAM, T e Haarlem, as famílias baseadas em *spoligotyping* mais comuns identificadas neste estudo, também foram encontradas em outros estudos no sul do Brasil (NOGUEIRA *et al*., 2016; PERDIGÃO *et al*., 2018; PRIM *et al*., 2015; SALVATO *et al*., 2019; SALVATO, *et al*., 2020). Nossa análise da presença do SNP fbpC103, considerado um marcador altamente específico para a família LAM, não apenas confirmou a identificação de todos as cepas classificadas como LAM por *spoligotyping*, mas identificou ainda três cepas SIT823/T1, atribuídos erroneamente à família T de acordo com SITVIT2, que apresentaram o marcador fbpC103 e estão de fato posicionados filogeneticamente como LAM na árvore filogenética.

De acordo com estudos anteriores (BJORN-MORTENSEN *et al*., 2016; WALKER *et al*., 2014), o uso de um limite conservador de cinco SNPs para definir clusters genômicos para inferência de episódios de transmissão recente, permitiu o agrupamento de 43,7% dos isolados incluídos no estudo, demonstrando que aproximadamente metade dos

casos de TB na região metropolitana de Florianópolis entre 2014 e 2016 foram associados à cadeias de transmissão recentes. Além disso, não foi encontrada associação entre agrupamento genômico (considerando os limites de 5, 12 ou 25 SNPs para definir os clusters genômicos) e fatores de risco, abandono ou falência do tratamento ou características clínicas.

O cenário de transmissão recente observado no capítulo IV evidencia a ocorrência de uma transmissão contínua da TB, provavelmente ainda em andamento. Uma limitação do estudo consiste no fato de que apenas 151 (57,4%) casos foram analisados de um total de 263 casos notificados para o mesmo período na Grande Florianópolis e, portanto, é provável que existam ligações perdidas nas cadeias de transmissão. No entanto, os dados reportados já permitiram a detecção de uma grande cadeia de transmissão (GC1) responsável por 7,9% dos casos analisados. Todos os isolados agrupados no GC1 pertenciam à sublinhagem 4.3.3 e eram classificados como SIT216/LAM5. Curiosamente, o SIT216/LAM5 foi descrito em Santa Catarina como o SIT mais prevalente em estabelecimentos prisionais, mas apenas um isolado no GC1 era proveniente de paciente que passou algum tempo em uma instituição prisional (MEDEIROS *et al*., 2018). Esses fatos destacam a importância epidemiológica das cepas SIT216/LAM5, agora em nível comunitário, e sua associação com a transmissão recente. O padrão de *spoligotyping* LAM5 também é o mais relatado no Rio Grande do Sul, porém pertencente ao SIT9334. Situação semelhante tem sido relatada com a cepa MDR SIT863 no Rio Grande do Sul, inicialmente identificada em estabelecimentos prisionais, mas hoje responsável pela maioria dos casos de TB-MDR no estado (DALLA COSTA *et al*., 2015; PERDIGÃO *et al*., 2018). A presença de isolados de *M. tuberculosis* de indivíduos reclusos nas cadeias de transmissão da TB favorece a disseminação de cepas entre a população em geral e os estabelecimentos prisionais e, embora a direcionalidade não seja clara neste momento, este último pode atuar como reservatório de cepas específicas e promover uma maior dispersão de clusters específicos (SACCHI *et al*., 2015). Aqui, o cluster GC1 apresentou uma ampla distribuição geográfica quando considerada a residência do paciente. Além disso, as baixas distâncias em SNPs entre os isolados nesses maiores clusters genômicos podem indicar que a transmissão ocorreu em um período muito recente e provavelmente está em andamento, sugerindo que, para evitar uma maior disseminação dessas cepas, uma vigilância mais próxima dessas clados filogeneticamente distintos é altamente importante.

Em resumo, no capítulo IV é enfatizada a importância das abordagens baseadas em WGS que permitem análises filogenéticas de alta resolução para investigar a transmissão de *M. tuberculosis* no Brasil. Os resultados foram importantes para elucidar a dinâmica da TB nesse cenário e diminuiu a falta de dados genômicos sobre cepas de *M. tuberculosis* circulantes no estado de Santa Catarina. Os dados mostraram que ocorreu uma transmissão descontrolada da TB na região metropolitana de Florianópolis e forneceu dados precisos para subsidiar as medidas de controle da TB nessa região.

Capítulo V - Caracterização genômica de variantes em genes do metabolismo do ácido micólico em isolados de *Mycobacterium tuberculosis* de Santa Catarina

M. tuberculosis possui uma parede celular complexa contendo ácidos micólicos, que desempenham um papel importante na patogênese, virulência e sobrevivência, protegendo a célula contra ambientes hostis (DRUSZCZYNSKA *et al*., 2017; GUENIN-MACÉ *et al*., 2009). Em estudos foi mostrado que genes codificadores de enzimas envolvidas na síntese de ácidos micólicos são essenciais para a funcionalidade micobacteriana (GLICKMAN *et al*., 2000). No capítulo V dessa tese nós usamos os dados genômicos de WGS para avaliar mutações em genes relacionados ao metabolismo de ácidos micólicos em isolados de *M. tuberculosis* de pacientes com TB pulmonar da Região Metropolitana de Florianópolis, Santa Catarina, e sua potencial associação com desfechos clínicos, características epidemiológicas e genótipos específicos. Estimulado pela importância do monitoramento do perfil molecular de cepas de *M. tuberculosis* circulantes, neste estudo buscamos complementar os relatos anteriores de caracterização genotípica de *M. tuberculosis* na região, e mais recentemente, a análise epidemiológica e genômica apresentada no capítulo IV desta tese.

Embora *M. tuberculosis* exiba níveis relativamente baixos de diversidade genética, sabe-se também que genes essenciais para a fisiologia micobacteriana podem estar sob diversas pressões de seleção. Isso pode ser esperado, pois o metabolismo dos ácidos micólicos é o alvo da maioria dos medicamentos usados no tratamento da TB (PEPPERELL *et al*., 2013; PORTEVIN *et al*., 2014). Com abordagens baseadas em WGS, é possível detectar microevoluções em regiões genômicas diferentes daquelas conhecidas

por serem altamente variáveis e comumente avaliadas em análises tradicionais (NIEMANN & SUPPLY, 2014; NIKOLAYEVSKYY *et al*., 2019).

Portevin *et al*., (2014) realizaram uma quantificação relativa dos subtipos de ácidos micólicos e analisaram sequências genômicas de isolados clínicos do CMTB, incluindo quatro linhagens filogenéticas principais (linhagens 1, 2, 4 e 6). Variações significativas nos padrões de ácido micólico foram encontradas entre as linhagens. Os autores descobriram que os padrões de linhagens ancestrais contrastavam com os de linhagens modernas e identificaram importantes SNPs possivelmente associados a padrões específicos de ácido micólico. Em nosso estudo, os 151 isolados foram classificados como Euro-Americanos (linhagem 4). Das principais linhagens, a linhagem 4 é a mais frequente em todo o mundo (STUCKI *et al*., 2016). Em estudos anteriores observou-se que a linhagem 4 apresenta uma importante diversidade genética e fenotípica, o que pode determinar a epidemiologia das suas sublinhagens (COSCOLLA & GAGNEUX, 2014; STUCKI *et al*., 2016).

As oito mutações identificadas na linhagem 4 por Portevin *et al*., (2014) como responsáveis por alterações de aminoácidos, que também foram detectadas nos isolados clínicos avaliados no presente estudo. Seis mutações - *accD2* Lys23Glu (46/151), *fas* Thr2595Asn (5/151), *kasA* Gly269Ser (46/151), *mmaA4* Asn165Ser (97/151), *Rv3057c* Asp112Ala (104/151) e *Rv3802c* Val50Phe (105/151), foram encontrados com maior frequência em cepas LAM, mas também foram observados em cepas Órfãs (padrões de *spoligotyping* que são exclusivos da cepa analisada e não são encontrados no banco de dados do SITVIT). No presente estudo, a mutação *fabH* Gln182His ocorreu exclusivamente em cepas SIT 137/X2, e *Rv3057c* Val275Ala foi identificada apenas em cepas da linhagem Haarlem. Os resultados observados em nosso estudo mostram uma diversidade intralinhagem assim como reportado por Portevin *et al*., (2014), sugerindo que a variabilidade nos padrões de ácidos micólicos podem contribuir para a interação entre diferentes cepas de CMTB e diferentes hospedeiros humanos. Embora não seja possível concluir que alterações em genes envolvidos nas vias biossintéticas dos ácidos micólicos estejam necessariamente relacionados com alterações nos próprios ácidos micólicos, são necessários estudos focados em analisar essas alterações e relacionar com as diferenças na composição química e abundância relativa desses ácidos micólicos.

O desfecho desfavorável no tratamento da TB pode ser devido a fatores associados ao hospedeiro e efetividade do programa de controle de TB (COSTA-VEIGA *et al*., 2018), bem como a fatores relacionados a micobactéria. Sambandan *et al*., (2013) assinalaram que as alterações na composição dos ácidos micólicos da parede celular são de particular importância no contexto do tratamento da TB. Por exemplo, mudanças na proporção de ácido metoxi-/ceto-micólico podem alterar a suscetibilidade basal a drogas antimicobacterianas na ausência de mecanismos de resistência a drogas. Mutações em genes relacionados ao metabolismo do ácido micólico já foram descritas e, embora nenhuma estivesse diretamente relacionada à falha do tratamento, ocorreu uma mutação em um gene considerado essencial, o gene *rip*. O gene *rip* regula a composição do envelope celular, o crescimento e a persistência in vivo (MAKINOSHIMA & GLICKMAN, 2005) e tem sido relatado como um determinante da resistência intrínseca a antibióticos compartilhada pela rifampicina, vancomicina e meropenem (XU *et al*., 2017). No estudo do capítulo V, observamos mutações intergênicas relacionadas a SITs/sublinhagens específicas. Essas mutações são pouco estudadas. De fato, os SNPs intergênicos são frequentemente chamados de SNPs neutros, porque, em geral, seu efeito no fenótipo é mínimo. No entanto, SNPs em regiões não codificantes podem ter importantes consequências funcionais (COSCOLLA & GAGNEUX, 2014).

