UNIVERSIDADE DO ESTADO DO PARÁ
INSTITUTO EVANDRO CHAGAS
CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA PARASITÁRIA NA AMAZÔNIA

DAVI JOSUÉ MARCON

EPIDEMIOLOGIA GENÔMICA DA TUBERCULOSE DROGA-RESISTENTE NO ESTADO DO PARÁ APÓS INTRODUÇÃO DA BEDAQUILINA

Linha de pesquisa: Morfologia e genética de agentes infecciosos, parasitos e vetores de endemias na Amazônia.

BELÉM – PA
2025

DAVI JOSUÉ MARCON

EPIDEMIOLOGIA GENÔMICA DA TUBERCULOSE DROGA-RESISTENTE NO ESTADO DO PARÁ APÓS INTRODUÇÃO DA BEDAQUILINA

Dissertação apresentada ao Programa de Pós-Graduação em Biologia Parasitária na da Amazônia da Universidade do Estado do Pará. como requisito para a obtenção do grau de Mestre em Biologia Parasitária da Amazônia.

Linha de pesquisa: Morfologia e genética de agentes infecciosos, parasitos e vetores de endemias na Amazônia

Orientadora: Dra. Karla Valéria Batista Lima
Co-orientador: Dr. Carlos Augusto Abreu Albério.

Banca Avaliadora:

_____

Dr. Carlos Alberto Marques de Carvalho

_____

Dra. Sintia Silva de Almeida

_____

Dr. Fabiano Cordeiro Moreira

Orientadora:_____

Profa. Dra. Karla Valéria Batista Lima

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema de Bibliotecas da UEPA / SIBIUEPA

Marcon, Davi Josué

Epidemiologia genômica da tuberculose droga-resistente no Estado do Pará após introdução da bedaquilina / Davi Josué Marcon. – Belém: UEPA, 2025.

160f.: il.

Orientadora: Prof. Dra. Karla Valéria Batista Lima

Coorientador: Dr. Dr. Carlos Augusto Abreu Albério

Dissertação (Mestrado) – Universidade do Estado do Pará, Centro de Ciências Biológicas e da Saúde, Programa de Pós-graduação em Biologia parasitária da Amazônia, Belém, 2025.

1. Tuberculose resistente. 2. Bedaquilina. 3. Linezolida. I. Universidade do Estado do Pará. II. Título.

CDD 22.ed. 614.542

Elaborada por Josicléia Garcia Vieira - CRB-2 /562

DEDICATÓRIA

A todos que acreditam na erradicação da Tuberculose.

AGRADECIMENTOS

A Deus pela saúde física e espiritual e pela oportunidade de trabalhar com o que gosto, acompanhado de amigos e colegas, sem precisar me preocupar com qualquer outra coisa que não fosse trabalhar e desenvolver esse projeto.

A minha Orientadora Dra. Karla Lima, por me integrar ao seu grupo (que descobri ser uma grande família), me permitindo colaborar e opinar com todos os projetos de seu grupo, pela confiança, investimento e apoio no desenvolvimento de um projeto tão complexo. Agradeço por ter me desafiado neste projeto, mas também por ter confiado em minha capacidade de realizá-lo, me fornecendo ferramentas para crescer.

Ao meu Coorientador, Dr. Carlos Albério, pelas excelentes discussões no ambulatório do Barros Barreto, por sempre fornecer comentários pertinentes de maneira simples e demonstrar que antes de tratamentos e exames, todas as pessoas em tratamento são humanos, com nome e história para serem avaliados.

Aos meus mentores Emilyn e Abhinav Sharma, por me pegarem na mão nas vezes que precisei, pelas broncas em algumas de nossas reuniões, e por oportunizar o meu desenvolvimento não apenas como acadêmico, mas como homem e como profissional. Obrigado pela companhia nos dias que tomamos café enquanto discutimos as intermináveis anotações no ITOL e por todos os ricos conselhos que moldaram minha carreira até esse exato momento, sempre me ensinando pelo exemplo.