Em geral, o desfecho variável da infecção tem sido atribuído principalmente ao hospedeiro e a fatores ambientais, pois a diversidade genética de *M. tuberculosis* é menos pronunciada do que em outros patógenos. No entanto, sabe-se que a diversidade da cepa pode desempenhar um papel importante durante uma infecção. Foi estabelecido que existem fenótipos metabólicos amplamente diferentes entre as cepas do CMTB. Estudos de correlação da variação genética do CMTB com o fenótipo têm identificado mutações clinicamente relevantes que conferem fenótipos com características importantes, como a resistência as drogas (HICKS *et al*., 2018). Ressalta-se a importância de relacionar a variação genética a fenótipos mais sutis, como no estudo de Øyås *et al*., (2020) que conseguiram prever os efeitos dos SNPs em genes que codificam enzimas em fenótipos metabólicos e identificaram SNPs que estão associados a vulnerabilidades metabólicas. Isso aumenta a importância de estudar o perfil de mutações relacionadas ao metabolismo do ácidos micólicos, que estão correlacionadas com fenótipos mais sutis, como aqueles relacionados à adaptação ao hospedeiro, transmissibilidade e resistência basal, que podem

ser utilizadas como uma possibilidade para a identificação de uma estratégia de tratamento mais seletiva.

Em síntese, no capítulo V foi possível identificar um amplo repertório de SNPs em genes relacionados ao metabolismo do ácido micólico em isolados de *M. tuberculosis* e descrever, pela primeira vez, a variabilidade entre diferentes SITs/sublinhagens da Linhagem 4 circulantes na Grande Florianópolis, Santa Catarina, Brasil. Embora tenha sido sugerido que o background genético influencia o metabolismo do ácidos micólicos, mais estudos são necessários para elucidar até que ponto as diferenças genômicas afetam as propriedades fenotípicas de *M. tuberculosis*, particularmente no que diz respeito à resistência antimicrobiana.

5 CONCLUSÃO

Os importantes achados quanto aos padrões de transmissão de *M. tuberculosis*, aquisição de resistência e demandas identificadas para o melhor manejo do controle da TB resistente, sinalizam a necessidade da inclusão da vigilância genômica da TB como estratégia permanente dos programas de controle da TB na esfera nacional e nas unidades federadas. Tal estratégia deve ser orientada por um fluxo estabelecido no sentido de orientar os laboratórios de saúde pública sobre como deve ser a seleção de amostras elegíveis e de interesse para o sequenciamento, quando submeter as amostras ao sequenciamento, e principalmente definir uma estratégia de análise dos dados gerados.

Neste trabalho, os resultados obtidos pelo sequenciamento e análise bioinformática dos dados geraram informações suficientes para auxiliar o direcionamento das medidas de controle da TB em diferentes contextos no sul do Brasil. Nos diferentes capítulos desta tese nós identificamos os principais genótipos causando resistência às drogas anti-TB e elucidamos as bases moleculares da resistência ocorrendo em Santa Catarina e Rio Grande do Sul. Os achados apontaram para a necessidade de ampliar o diagnóstico da resistência por testes moleculares para detecção oportuna daqueles genótipos que podem não ser identificados pelos testes fenotípicos, como no caso de *disputed mutations*, evitando a subnotificação de casos de TB-MDR. Em relação ao contexto filogenético das principais cepas circulando no sul do Brasil, as análises baseadas em NGS foram capazes de melhor classificar essas cepas entre as linhagens e sublinhagens do CMTB e proporcionaram uma

melhor caracterização das sublinhagens que estão originando os principais eventos de transmissão contínua ocorrendo na região e os genótipos mais associados à resistência.

Os resultados aqui apresentados também elucidaram os principais eventos de aquisição de resistência ocorrendo em cepas multirresistentes, e revelou a amplificação da resistência em diferentes cadeias de transmissão pela aquisição de mutações provávelmente ocorrendo durante o tratamento ou após o abandono de tratamento. Esse fato evidência a necessidade de uma análise minuciosa de quais fatores podem estar contribuindo para esses eventos de aquisição de mutações e consequente amplificação da resistência, para assim, tornar possível direcionar as medidas de controle da TB-MDR na região. Da mesma forma, ao elucidarmos a dinâmica de transmissão de *M. tuberculosis* através da identificação de clusters genômicos e dispersão geotemporal, evidenciamos que se faz necessária a adoção de medidas visando a interrupção das importantes cadeias de transmissão aqui relatadas, incluindo cepas sensíveis as drogas como em Santa Catarina e multirresistentes no Rio Grande do Sul.

6 REFERÊNCIAS BIBLIOGRÁFICAS

ALLIX-BÉGUEC, C.; ARANDJELOVIC, I.; BI, L.; BECKERT, P.; BONNET, M.; BRADLEY, P.; CABIBBE, A. M.; CANCINO-MUÑOZ, I.; CAULFIELD, M. J.; CHAIPRASERT, A.; CIRILLO, D. M.; CLIFTON, D. A.; COMAS, I.; CROOK, D. W.; DE FILIPPO, M. R.; DE NEELING, H.; DIEL, R.; DROBNIEWSKI, F. A.; FAKSRI, K.; … ZHU, B. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. *New England Journal of Medicine*, *379*(15), 1403–1415, 2018.

BJORN-MORTENSEN, K.; SOBORG, B.; KOCH, A.; LADEFOGED, K.; MERKER, M.; LILLEBAEK, T.; ANDERSEN, A. B.; NIEMANN, S.; & KOHL, T. A. Tracing Mycobacterium tuberculosis transmission by whole genome sequencing in a high incidence setting: A retrospective population-based study in East Greenland. *Scientific Reports*, 6, 33180, 2016.

BLOUIN, Y.; HAUCK, Y.; SOLER, C.; FABRE, M.; VONG, R.; DEHAN, C.; CAZAJOUS, G.; MASSOURE, P. L.; KRAEMER, P.; JENKINS, A.; GARNOTEL, E.; POURCEL, C.; & VERGNAUD, G. Significance of the Identification in the Horn of Africa of an Exceptionally Deep Branching Mycobacterium tuberculosis Clade. *PLOS ONE*, *7*(12), e52841, 2012.

BOS, K. I.; HARKINS, K. M.; HERBIG, A.; COSCOLLA, M.; WEBER, N.; COMAS, I.; FORREST, S. A.; BRYANT, J. M.; HARRIS, S. R.; SCHUENEMANN, V. J.; CAMPBELL, T. J.; MAJANDER, K.; WILBUR, A. K.; GUICHON, R. A.; STEADMAN, D. L. W.; COOK, D. C.; NIEMANN, S.; BEHR, M. A.; ZUMARRAGA, M.; … KRAUSE, J. Pre-Columbian mycobacterial genomes reveal seals as a source of New World human tuberculosis. *Nature*, 514(7523), 494–497, 2014.

BOTTAI, D.; FRIGUI, W.; SAYES, F.; DI LUCA, M.; SPADONI, D.; PAWLIK, A.; ZOPPO, M.; ORGEUR, M.; KHANNA, V.; HARDY, D.; MANGENOT, S.; BARBE, V.; MEDIGUE, C.; MA, L.; BOUCHIER, C.; TAVANTI, A.; LARROUY-MAUMUS, G.; & BROSCH, R. TbD1 deletion as a driver of the evolutionary success of modern epidemic Mycobacterium tuberculosis lineages. *Nature Communications*, 11(1), 1–14, 2020.

BOTTAI, D.; STINEAR, T. P.; SUPPLY, P.; & BROSCH, R. Mycobacterial Pathogenomics and Evolution. *Microbiology Spectrum*, 2(1), 2014.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. *Boletim Epidemiológico da Tuberculose 2020.* Ministério da Saúde, 2020.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. *Boletim Epidemiológico da Tuberculose 2022.* Ministério da Saúde, 2022.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. *Boletim Epidemiológico da Tuberculose 2017.* Ministério da Saúde, 2017.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. *Manual de recomendações para o controle da tuberculose no Brasil*. 1a ed. Ministério da Saúde, 2011.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. *Manual de recomendações para o controle da tuberculose no Brasil*. 2a ed. Ministério da Saúde, 2019.

BRITES, D.; & GAGNEUX, S. Co-evolution of Mycobacterium tuberculosis and Homo sapiens. *Immunological Reviews*, 264(1), 6–24, 2015.

BRITES, D.; LOISEAU, C.; MENARDO, F.; BORRELL, S.; BONIOTTI, M. B.; WARREN, R.; DIPPENAAR, A.; PARSONS, S. D. C.; BEISEL, C.; BEHR, M. A.; FYFE, J. A.; COSCOLLA, M.; & GAGNEUX, S. A new phylogenetic framework for the animal-adapted mycobacterium tuberculosis complex. *Frontiers in Microbiology*, 9, 2820, 2018.

BROSCH, R.; GORDON, S. V.; MARMIESSE, M.; BRODIN, P.; BUCHRIESER, C.; EIGLMEIER, K.; GARNIER, T.; GUTIERREZ, C.; HEWINSON, G.; KREMER, K.; PARSONS, L. M.; PYM, A. S.; SAMPER, S.; VAN SOOLINGEN, D.; & COLE, S. T. A new evolutionary scenario for the Mycobacterium tuberculosis complex. *Proceedings of the National Academy of Sciences of the United States of America*, 99(6), 3684–3689, 2002.

BRUDEY, K.; DRISCOLL, J. R.; RIGOUTS, L.; PRODINGER, W. M.; GORI, A.; AL-HAJOJ, S. A.; ALLIX, C.; ARISTIMUÑO, L.; ARORA, J.; BAUMANIS, V.; BINDER, L.; CAFRUNE, P.; CATALDI, A.; CHEONG, S.; DIEL, R.; ELLERMEIER, C.; EVANS, J. T.; FAUVILLE-DUFAUX, M.; FERDINAND, S.; … SOLA, C. Mycobacterium tuberculosis complex genetic diversity: Mining the fourth international spoligotyping database (SpolDB4) for classification, population genetics and epidemiology. *BMC Microbiology*, 6, 1–17, 2006.

BRYNILDSRUD, O. B.; PEPPERELL, C. S.; SUFFYS, P.; GRANDJEAN, L.; MONTESERIN, J.; DEBECH, N.; BOHLIN, J.; ALFSNES, K.; PETTERSSON, J. O. H.; KIRKELEITE, I.; FANDINHO, F.; DA SILVA, M. A.; PERDIGAO, J.; PORTUGAL, I.; VIVEIROS, M.; CLARK, T.; CAWS, M.; DUNSTAN, S.; THAI, P. V. K.; … ELDHOLM, V. Global expansion of Mycobacterium tuberculosis lineage 4 shaped by colonial migration and local adaptation. *Science Advances*, 4(10), 1–12, 2018.

CAMUS, J. C.; PRYOR, M. J.; MÉDIGUE, C.; & COLE, S. T. Re-annotation of the genome sequence of Mycobacterium tuberculosis H37Rv. *Microbiology (Reading, England)*, 148(10), 2967–2973, 2002.

CHAVES TORRES, N. M.; QUIJANO RODRÍGUEZ, J. J.; PORRAS ANDRADE, P. S.; ARRIAGA, M. B.; & NETTO, E. M. Factors predictive of the success of tuberculosis treatment: A systematic review with meta-analysis. *PloS One*, 14(12), e0226507, 2019.

COLE, S. T.; BROSCH, R.; PARKHILL, J.; GARNIER, T.; CHURCHER, C.; HARRIS, D.; GORDON, S. V.; EIGLMEIER, K.; GAS, S.; BARRY, C. E.; TEKAIA, F.; BADCOCK, K.; BASHAM, D.; BROWN, D.; CHILLINGWORTH, T.;

CONNOR, R.; DAVIES, R.; DEVLIN, K.; FELTWELL, T.; … BARRELL, B. G. Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence. *Nature*, 393(6685), 537–544, 1998.

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. *Nature Communications*, 5, 4–8, 2014.

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.; PORTELLI, S.; OPPONG, Y.; ALVES, A.; BESSA, T. B.; CAMPINO, S.; CAWS, M.; CHATTERJEE, A.; CRAMPIN, A. C.; DHEDA, K.; … CLARK, T. G. Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis. *Nature Genetics*, 50(2), 307–316, 2018.

COMAS, I.; & GAGNEUX, S. The Past and Future of Tuberculosis Research. *PLoS Pathogens*, 5(10), 2009.

COMAS, I.; HOMOLKA, S.; NIEMANN, S.; & GAGNEUX, S. Genotyping of genetically monomorphic bacteria: DNA sequencing in Mycobacterium tuberculosis highlights the limitations of current methodologies. *PLoS ONE*, 4(11), 2009.

COSCOLLA, M. Biological and epidemiological consequences of MTBC diversity. *Advances in Experimental Medicine and Biology*, 1019, 95-116, 2017.

COSCOLLA, M.; & GAGNEUX, S. Consequences of genomic diversity in mycobacterium tuberculosis. *Seminars in Immunology*, 26(6), 431–444, 2014.

COSTA-VEIGA, A.; BRIZ, T.; & NUNES, C. Unsuccessful treatment in pulmonary tuberculosis: factors and a consequent predictive model. *European Journal of Public Health*, 28(2), 352–358, 2018.

COUVIN, D.; & RASTOGI, N. The establishment of databases on circulating genotypes of Mycobacterium tuberculosis complex and web tools for an effective response to better monitor, understand and control the tuberculosis epidemic worldwide. *EuroReference - Journal of Reference*, 12, 36–48, 2014.

COWAN, L. S.; DIEM, L.; MONSON, T.; WAND, P.; TEMPORADO, D.; OEMIG, T. V.; & CRAWFORD, J. T. Evaluation of a Two-Step Approach for Large-Scale, Prospective Genotyping of Mycobacterium tuberculosis Isolates in the United States. *Journal of Clinical Microbiology*, 43(2), 688, 2005.

DALLA COSTA, E. R.; VASCONCELOS, S. E. G.; ESTEVES, L. S.; GOMES, H. M.; GOMES, L. L.; DA SILVA, P. A.; PERDIGÃO, J.; PORTUGAL, I.; VIVEIROS, M.; MCNERNEY, R.; PAIN, A.; CLARK, T. G.; RASTOGI, N.; UNIS, G.; ROSSETTI, M. L. R.; & SUFFYS, P. N. Multidrug-Resistant Mycobacterium tuberculosis of the Latin American Mediterranean Lineage, Wrongly Identified as Mycobacterium pinnipedii (Spoligotype International Type 863 [SIT863]), Causing Active Tuberculosis in South Brazil. *Journal of Clinical Microbiology*, 53(12), 3805–3811, 2015.

DE JONG, B. C.; ANTONIO, M.; & GAGNEUX, S. Mycobacterium africanum—Review of an Important Cause of Human Tuberculosis in West Africa. *PLOS Neglected Tropical Diseases*, 4(9), e744, 2010.

DEMAY, C.; LIENS, B.; BURGUIÈRE, T.; HILL, V.; COUVIN, D.; MILLET, J.; MOKROUSOV, I.; SOLA, C.; ZOZIO, T.; & RASTOGI, N. SITVITWEB - A publicly available international multimarker database for studying Mycobacterium tuberculosis genetic diversity and molecular epidemiology. *Infection, Genetics and Evolution*, 12(4), 755–766, 2012.

DEVASIA, R. A.; BLACKMAN, A.; GEBRETSADIK, T.; GRIFFIN, M.; SHINTANI, A.; MAY, C.; SMITH, T.; HOOPER, N.; MARURI, F.; WARKENTIN, J.; MITCHEL, E.; & STERLING, T. R. Fluoroquinolone resistance in Mycobacterium tuberculosis: the effect of duration and timing of fluoroquinolone exposure. *American Journal of Respiratory and Critical Care Medicine*, 180(4), 365-70, 2009.

DONOGHUE, H. D. Insights gained from palaeomicrobiology into ancient and modern tuberculosis. *Clinical Microbiology and Infection : The Official Publication of the European Society of Clinical Microbiology and Infectious Diseases*, 17(6), 821–829, 2011.

DOOKIE, N.; RAMBARAN, S.; PADAYATCHI, N.; MAHOMED, S.; & NAIDOO, K. Evolution of drug resistance in Mycobacterium tuberculosis: A review on the molecular determinants of resistance and implications for personalized care. *Journal of Antimicrobial Chemotherapy*, 73(5), 1138–1151, 2018.

DRUSZCZYNSKA, M.; WAWROCKI, S.; SZEWCZYK, R.; & RUDNICKA, W. Mycobacteria-derived biomarkers for tuberculosis diagnosis. *The Indian Journal of Medical Research*, 146(6), 700–707, 2017.

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.; & ROSSETTI, M. L. R. Genetic diversity of Mycobacterium tuberculosis isoniazid monoresistant and multidrug-resistant in Rio Grande do Sul, a tuberculosis high-burden state in Brazil. *Tuberculosis*, 110, 36–43, 2018.

FIRDESSA, R.; BERG, S.; HAILU, E.; SCHELLING, E.; GUMI, B.; ERENSO, G.; GADISA, E.; KIROS, T.; HABTAMU, M.; HUSSEIN, J.; ZINSSTAG, J.; ROBERTSON, B. D.; AMENI, G.; LOHAN, A. J.; LOFTUS, B.; COMAS, I.; GAGNEUX, S.; TSCHOPP, R.; YAMUAH, L.; … ASEFFA, A. Mycobacterial lineages causing pulmonary and extrapulmonary tuberculosis, Ethiopia. *Emerging Infectious Diseases*, 19(3), 460–463, 2013.

FLEMING, M. F.; KRUPITSKY, E.; TSOY, M.; ZVARTAU, E.; BRAZHENKO, N.; JAKUBOWIAK, W.; & MCCAUL, M. E. Alcohol and drug use disorders, HIV status and drug resistance in a sample of Russian TB patients. *The International Journal of Tuberculosis and Lung Disease*, 10(5), 565–570, 2006.

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. *Nature Genetics*, 45(7), 784–790, 2013.

GAGNEUX, S.; & SMALL, P. M. Global phylogeography of Mycobacterium tuberculosis and implications for tuberculosis product development. *Lancet Infectious Diseases*, 7(5), 328–337, 2007.

GAGNEUX, S.; DERIEMER, K.; VAN, T.; KATO-MAEDA, M.; DE JONG, B. C.; NARAYANAN, S.; NICOL, M.; NIEMANN, S.; KREMERI, K.; GUTIERREZ, M. C.; HILTY, M.; HOPEWELL, P. C.; & SMALL, P. M. Variable host-pathogen compatibility in Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America*, 103(8), 2869–2873, 2006.