Aos meus pais Fábio e Silvane, por me permitirem trabalhar sem me preocupar com bolsas ou financiamento, me apoiando no desenvolvimento do projeto e ouvindo todos os resultados preliminares deste trabalho, me permitindo ter tranquilidade e conforto para trabalhar nele. Aos meus irmãos, Israel, Giulia e Laura, pela companhia durante os dias que pude aproveitar em casa, pelos passeios dos Irmões, e as conversas durante o fim de semana.

Aos professores e pesquisadores: Ana Judith, Agenor Valadares, Carlos Marques, Cyntia Souza, Danielle Brasiliense, Fabiano Moreira, Luciana Xavier, Rafael Baraúna, e Silvia Silva. Pelos ensinamentos transmitidos nas disciplinas que participei, e as excelentes conversas em corredores sobre este e outros trabalhos, que foram essenciais para minha formação enquanto aluno e na realização de todos os projetos que me envolvi após minha graduação, inclusive, coordenados por eles.

A equipe do Hospital Universitário João de Barros Barreto e a Equipe do Laboratório Central do Estado do Pará pela disponibilização das amostras e resultados laboratoriais para este projeto. Em especial as “tuberculetes” do LACEN: Luna, Rafaella, Roses e Simone, que

sempre me receberam com muito carinho na sala técnica da tuberculose (até cantando parabéns em meu aniversário), sempre se dedicaram na busca de amostras, na organização de cepas e até mesmo, me explicando muitas das metodologias de cultivo e testagem de sensibilidade que são descritas neste trabalho.

As pesquisadoras do projeto REVIGET: Cristina, Cristiane e Theolis, com quem compartilhei conversas, almoços e discussões sobre o projeto e que oportunizaram minha participação em outros projetos.

Aos colegas de Trabalho em Bioinformática no Instituto Evandro Chagas: Sandro Patroca, Luciano Franco, Amália Lobato e Sintia Almeida. Com os quais discuti diversos problemas enfrentados e encontrei solução para a maioria destes, com a sugestão de metodologias e discussões sobre quais os melhores caminhos para esse projeto.

A Seção de Tecnologia de Informação do Instituto Evandro Chagas, em especial Reginaldo Júnior e Roberto Simões. Por terem respondido aos inúmeros chamados que abri durante meu projeto, permitindo que eu utilizasse o HPC do Instituto Evandro Chagas da melhor maneira possível, mesmo quando minhas solicitações pareciam fora do que era planejado. Agradeço pelas horas de estudo que foram necessárias para a implementação da maioria dos softwares utilizados neste projeto.

As técnicas do laboratório de Tuberculose do Instituto Evandro Chagas: Michelle Brito e Jeanne Cabral. Que muitas vezes me auxiliaram no cadastro das amostras no sistema do instituto e na manipulação das amostras obtidas no LACEN.

Ao técnico Alex Brito, que não apenas auxiliou minha chegada no Laboratório de Tuberculose, me fornecendo materiais e me ensinando que eu precisava para realizar meu projeto. Mas me acompanhou em almoços, lanches e até na Bahia, sempre compartilhando suas experiências e conselhos valorosos para minha formação, além das caronas e risadas.

Ao Dr. Ricardo, pela acessibilidade e disponibilidade em elaborar os mapas que utilizamos para discussão destes resultados, por rapidamente nos responder e aceitar todas as inúmeras alterações que solicitamos.

As “meninas da Karla”: Emmily, Emilly, Layana e Isabel, que me acompanharam durante todo este projeto e nas atividades do projeto REVIGET desde o começo, pela colaboração, almoços e conversas durante minha integração no grupo de pesquisa.

Aos meus amigos e colegas que me acompanharam durante o Mestrado: Ana, Ana Paula, Brenda, Bruna, Carlos, Livia, Giovanna, Iasmin, Jady, Mateus e Mayza pela companhia e colaboração durante a maioria dos trabalhos que desenvolvemos durante nossas aulas.

Aos meus amigos da graduação me acompanharam, discutindo e propondo idéias e sempre me dando espaço para participar de seus próprios projetos: Bia, Gustavo, Giovane, Joaquim, Naiana, Thiago, Valéria. Obrigado pelos passeios mensais, as conversas e todo o tempo que aproveitamos.

Aos meus amigos: Cayo, Matheus, Rafaelly e Tiago, com quem muitas vezes saí para conversar ou passar um tempo durante a realização do projeto.