GAGNEUX, S.; LONG, C. D.; SMALL, P. M.; VAN, T.; SCHOOLNIK, G. K.; & BOHANNAN, B. J. M. The competitive cost of antibiotic resistance in Mycobacterium tuberculosis. *Science*, 312(5782), 1944–1946, 2006.

GEHRE, F.; ANTONIO, M.; FAÏHUN, F.; ODOUN, M.; UWIZEYE, C.; DE RIJK, P.; DE JONG, B. C.; & AFFOLABI, D. The First Phylogeographic Population Structure and Analysis of Transmission Dynamics of M. africanum West African 1—Combining Molecular Data from Benin, Nigeria and Sierra Leone. *PLOS ONE*, 8(10), e77000, 2013.

GETAHUN, H.; MATTEELLI, A.; CHAISSON, R. E.; & RAVIGLIONE, M. Latent Mycobacterium tuberculosis Infection. *New England Journal of Medicine*, 372(22), 2127–2135, 2015.

GLICKMAN, M. S.; COX, J. S.; & JACOBS, W. R. A novel mycolic acid cyclopropane synthetase is required for cording, persistence, and virulence of Mycobacterium tuberculosis. *Molecular Cell*, 5(4), 717–727, 2000.

GUENIN-MACÉ, L.; SIMÉONE, R.; & DEMANGEL, C. Lipids of Pathogenic Mycobacteria: Contributions to Virulence and Host Immune Suppression. *Transboundary and Emerging Diseases*, 56(6–7), 255–268, 2009.

HERSHKOVITZ, I.; DONOGHUE, H. D.; MINNIKIN, D. E.; MAY, H.; LEE, O. Y. C.; FELDMAN, M.; GALILI, E.; SPIGELMAN, M.; ROTHSCHILD, B. M.; & BAR-GAL, G. K. Tuberculosis origin: The Neolithic scenario. *Tuberculosis*, 95(1), 122–126, 2015.

HICKS, N. D.; YANG, J.; ZHANG, X.; ZHAO, B.; GRAD, Y. H.; LIU, L.; OU, X.; CHANG, Z.; XIA, H.; ZHOU, Y.; WANG, S.; DONG, J.; SUN, L.; ZHU, Y.; ZHAO, Y.; JIN, Q.; & FORTUNE, S. M. Clinically prevalent mutations in Mycobacterium tuberculosis alter propionate metabolism and mediate multidrug tolerance. *Nature Microbiology*, 3(9), 1032–1042, 2018.

HUITRIC, E.; WERNGREN, J.; JURÉEN, P.; & HOFFNER, S. Resistance levels and rpoB gene mutations among in vitro-selected rifampin-resistant Mycobacterium tuberculosis mutants. *Antimicrobial Agents and Chemotherapy*, 50(8), 2860–2862, 2006.

IBGE. Instituto Brasileiro de Geografia e Estatística. Cidades. Disponível em: https://cidades.ibge.gov.br/brasil/panorama.

JO, K. W.; LEE, S.; KANG, M. R.; SUNG, H.; KIM, M. N.; & SHIM, T. S. Frequency and type of disputed rpoB mutations in Mycobacterium tuberculosis isolates from South Korea. *Tuberculosis and Respiratory Diseases*, 80(3), 272–276, 2017.

JONSSON, J.; HOFFNER, S.; BERGGREN, I.; BRUCHFELD, J.; GHEBREMICHAEL, S.; PENNHAG, A.; & GROENHEIT, R. Comparison between RFLP and MIRU-VNTR Genotyping of Mycobacterium tuberculosis Strains Isolated in Stockholm 2009 to 2011. *PLOS ONE*, 9(4), e95159, 2014.

KAMERBEEK, J.; SCHOULS, L. E. O.; KOLK, A.; AGTERVELD, M. V. A. N.; SOOLINGEN, D. V. A. N.; KUIJPER, S.; BUNSCHOTEN, A.; MOLHUIZEN, H.; SHAW, R.; & GOYAL, M. *Simultaneous Detection and Strain Differentiation of Mycobacterium tuberculosis for Diagnosis and Epidemiology*. 35(4), 907–914, 1997.

KIM, H.; MOK, J. H.; KANG, B.; LEE, T.; LEE, H.-K.; JANG, H. J.; CHO, Y. J.; & JEON, D. Trend of multidrug and fluoroquinolone resistance in Mycobacterium tuberculosis isolates from 2010 to 2014 in Korea: a multicenter study. *The Korean Journal of Internal Medicine*, 34(2), 344–352, 2019.

KOLATTUKUDY, P. E.; FERNANDES, N. D.; AZAD, A. K.; FITZMAURICE, A. M.; & SIRAKOVA, T. D. Biochemistry and molecular genetics of cell-wall lipid biosynthesis in mycobacteria. *Molecular Microbiology*, 24(2), 263–270, 1997.

KREMER, K.; VAN SOOLINGEN, D.; FROTHINGHAM, R.; HAAS, W. H.; HERMANS, P. W. M.; MARTÍN, C.; PALITTAPONGARNPIM, P.; PLIKAYTIS, B. B.; RILEY, L. W.; YAKRUS, M. A.; MUSSER, J. M.; & VAN EMBDEN, J. D. A. Comparison of Methods Based on Different Molecular Epidemiological Markers for Typing of Mycobacterium tuberculosis Complex Strains: Interlaboratory Study of Discriminatory Power and Reproducibility. *Journal of Clinical Microbiology*, 37(8), 2607, 1999.

LEHMANN, J. Para-aminosalicylic acid in the treatment of tuberculosis. *Lancet*, 1(6384), 15–16, 1946.

LIN, W. H.; LEE, W. T.; TSAI, H. Y.; & JOU, R. Disputed rpoB Mutations in Mycobacterium tuberculosis and Tuberculosis Treatment Outcomes. *Antimicrobial Agents and Chemotherapy*, 65(7), 2021.

LIU, L.; JIANG, F.; CHEN, L.; ZHAO, B.; DONG, J.; SUN, L.; ZHU, Y.; LIU, B.; ZHOU, Y.; YANG, J.; ZHAO, Y.; JIN, Q.; & ZHANG, X. The impact of combined gene mutations in *inhA* and *ahpC* genes on high levels of isoniazid resistance amongst *katG* non-315 in multidrug-resistant tuberculosis isolates from China. *Emerging Microbes & Infections*, 7(1), 1–10, 2018.

LONG, R.; CHONG, H.; HOEPPNER, V.; SHANMUGANATHAN, H.; KOWALEWSKA-GROCHOWSKA, K.; SHANDRO, C.; MANFREDA, J.; SENTHILSELVAN, A.; ELZAINY, A.; & MARRIE, T. Empirical Treatment of Community-Acquired Pneumonia and the Development of Fluoroquinolone-Resistant Tuberculosis. *Clinical Infectious Diseases*, 48(10), 1354–1360, 2009.

MACHADO, D.; COUTO, I.; PERDIGÃO, J.; RODRIGUES, L.; PORTUGAL, I.; BAPTISTA, P.; VEIGAS, B.; AMARAL, L.; & VIVEIROS, M. Contribution of Efflux to the Emergence of Isoniazid and Multidrug Resistance in Mycobacterium tuberculosis. *PLOS ONE*, 7(4), e34538, 2012.

MACHADO, D.; PERDIGÃO, J.; PORTUGAL, I.; PIERONI, M.; SILVA, P. A.; COUTO, I.; & VIVEIROS, M. Efflux Activity Differentially Modulates the Levels of Isoniazid and Rifampicin Resistance among Multidrug Resistant and Monoresistant Mycobacterium tuberculosis Strains. *Antibiotics*, 7(1), 1–16, 2018.

MAGIS-ESCURRA, C.; GÜNTHER, G.; LANGE, C.; ALEXANDRU, S.; ALTET, N.; AVSAR, K.; BANG, D.; BARBUTA, R.; BOTHAMLEY, G.; CIOBANU, A.; CRUDU, V.; DAVILOVITS, M.; DEDICOAT, M.; DUARTE, R.; GUALANO, G.; KUNST, H.; DE LANGE, W.; LEIMANE, V.; MCLAUGHLIN, A. M.; … VAN LETH, F. Treatment outcomes of MDR-TB and HIV co-infection in Europe. *European Respiratory Journal*, 49(6), 1602363. European Respiratory Society, 2017.

MAIDEN, M. C. J.; BYGRAVES, J. A.; FEIL, E.; MORELLI, G.; RUSSELL, J. E.; URWIN, R.; ZHANG, Q.; ZHOU, J.; ZURTH, K.; CAUGANT, D. A.; FEAVERS, I. M.; ACHTMAN, M.; & SPRATT, B. G. Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America*, 95(6), 3140, 1998.

MAKINOSHIMA, H.; & GLICKMAN, M. S. Regulation of Mycobacterium tuberculosis cell envelope composition and virulence by intramembrane proteolysis. *Nature*, 436(7049), 406–409, 2005.

MARIAM, D. H.; MENGISTU, Y.; HOFFNER, S. E.; & ANDERSSON, D. I. Effect of rpoB Mutations Conferring Rifampin Resistance on Fitness of Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy*, 48(4), 1289–1294, 2004.

MATHEMA, B.; KUREPINA, N. E.; BIFANI, P. J.; & KREISWIRTH, B. N. Molecular epidemiology of tuberculosis: current insights. *Clinical Microbiology Reviews*, 19(4), 658–685, 2006.

MAXAM, A. M.; & GILBERT, W. A new method for sequencing DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 74(2), 560–564, 1977.

MAZARS, E.; LESJEAN, S.; BANULS, A. L.; GILBERT, M.; VINCENT, V.; GICQUEL, B.; TIBAYRENC, M.; LOCHT, C.; & SUPPLY, P. High-resolution minisatellite-based typing as a portable approach to global analysis of Mycobacterium tuberculosis molecular epidemiology. *Proceedings of the National Academy of Sciences of the United States of America*, 98(4), 1901, 2001.

MEDEIROS, T. F.; NOGUEIRA, C. L.; PRIM, R. I.; SCHEFFER, M. C.; ALVES, E. V.; ROVARIS, D. B.; ZOZIO, T.; RASTOGI, N.; & BAZZO, M. L. Molecular epidemiology of Mycobacterium tuberculosis strains from prison populations in Santa Catarina, Southern Brazil. *Infection, Genetics and Evolution*, 58, 34–39, 2018.

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.; OFORI-ANYINAM, B.; DREYER, V.; SUPPLY, P.; SURESH, A.; UTPATEL, C.; VAN SOOLINGEN, D.; ZHOU, Y.; ASHTON, P. M.; BRITES, D.; … VAN RIE, A. Whole genome sequencing of Mycobacterium tuberculosis: current standards and open issues. *Nature Reviews Microbiology*, 17, 533–545, 2019.

MIOTTO, P.; CABIBBE, A. M.; BORRONI, E.; DEGANO, M.; & CIRILLOA, D. M. Role of disputed mutations in the rpoB gene in interpretation of automated liquid MGIT culture results for rifampin susceptibility testing of mycobacterium tuberculosis. *Journal of Clinical Microbiology*, 56(5), 2018.

MIOTTO, P.; TESSEMA, B.; TAGLIANI, E.; CHINDELEVITCH, L.; STARKS, A. M.; EMERSON, C.; HANNA, D.; KIM, P. S.; LIWSKI, R.; ZIGNOL, M.; GILPIN, C.; NIEMANN, S.; DENKINGER, C. M.; FLEMING, J.; WARREN, R. M.; CROOK, D.; POSEY, J.; GAGNEUX, S.; HOFFNER, S.; … RODWELL, T. C. A standardised method for interpreting the association between mutations and phenotypic drug resistance in Mycobacterium tuberculosis. *European Respiratory Journal*, 50(6), 2017.

MIYAHARA, R.; SMITTIPAT, N.; JUTHAYOTHIN, T.; YANAI, H.; DISRATTHAKIT, A.; IMSANGUAN, W.; INTRALAWAN, D.; NEDSUWAN, S.; CHAIYASIRINROJE, B.; BUPACHAT, S.; TOKUNAGA, K.; MAHASIRIMONGKOL, S.; & PALITTAPONGARNPIM, P. Risk factors associated with large clusters of tuberculosis patients determined by whole-genome sequencing in a high-tuberculosis-burden country. *Tuberculosis*, *125*, 101991, 2020.

MVELASE, N. R.; PILLAY, M.; SIBANDA, W.; NGOZO, J. N.; BRUST, J. C. M.; & MLISANA, K. P. rpoB Mutations Causing Discordant Rifampicin Susceptibility in Mycobacterium tuberculosis: Retrospective Analysis of Prevalence, Phenotypic, Genotypic, and Treatment Outcomes. *Open Forum Infectious Diseases*, 6(4), 2019.

MYERS, B.; BOUTON, T. C.; RAGAN, E. J.; WHITE, L. F.; MCILLERON, H.; THERON, D.; PARRY, C. D. H.; HORSBURGH, C. R.; WARREN, R. M.; & JACOBSON, K. R. Impact of alcohol consumption on tuberculosis treatment outcomes: A prospective longitudinal cohort study protocol. *BMC Infectious Diseases*, 18(1), 1–9, 2018.

NAPIER, G.; CAMPINO, S.; MERID, Y.; ABEBE, M.; WOLDEAMANUEL, Y.; ASEFFA, A.; HIBBERD, M. L.; PHELAN, J.; & CLARK, T. G. Robust barcoding and identification of Mycobacterium tuberculosis lineages for epidemiological and clinical studies. *Genome Medicine*, 12(1), 1–10, 2020.

NGABONZIZA, J. C. S.; LOISEAU, C.; MARCEAU, M.; JOUET, A.; MENARDO, F.; TZFADIA, O.; ANTOINE, R.; NIYIGENA, E. B.; MULDERS, W.; FISSETTE, K.; DIELS, M.; GAUDIN, C.; DUTHOY, S.; SSENGOOBA, W.; ANDRÉ, E.; KASWA, M. K.; HABIMANA, Y. M.; BRITES, D.; AFFOLABI, D.; … SUPPLY, P. A sister lineage of the Mycobacterium tuberculosis complex discovered in the African Great Lakes region. *Nature Communications*, 11(1), 2020.

NGUYEN, L. Antibiotic resistance mechanisms in M. tuberculosis: an update. *Archives of Toxicology*, 90(7), 1585–1604, 2016.

NIEMANN, S.; & SUPPLY, P. Diversity and evolution of Mycobacterium tuberculosis: moving to whole-genome-based approaches. *Cold Spring Harbor Perspectives in Medicine*, 4(12), 2014.

NIKOLAYEVSKYY, V.; NIEMANN, S.; ANTHONY, R.; VAN SOOLINGEN, D.; TAGLIANI, E.; KÖDMÖN, C.; VAN DER WERF, M. J.; & CIRILLO, D. M. Role and value of whole genome sequencing in studying tuberculosis transmission. *Clinical Microbiology and Infection*, 25(11), 1377–1382, 2019.

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*, 97, 57–64, 2016.

NUNES-ALVES, C.; BOOTY, M. G.; CARPENTER, S. M.; JAYARAMAN, P.; ROTHCHILD, A. C.; & BEHAR, S. M. In search of a new paradigm for protective immunity to TB. *Nature Reviews. Microbiology*, 12(4), 289–299, 2014.

ØYÅS, O.; BORRELL, S.; TRAUNER, A.; ZIMMERMANN, M.; FELDMANN, J.; LIPHARDT, T.; GAGNEUX, S.; STELLING, J.; SAUER, U.; & ZAMPIERI, M. Model-based integration of genomics and metabolomics reveals SNP functionality in Mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences of the United States of America*, 117(15), 8494–8502, 2020.

PANKHURST, L. J.; DEL OJO ELIAS, C.; VOTINTSEVA, A. A.; WALKER, T. M.; COLE, K.; DAVIES, J.; FERMONT, J. M.; GASCOYNE-BINZI, D. M.; KOHL, T. A.; KONG, C.; LEMAITRE, N.; NIEMANN, S.; PAUL, J.; ROGERS, T. R.; ROYCROFT, E.; SMITH, E. G.; SUPPLY, P.; TANG, P.; WILCOX, M. H.; … CROOK, D. W. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: A prospective study. *The Lancet Respiratory Medicine*, 4(1), 49–58, 2016.

PAREEK, C. S.; SMOCZYNSKI, R.; & TRETYN, A. Sequencing technologies and genome sequencing. *Journal of Applied Genetics*, 52(4), 413–435, 2011.

PEPPERELL, C. S.; CASTO, A. M.; KITCHEN, A.; GRANKA, J. M.; CORNEJO, O. E.; HOLMES, E. C.; BIRREN, B.; GALAGAN, J.; & FELDMAN, M. W. The Role of Selection in Shaping Diversity of Natural M. tuberculosis Populations. *PLOS Pathogens*, 9(8), e1003543, 2013.

PERDIGÃO, J.; GOMES, P.; MIRANDA, A.; MALTEZ, F.; MACHADO, D.; SILVA, C.; PHELAN, J. E.; BRUM, L.; CAMPINO, S.; COUTO, I.; VIVEIROS, M.; CLARK, T. G.; & PORTUGAL, I. Using genomics to understand the origin and dispersion of multidrug and extensively drug resistant tuberculosis in Portugal. *Scientific Reports*, 10(1), 2600, 2020.

PERDIGÃO, J.; SILVA, C.; DINIZ, J.; PEREIRA, C.; MACHADO, D.; RAMOS, J.; SILVA, H.; ABILLEIRA, F.; BRUM, C.; REIS, A. J.; MACEDO, M.; SCAINI, J. L.; SILVA, A. B.; ESTEVES, L.; MACEDO, R.; MALTEZ, F.; CLEMENTE, S.; COELHO, E.; VIEGAS, S.; … PORTUGAL, I. Clonal expansion across the seas as seen through CPLP-TB database: A joint effort in cataloguing Mycobacterium

tuberculosis genetic diversity in Portuguese-speaking countries. *Infection, Genetics and Evolution*, 72, 44-58, 2018.

PERIZZOLO, P. F.; DALLA COSTA, E. R.; RIBEIRO, A. W.; SPIES, F. S.; RIBEIRO, M. O.; DIAS, C. F.; UNIS, G.; ALMEIDA DA SILVA, P.; GOMES, H. M.; SUFFYS, P. N.; & ROSSETTI, M. L. R. Characteristics of multidrug-resistant Mycobacterium tuberculosis in southern Brazil. *Tuberculosis*, 92(1), 56–59, 2012.

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.; HUGGETT, J. F.; & CLARK, T. G. Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. *Genome Medicine*, 11(1), 41, 2019.

PORTEVIN, D.; SUKUMAR, S.; COSCOLLA, M.; SHUI, G.; LI, B.; GUAN, X. L.; BENDT, A. K.; YOUNG, D.; GAGNEUX, S.; & WENK, M. R. Lipidomics and genomics of Mycobacterium tuberculosis reveal lineage-specific trends in mycolic acid biosynthesis. *MicrobiologyOpen*, 3(6), 823–835, 2014.