A Raphaela Costa pelos cafés, conversas, passeios, discussões e contribuições na reta final deste projeto, seja sobre o projeto, ou sobre outros assuntos.

Aos desenvolvedores das ferramentas de código aberto utilizados no desenvolvimento deste trabalho e aos responsáveis por *shadow libraries* como libgen, *anna's archive* e scihub, de onde obtive a maioria dos livros e artigos citados neste texto.

Aos alunos de graduação que pude colaborar na formação durante minha própria formação como mestre, discutindo metodologias e propondo idéias: Danna, José e Beatriz.

Ao Instituto Evandro Chagas, a Universidade do Estado do Pará e o Programa de Pós Graduação em Biologia Parasitária na Amazônia pela estrutura e oportunidade deste projeto.

Ao CNPq, Capes pelo financiamento do projeto e a concessão de minha Bolsa

A REVIGET pelas oportunidades, experiências em treinamento, financiamentos, discussões e pela oportunidade de conhecer a Bahia.

EPÍGRAFE

“Existe uma única luz na ciência, e iluminar em qualquer lugar, é iluminar todo lugar.”

Isaac Asimov

RESUMO

Recentemente a bedaquilina foi introduzida para o tratamento da tuberculose droga resistente (TB-DR), sendo a primeira inclusão após 40 anos sem novos fármacos. Apesar de promissor, este medicamento já possui relatos de resistência em outros países, e a farmacovigilância torna-se necessária para o uso racional do fármaco e a compreensão dos fenômenos de aquisição de resistência. Este Trabalho tem como objetivo implementar fluxo de análise genômico acessível para usuários com pouca experiência para a análise de genomas do Complexo *Mycobacterium Tuberculosis* CMTB e propor um fluxo de sequenciamento de genoma completo a partir da cultura utilizada na rotina de diagnóstico e a análise de seus resultados. Este trabalho é considerado um projeto piloto para integrar o sequenciamento genômico à rotina laboratorial da tuberculose no Pará, um importante passo na implementação da rede de vigilância norte-nordeste da tuberculose (REVIGET). Isolados utilizados na rotina de diagnóstico do LACEN-PA e identificadas como resistentes a rifampicina e/ou isoniazida tiveram o DNA extraído utilizando o kit Instagene Matrix e sequenciadas na plataforma Illumina Nextseq550. As análise foram realizadas com o pipeline MAGMA. Resultando na detecção dos primeiros casos de Tuberculose Extensivamente Droga Resistente (TB-XDR) no Pará, com 2 casos de resistência à bedaquilina e 1 caso de resistência à linezolida, além de revelar a rede de transmissão para os casos em questão. Ao integrar dados clínicos, epidemiológicos e genéticos este trabalho ampliou a compreensão sobre a dinâmica da tuberculose droga resistente no Pará, trazendo metodologias modernas aplicadas em outros países para a região amazônica, com a expectativa de auxiliar na tomada de decisões para o enfrentamento da TB no estado e lançar base para o fortalecimento da rede de vigilância norte nordeste da tuberculose no pará.

Palavras-chave: Sequenciamento completo do genoma (WGS); Tuberculose Extensivamente Resistente a Medicamentos, Bedaquilina, Linezolida.

ABSTRACT

Recently, bedaquiline was introduced for the treatment of drug-resistant tuberculosis (DR-TB), marking the first inclusion of a new drug after 40 years without novel anti-TB medications. Despite its promising inclusion, resistance to this drug has already been reported in other countries, highlighting the need for pharmacovigilance to ensure its rational use and to understand the mechanisms of resistance acquisition. This study aims to implement an accessible analysis workflow for users without prior experience to analyze MTBC genomes and propose a workflow for WGS using culture from the diagnostics routine and to analyze its results. This study is a pilot project to integrate genomic sequencing into the routine tuberculosis laboratory workflow in Pará, and represents a significant step in the implementation of the North-Northeast Tuberculosis Surveillance Network (REVIGET). Strains used in the routine diagnostics of LACEN-PA, identified as resistant to rifampicin and/or isoniazid, were extracted using the Instagene Matrix kit, sequenced on the Illumina NextSeq550 platform, and analyzed using the MAGMA pipeline. This resulted in the detection of the first XDR-TB cases in Pará, with two cases of bedaquiline resistance and one case of linezolid resistance, as well as the identification of the transmission network associated with these cases. By integrating clinical, epidemiological, and genetic data, this study expanded the understanding of drug-resistant tuberculosis dynamics in Pará, bringing modern methodologies applied in other countries to the Amazon region. The findings are expected to support decision-making in tackling TB in the state and lay the foundation for strengthening the North-Northeast Tuberculosis Surveillance Network in Pará.