PRIM, R. I.; SCHÖRNER, M. A.; SENNA, S. G.; NOGUEIRA, C. L.; FIGUEIREDO, A. C. C.; DE OLIVEIRA, J. G.; ROVARIS, D. B.; & BAZZO, M. L. Molecular profiling of drug resistant isolates of Mycobacterium tuberculosis in the state of Santa Catarina, southern Brazil. *Memorias Do Instituto Oswaldo Cruz*, 110(5), 618–623, 2015.

PURYEAR, S. B.; BALZER, L. B.; AYIEKO, J.; KWARISIIMA, D.; HAHN, J. A.; CHARLEBOIS, E. D.; CLARK, T. D.; COHEN, C. R.; BUKUSI, E. A.; KAMYA, M. R.; PETERSEN, M. L.; HAVLIR, D. V.; & CHAMIE, G. Associations between alcohol use and HIV care cascade outcomes among adults undergoing population-based HIV testing in East Africa. *AIDS*, 34(3), 405–413, 2020.

REILING, N.; HOMOLKA, S.; WALTER, K.; BRANDENBURG, J.; NIWINSKI, L.; ERNST, M.; HERZMANN, C.; LANGE, C.; DIEL, R.; EHLERS, S.; & NIEMANN, S. Clade-specific virulence patterns of Mycobacterium tuberculosis complex strains in human primary macrophages and aerogenically infected mice. *MBio*, 4(4), 2013.

ROSSETTI, M. L.; ALMEIDA DA SILVA, P. E.; SALVATO, R. S.; REIS, A. J.; SCHIEFELBEIN, S. H.; VON GROLL, A.; BARCELLOS, R. B.; MASCHMANN, R.; ESTEVES, L. S.; SPIES, F.; TRESPACH, R. R.; DALLA COSTA, E. R.; & NEUBAUER DE AMORIM, H. L. A highly rifampicin resistant Mycobacterium tuberculosis strain emerging in Southern Brazil. *Tuberculosis*, 125, 102015, 2020.

SACCHI, F. P. C.; PRACA, R. M.; TATARA, M. B.; SIMONSEN, V.; FERRAZOLI, L.; CRODA, M. G.; SUFFYS, P. N.; KO, A. I.; ANDREWS, J. R.; & CRODA, J. Prisons as reservoir for community transmission of tuberculosis, Brazil. *Emerging Infectious Diseases*, 21(3), 452–455, 2015.

SAKULA, A. Robert Koch: centenary of the discovery of the tubercle bacillus, 1882. *Thorax*, 37(4), 246–251, 1982.

SALVATO, R. S.; COSTA, E. R. D.; REIS, A. J.; SCHIEFELBEIN, S. H.; HALON, M. L.; BARCELLOS, R. B.; UNIS, G.; DIAS, C. F.; VIVEIROS, M.; PORTUGAL, I.; DA SILVA, P. E. A.; KRITSKI, A. L.; PERDIGÃO, J.; & ROSSETTI, M. L. R. First insights into circulating XDR and pre-XDR Mycobacterium tuberculosis in Southern Brazil. *Infection, Genetics and Evolution*, 78, 104127, 2020.

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.; DE ALMEIDA, I. N.; DE ASSIS FIGUEREDO, L. J.; SILVA, E. C.; KRITSKI, A. L.; DALLA COSTA, E. R.; & ROSSETTI, M. L. R. Molecular characterisation of multidrug-resistant Mycobacterium tuberculosis isolates from a high-burden tuberculosis state in Brazil. *Epidemiology and Infection*, 147, e216, 2019.

SAMBANDAN, D.; DAO, D. N.; WEINRICK, B. C.; VILCHÈZE, C.; GURCHA, S. S.; OJHA, A.; KREMER, L.; BESRA, G. S.; HATFULL, G. F.; & JACOBS, W. R. Keto-mycolic acid-dependent pellicle formation confers tolerance to drug-sensitive Mycobacterium tuberculosis. *MBio*, 4(3), 2013.

SANDGREN, A.; LONNROTH, K.; PIRET, V.; PIERPAOLO, DE C.; & ALJONA, K. Collaborative action on tuberculosis and alcohol abuse in Estonia First report of a demonstration project. *WHO Regional Office for Europe*, 2013.

SANGER, F.; NICKLEN, S.; & COULSON, A. R. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, 74(12), 5463–5467, 1977.

SCHATZ, A.; BUGIE, E.; & WAKSMAN, S. A. Streptomycin, a substance exhibiting antibiotic activity against gram-positive and gram-negative bacteria. 1944. *Clinical Orthopaedics and Related Research*, 437(437), 3–6, 2005.

SEIFERT, M.; CATANZARO, D.; CATANZARO, A.; & RODWELL, T. C. Genetic mutations associated with isoniazid resistance in Mycobacterium tuberculosis: A systematic review. *PLoS ONE*, 10(3), 1–13, 2015.

SINGH, R.; DWIVEDI, S. P.; GAHARWAR, U. S.; MEENA, R.; RAJAMANI, P.; & PRASAD, T. Recent updates on drug resistance in Mycobacterium tuberculosis. *Journal of Applied Microbiology*, 128(6), 1547–1567, 2020.

STUCKI, D.; BRITES, D.; JELJELI, L.; COSCOLLA, M.; LIU, Q.; TRAUNER, A.; FENNER, L.; RUTAIHWA, L.; BORRELL, S.; LUO, T.; GAO, Q.; KATO-MAEDA, M.; BALLIF, M.; EGGER, M.; MACEDO, R.; MARDASSI, H.; MORENO, M.; VILANOVA, G. T.; FYFE, J.; … GAGNEUX, S. Mycobacterium tuberculosis lineage 4 comprises globally distributed and geographically restricted sublineages. *Nature Genetics*, 48(12), 1535–1543, 2016.

SUPPLY, P.; ALLIX, C.; LESJEAN, S.; CARDOSO-OELEMANN, M.; RÜSCH-GERDES, S.; WILLERY, E.; SAVINE, E.; DE HAAS, P.; VAN DEUTEKOM, H.; RORING, S.; BIFANI, P.; KUREPINA, N.; KREISWIRTH, B.; SOLA, C.; RASTOGI, N.; VATIN, V.; GUTIERREZ, M. C.; FAUVILLE, M.; NIEMANN, S.; … VAN SOOLINGEN, D. Proposal for standardization of optimized mycobacterial interspersed repetitive unit-variable-number tandem repeat typing of Mycobacterium tuberculosis. *Journal of Clinical Microbiology*, 44(12), 4498–4510, 2006.

THIERRY, D.; CAVE, M. D.; EISENACH, K. D.; CRAWFORD, J. T.; BATES, J. H.; GICQUEL, B.; & GUESDON, J. L. IS6110, an IS-like element of Mycobacterium tuberculosis complex. *Nucleic Acids Research*, 18(1), 188, 1990.

TORREA, G.; NG, K. C. S.; VAN DEUN, A.; ANDRÉ, E.; KAISERGRUBER, J.; SSENGOOBA, W.; DESMARETZ, C.; GABRIELS, S.; DRIESEN, M.; DIELS, M.; ASNONG, S.; FISSETTE, K.; GUMUSBOGA, M.; RIGOUTS, L.; AFFOLABI, D.; JOLOBA, M.; & DE JONG, B. C. Variable ability of rapid tests to detect Mycobacterium tuberculosis rpoB mutations conferring phenotypically occult rifampicin resistance. *Scientific Reports 2019*, 9(1), 1–9, 2019.

UMPELEVA, T.; BELOUSOVA, K.; GOLUBEVA, L.; BOTEVA, T.; MOROZOVA, I.; VYAZOVAYA, A.; MOKROUSOV, I.; EREMEEVA, N.; & VAKHRUSHEVA, D. Molecular characteristics of Mycobacterium tuberculosis in the "closed" Russian town with limited population migration. *Infection, Genetics and Evolution*, 79, 104174, 2020.

VAN DEUN, A.; AUNG, K. J. M.; BOLA, V.; LEBEKE, R.; HOSSAIN, M. A.; DE RIJK, W. B.; RIGOUTS, L.; GUMUSBOGA, A.; TORREA, G.; & DE JONG, B. C. Rifampin Drug Resistance Tests for Tuberculosis: Challenging the Gold Standard. *Journal of Clinical Microbiology*, 51(8), 2633–2640, 2013.

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.; & SMALL, P. M. Strain identification of Mycobacterium tuberculosis by DNA fingerprinting: Recommendations for a standardized methodology. *Journal of Clinical Microbiology*, 31(2), 406–409, 1993.

VERZA, M.; SCHEFFER, M. C.; SALVATO, R. S.; SCHORNER, M. A.; BARAZZETTI, F. H.; MACHADO, H. DE M.; MEDEIROS, T. F.; ROVARIS, D. B.; PORTUGAL, I.; VIVEIROS, M.; PERDIGÃO, J.; KRITSKI, A.; & BAZZO, M. L. Genomic epidemiology of Mycobacterium tuberculosis in Santa Catarina, Southern Brazil. *Scientific Reports*, 10(1), 12891, 2020.

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.; PARKHILL, J.; HARRIS, D.; WALKER, A. S.; BOWDEN, R.; MONK, P.; SMITH, E. G.; & PETO, T. E. A. Whole-genome sequencing to delineate Mycobacterium tuberculosis outbreaks: A retrospective observational study. *The Lancet Infectious Diseases*, 13(2), 137–146, 2013.

WALKER, T. M.; KOHL, T. A.; OMAR, S. V.; HEDGE, J.; DEL OJO ELIAS, C.; BRADLEY, P.; IQBAL, Z.; FEUERRIEGEL, S.; NIEHAUS, K. E.; WILSON, D. J.; CLIFTON, D. A.; KAPATAI, G.; IP, C. L. C.; BOWDEN, R.; DROBNIEWSKI, F. A.; ALLIX-BÉGUEC, C.; GAUDIN, C.; PARKHILL, J.; DIEL, R.; … MUNANG, M. Whole-genome sequencing for prediction of Mycobacterium tuberculosis drug susceptibility and resistance: A retrospective cohort study. *The Lancet Infectious Diseases*, 15(10), 1193–1202, 2015.

WALKER, T. M.; LALOR, M. K.; BRODA, A.; ORTEGA, L. S.; MORGAN, M.; PARKER, L.; CHURCHILL, S.; BENNETT, K.; GOLUBCHIK, T.; GIESS, A. P.; DEL OJO ELIAS, C.; JEFFERY, K. J.; BOWLER, I. C. J. W.; LAURENSON, I. F.;

BARRETT, A.; DROBNIEWSKI, F.; MCCARTHY, N. D.; ANDERSON, L. F.; ABUBAKAR, I.; … CONLON, C. P. Assessment of Mycobacterium tuberculosis transmission in Oxfordshire, UK, 2007-12, with whole pathogen genome sequences: An observational study. *The Lancet Respiratory Medicine*, 2(4), 285–292, 2014.

WALTER, K. S.; PEREIRA DOS SANTOS, P. C.; GONÇALVES, T. O.; DA SILVA, B. O.; SANTOS, A. DA S.; LEITE, A. DE C.; DA SILVA, A. M.; MOREIRA, F. M. F.; DE OLIVEIRA, R. D.; LEMOS, E. F.; CUNHA, E.; LIU, Y. E.; KO, A. I.; COLIJN, C.; COHEN, T.; MATHEMA, B.; CRODA, J.; & ANDREWS, J. R. The role of prisons in disseminating tuberculosis in Brazil: A genomic epidemiology study. *The Lancet Regional Health - Americas*, 9, 100186, 2022.

WANG, S.; ZHAO, B.; SONG, Y.; ZHOU, Y.; PANG, Y.; OU, X.; LI, Q.; XIA, H.; ZHAO, Y.; WANG, S.; ZHAO, B.; SONG, Y.; ZHOU, Y.; PANG, Y.; OU, X.; LI, Q.; XIA, H.; & ZHAO, Y. Molecular characterization of the rpoB gene mutations of Mycobacterium tuberculosis isolated from China. *Journal of Tuberculosis Research*, 1(1), 1–8, 2013.

WANG, Z.; XIE, T.; MU, C.; WANG, C.; JU, H.; ZHAO, H.; & SUN, R. Molecular characteristics of ofloxacin mono-resistant Mycobacterium tuberculosis isolates from new and previously treated tuberculosis patients. *Journal of Clinical Laboratory Analysis*, 32(1), 2018.

WHO. *Consolidated Guidelines on Tuberculosis, Module 4: Treatment - Drug-Susceptible Tuberculosis Treatment*. World Health Organization, 2022.

WHO. *Global Tuberculosis Report 2021*. World Health Organization, 2021.

WHO. *Implementing Tuberculosis Diagnostics Policy framework.* World Health Organization, 2015.

XU, W.; DEJESUS, M. A.; RÜCKER, N.; ENGELHART, C. A.; WRIGHT, M. G.; HEALY, C.; LIN, K.; WANG, R.; PARK, S. W.; IOERGER, T. R.; SCHNAPPINGER, D.; & EHRT, S. Chemical genetic interaction profiling reveals determinants of intrinsic antibiotic resistance in Mycobacterium tuberculosis. *Antimicrobial Agents and Chemotherapy*, 61(12), 2017.

ZHANG, J.; CHIODINI, R.; BADR, A.; & ZHANG, G. The impact of next-generation sequencing on genomics. *Journal of Genetics and Genomics*, 38(3), 95–109, 2011.

ZIGNOL, M.; DEAN, A. S.; ALIKHANOVA, N.; ANDRES, S.; CABIBBE, A. M.; CIRILLO, D. M.; DADU, A.; DREYER, A.; DRIESEN, M.; GILPIN, C.; HASAN, R.; HASAN, Z.; HOFFNER, S.; HUSAIN, A.; HUSSAIN, A.; ISMAIL, N.; KAMAL, M.; MANSJÖ, M.; MVUSI, L.; … RAVIGLIONE, M. C. Population-based resistance of Mycobacterium tuberculosis isolates to pyrazinamide and fluoroquinolones: results from a multicountry surveillance project. *The Lancet Infectious Diseases*, 16(10), 1185–1192, 2016.

Curriculum Vitae resumido

1. DADOS PESSOAIS

Nome: Richard Steiner Salvato

Local e Data de Nascimento: Porto Alegre, Rio Grande do Sul, Brasil, 05/12/1994

Endereço Profissional:

Centro Estadual de Vigilância em Saúde / SES-RS

Av. Ipiranga nº 5400. Porto Alegre - RS. CEP: 90.610-000.

Telefone Profissional: +55 51 32884036 **E-mail:** richardsalvato@hotmail.com

2. FORMAÇÃO ACADÊMICA

Doutorado em Biologia Celular e Molecular - Universidade Federal do Rio Grande do Sul (UFRGS) - 2022

Tese: "Epidemiologia molecular e genômica de *Mycobacterium tuberculosis* na Região Sul do Brasil".

Mestrado em Biologia Celular e Molecular Aplicada à Saúde - Universidade Luterana do Brasil (ULBRA) – 2019

Dissertação: "Whole genome sequencing in the prediction of resistance and phylogenetic inferences in Mycobacterium tuberculosis isolated in Rio Grande do Sul, Brazil".

Graduação em enfermagem - Universidade Luterana do Brasil (ULBRA) – 2017

3. ESTÁGIOS

08/2016 –08/2017: Universidade Luterana do Brasil
Bolsista de Iniciação Científica, principais atividades:
Validação do instrumento de qualidade de vida interrai-qol para transtornos mentais e adições (qol-mha); Avaliação do apoio matricial em saúde mental em Porto Alegre, RS/Brasil.

04/2014 – 04/2016: Vigilância Epidemiológica de Alvorada, RS
Estagiário, principais atividades:
Alimentação e manutenção do SINAN; Investigações epidemiológicas de agravos de notificação compulsória; Vacinação; Investigações epidemiológicas de causas de óbitos indefinidas.

10/2013 – 04/2014: Coordenadoria geral de vigilância em saúde (CGVS)/SMS de Porto Alegre
Estagiário, principais atividades:
Alimentação dos bancos de dados e demais rotinas do SISPRÉNATAL e SINASC.

4. CURSOS

XIX Curso Internacional de Epidemiologia Molecular em Doenças Infecciosas, 2019. (Carga horária: 48h). FIOCRUZ, FIOCRUZ-BA, Salvador, Brasil.

25th Bioinformatics Workshop on Virus Evolution and Molecular Epidemiology (VEME), 2021. (Carga horária: 80h). Belo Horizonte, Brasil.

MiSeq System Operational Training provided by Illumina, 2022. (Carga horária: 24h). Porto Alegre, Brasil.

5. DISTINÇÕES

- ✓ Atualmente coordena a Vigilância Genômica na Secretaria Estadual da Saúde do Rio Grande do Sul.
- ✓ Responsável pela implementação e coordenação da vigilância de variantes do SARS-CoV-2 no Rio Grande do Sul, bem como por integrar os Grupos Técnicos GT Saúde e GT Protocolos do Governo do Estado durante a pandemia da COVID-19.
- ✓ Mais de 50 textos/entrevistas na imprensa a respeito da vigilância genômica no Rio Grande do Sul.
- ✓ Integra a Rede Genômica FIOCRUZ.

6. EXPERIÊNCIA PROFISSIONAL OU DIDÁTICA ANTERIOR

06/2020 – Atual: Centro Estadual de Vigilância em Saúde/ Secretaria Estadual da Saúde do Rio Grande do Sul
Cargo: Especialista em Saúde, principais atividades:
Coordenação da Vigilância Genômica no Governo do Estado;
Diagnóstico molecular de doenças causadas por vírus respiratórios e arboviroses;
Sequenciamento de nova geração e análises bioinformáticas;
Projetos de pesquisa relacionados aos aspectos moleculares e epidemiológicos de doenças infecciosas.

07/2017 – Atual: Centro de Desenvolvimento Científico e Tecnológico/CEVS/SES-RS
Cargo: Pesquisador colaborador, principais atividades:
Análise bioinformática dos dados de sequenciamento de nova geração (NGS) de genomas de *Mycobacterium tuberculosis;*
Condução de projetos de pesquisa onde objetivam a caracterização molecular de isolados de *Mycobacterium tuberculosis*;
Análise da situação epidemiológica da tuberculose no estado;

12/2019 - Universidade Federal de Santa Catarina (UFSC), Programa de Pós-graduação em Farmácia.
Disciplina "Tópicos em análise bioinformática de dados do sequenciamento do genoma completo de bactérias", Créditos: 2.

11/2019 – Universidade de Santa Cruz do Sul (UNISC).
Curso "Sequenciamento de Nova Geração (NGS) aplicado à epidemiologia molecular da tuberculose", CH: 15 horas.

10/2019 - Universidade Federal de Minas Gerais (UFMG), Faculdade de Medicina.
Curso "Whole Genome Sequencing (WGS) da Tuberculose", CH: 54 horas.

07/2022 – Rede Genômica FIOCRUZ.
Curso "Curso do ViralFlow", CH: 8 horas.

7. ARTIGOS COMPLETOS PUBLICADOS

17 SALVATO, R. S.; RODRIGUES IKEDA, M. L.; BARCELLOS, R. B.; *et al*. Possible Occupational Infection of Healthcare Workers with Monkeypox Virus, Brazil. Emerging infectious diseases, v. 28, n. 12, 2022.

16 SILVA, T. DE S.; SALVATO, R. S.; GREGIANINI, T. S.; *et al*. Molecular characterization of a new SARS-CoV-2 recombinant cluster XAG identified in Brazil. Frontiers of medicine, v. 0, 2022. Frontiers.

15 ARANTES, I.; GOMES, N. F.; GRÄF, T.; *et al*. Emergence and Spread of the SARS-CoV-2 Variant of Concern Delta across Different Brazilian Regions. Microbiology spectrum, 2022. Microbiol Spectr.

14 ARANTES, I. G.; SALVATO, R. S.; GREGIANINI, T. S.; *et al*. Multiple Introductions of SARS-CoV-2 C.37 Lambda lineage in the Southern Brazilian region. Journal of Travel Medicine, 2021.

13 GRÄF, T.; BELLO, G.; NAVECA, F. G.; *et al*. Phylogenetic-based inference reveals distinct transmission dynamics of SARS-CoV-2 lineages Gamma and P.2 in Brazil. iScience, v. 25, n. 4, p. 104156, 2022.

12 DEZORDI, F. Z.; RESENDE, P. C.; NAVECA, F. G.; *et al*. Unusual SARS-CoV-2 intrahost diversity reveals lineage superinfection. Microbial genomics, v. 8, n. 3, 2022.

11 GRÄF, T.; BELLO, G.; VENAS, T. M. M.; *et al*. Identification of a novel SARS-CoV-2 P.1 sub-lineage in Brazil provides new insights about the mechanisms of emergence of variants of concern. Virus evolution, v. 7, n. 2, p. veab091, 2021.

10 VARELA, A. P. M.; PRICHULA, J.; MAYER, F. Q.; *et al*. SARS-CoV-2 introduction and lineage dynamics across three epidemic peaks in Southern Brazil: massive spread of P.1. Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases, v. 96, p. 105144, 2021.

9 SALVATO, R. S.; REIS, A. J.; SCHIEFELBEIN, S. H.; *et al*. Genomic-based surveillance reveals high ongoing transmission of multi-drug-resistant Mycobacterium tuberculosis in Southern Brazil. International journal of antimicrobial agents, v. 58, n. 4, p. 106401, 2021.

8 SALVATO, R. S.; GREGIANINI, T. S.; CAMPOS, A. A. S.; *et al*. Epidemiological investigation reveals local transmission of SARS-CoV-2 lineage P.1 in Southern Brazil.

7 CONCEIÇÃO, E. C.; SALVATO, R. S.; GOMES, K. M.; *et al*. Molecular epidemiology of Mycobacterium tuberculosis in Brazil before the whole genome sequencing era: a literature review. Memorias do Instituto Oswaldo Cruz, v. 116, p. e200517, 2021.

6 MEDEIROS, T. F.; SCHEFFER, M. C.; VERZA, M.; *et al*. Genomic characterization of variants on mycolic acid metabolism genes in Mycobacterium tuberculosis isolates from Santa Catarina, Southern Brazil. Infection, Genetics and Evolution, 2021.

5 VERZA, M.; SCHEFFER, M. C.; SALVATO, R. S.; *et al*. Genomic epidemiology of Mycobacterium tuberculosis in Santa Catarina, Southern Brazil. Scientific reports, v. 10, n. 1, p. 12891, 2020.

4 ROSSETTI, M. L.; DA SILVA PE, A.; SALVATO, R. S.; *et al*. A highly rifampicin resistant Mycobacterium tuberculosis strain emerging in Southern Brazil. Tuberculosis , v. 125, 2020. Tuberculosis (Edinb).

3 SALVATO, R. S.; COSTA, E. R. D.; REIS, A. J.; *et al*. First insights into circulating XDR and pre-XDR Mycobacterium tuberculosis in Southern Brazil. Infection, genetics and

evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases, v. 78, 2020.

2 SALVATO, R. S.; SCHIEFELBEIN, S.; BARCELLOS, R. B.; *et al*. Molecular characterisation of multidrug-resistant Mycobacterium tuberculosis isolates from a high-burden tuberculosis state in Brazil. Epidemiology and infection, v. 147, p. e216, 2019. Cambridge University Press.

1 DE CAMARGO SILVA, J.; SALVATO, R. S.; DA SILVA, D. M. Cuidados da equipe de enfermagem ao paciente com insuficiência renal crônica durante a sessão de hemodiálise: revisão integrativa. Revista Ampliar, v. 3, n. 3, 2017.

PREPRINTS DE ARTIGOS SUBMETIDOS RECENTEMENTE

GREGIANINI, T. S.; SALVATO, R. S.; BARCELLOS, R. B.; *et al*. Chikungunya virus transmission in the Southernmost state of Brazil was characterized by self-limited cases (2017–2019) and a larger 2021 outbreak. BioRxiv 2022. doi:10.1101/2022.09.30.510389. *Under review* na revista Emerging Infectious Diseases.

SANT'ANNA F.H.; ANDREIS T.F.; SALVATO R.S.; *et al*. Incipient parallel evolution of SARS-CoV-2 Deltacron variant in South Brazil. BioRxiv 2022. doi:10.1101/2022.10.06.511203. *Under review* na revista Emerging Infectious Diseases.

CHEN N.F.G, CHAGUZA C., GAGNE L.; *et al*. Multi-site validation of an amplicon-based sequencing approach for human monkeypox virus. Preprint. medRxiv. 2022. doi:10.1101/2022.10.14.22280783. *Under review* na revista Nature Microbiology.

8. RESUMOS E TRABALHOS APRESENTADOS EM CONGRESSOS

SALVATO, R. S.; SCHIEFELBEIN, S. ; PRAETZEL, B. M. ; ANUSCA, I. S. ; ESTEVES, L. S. ; DALLA COSTA, E. R. ; RIBEIRO, M. O. ; ROSSETTI, M. L. . DIVERSIDADE GENÉTICA DA TUBERCULOSE MULTIRRESISTENTE NO RIO GRANDE DO SUL. In: Congresso da Sociedade Brasileira de Medicina Tropical, 2018, Recife. Anais MEDTROP 2018, 2018.

SALVATO, R. S.; SCHIEFELBEIN, S. ; PRAETZEL, B. M. ; BELLO, G. ; DALLA COSTA, E. R. ; RIBEIRO, M. O. ; UNIS, G. ; DIAS, C. ; ROSSETTI, M. L. . PERFIL CLÍNICO-EPIDEMIOLÓGICO E DISTRIBUIÇÃO ESPACIAL DA TUBERCULOSE RESISTENTE NO RIO GRANDE DO SUL. In: Congresso da Sociedade Brasileira de Medicina Tropical, 2018, Recife. Anais MEDTROP 2018, 2018.

9. ORIENTAÇÕES

Monografia de conclusão de curso de aperfeiçoamento/especialização

1. Nicole Reis. O PERFIL DE RESISTÊNCIA DO MYCOBACTERIUM TUBERCULOSIS NO HOSPITAL NOSSA SENHORA DA CONCEIÇÃO: UM ESTUDO TRANSVERSAL RETROSPECTIVO. 2021. Monografia. (Residência Médica em Infectologia) - Hospital Nossa Senhora da Conceição. Co-orientador: Richard Steiner Salvato.

Trabalho de conclusão de curso de graduação

1. Dandára Costa Fanfa. Análise das estirpes resistentes de Mycobacterium tuberculosis na População Privada de Liberdade do Rio Grande do Sul. 2021. Trabalho de Conclusão de Curso. (Graduação em Farmácia) - Universidade de Santa Cruz do Sul. Co-orientador: Richard Steiner Salvato.

2. Stephanie Steiner Salvato. EPIDEMIOLOGIA DA TUBERCULOSE RESISTENTE NO RIO GRANDE DO SUL. 2021. Trabalho de Conclusão de Curso. (Graduação em Farmácia) - Centro Universitário Metodista. Co-orientador: Richard Steiner Salvato.

3. Bruno Marques Praetzel. Avaliação da relação entre a frequência de mutações de resistência no gene gyrA e a MIC de isolados multirresistentes de Mycobacterium tuberculosis na população gaúcha. 2019. Trabalho de Conclusão de Curso. (Graduação em Biomedicina) - Fundação Universidade Federal de Ciências da Saúde de Porto Alegre. Co-orientador: Richard Steiner Salvato.

Iniciação científica

1. Stephanie Steiner Salvato. Epidemiologia molecular da Tuberculose no Rio Grande do Sul. 2019. Iniciação Científica. (Graduação em Farmácia) - CENTRO DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO DO RS. Orientador: Richard Steiner Salvato.

2. Larisa Vitória da Silva. Epidemiologia molecular da tuberculose no Rio Grande do Sul. 2019. Iniciação Científica - CENTRO DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO DO RS. Orientador: Richard Steiner Salvato.

3. Bruno Marques Praetzel. Avaliação da relação entre a frequência de mutações de resistência no gene gyrA e a MIC de isolados multirresistentes de Mycobacterium tuberculosis na população gaúcha. 2018. Iniciação Científica - CENTRO DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO DO RS. Orientador: Richard Steiner Salvato.

10. REVISOR PERIÓDICOS

2019 - Atual

Periódico: INFECTION GENETICS AND EVOLUTION

2019 - Atual

Periódico: Journal of Global Antimicrobial Resistance

2019 - Atual

Periódico: Revista do Instituto de Medicina Tropical de São Paulo

2020 - Atual

Periódico: Epidemiology and Infection

2019 - Atual

Periódico: Annals of Clinical Microbiology and Antimicrobials