Keywords: Whole Genome Sequecing (WGS); Extensively Drug Resistant Tuberculosis, Bedaquiline, Linezolid.

LISTA DE ABREVIATURAS E SIGLAS

BCG	Bacilo Calmette-Guérin
CIM	Concentração Inibitória Mínima
CMTB	Complexo *Mycobacterium tuberculosis*
GEMIBRA	Genomas de Micobactérias do Brasil
GWAS	Genome-Wide Association Study
HIV	*Human Imunodeficiency Virus*
LACEN	Laboratório Central do Estado
LPA	*Line Probe Assay*
MAGMA	*Maximum Accessible Genome for Mtb Analysis*
MGIT	*Mycobacteria Growth Indicator Tube*
MIRU-VNTR	*Micobacterial Interspaced Repetitive Unitis - Variable Number Tandem Repeats*
MNT	Micobactérias Não-Tuberculosas
MTB	Mycobacterium tuberculosis
OMS	Organização Mundial da Saúde
REVIGET	Rede de Vigilância Genomica da Tuberculose
SCG	Sequenciamento Completo de Genoma
SNP	Single Nucleotide Polymorfisms
SUS	Sistema Único de Saúde
TB	Tuberculose
TB-DR	Tuberculose - Droga-Resistente
TB-MDR	Tuberculose - Multi-Droga-Resistente
TB-RR	Tuberculose - Rifampicina-Resistente
TB-XDR	Tuberculose - Extensivamente-Resistente
TRM	Teste Rápido Molecular
TSA	Teste de Susceptibilidade Antimicrobiano

UBS	Unidade Básica de Saúde
UMS	Unidade Municipal de Saúde
URES	Unidade de Referência Especializada em Saúde
WGS	*Whole Genome Sequencing*

SUMÁRIO

1. INTRODUÇÃO.....6

2. JUSTIFICATIVA.....7

3. REVISÃO DA LITERATURA.....9

3.1 ASPECTOS GERAIS DA TUBERCULOSE.....9

3.2 CARACTERIZAÇÃO DO AGENTE ETIOLÓGICO.....9

3.3 ASPECTOS EPIDEMIOLÓGICOS DA TUBERCULOSE.....12

3.4 O DIAGNÓSTICO DA TUBERCULOSE.....12

3.5 A ESTRUTURA OPERACIONAL DO DIAGNÓSTICO NO ESTADO DO PARÁ.....13

3.6 A ROTINA DE DETECÇÃO DE RESISTÊNCIA.....15

3.7 A RESISTÊNCIA ÀS DROGAS ANTITUBERCULOSE.....18

3.8 ESQUEMAS DE TRATAMENTO DA TUBERCULOSE E DA TUBERCULOSE DROGA-RESISTENTE.....18

3.9 BEDAQUILINA E SUA RESISTÊNCIA.....20

3.10 VIGILÂNCIA GENÔMICA DA TUBERCULOSE.....21

4. OBJETIVOS.....23

4.1 OBJETIVO GERAL.....23

4.2 OBJETIVOS ESPECÍFICOS.....23

5 PRODUÇÃO TÉCNICA E CIENTÍFICA.....24

5.1 ARTIGOS.....24

5.2. PRODUÇÃO TÉCNICA.....86

5.3. APRESENTAÇÃO DE TRABALHOS EM CONGRESSO.....127

5.4 CURSOS MINISTRADOS.....146

6 CONCLUSÕES.....147

7. FONTES DE FINANCIAMENTO.....148

REFERÊNCIAS BIBLIOGRÁFICAS.....149

ANEXO A: CERTIFICADOS DAS ATIVIDADES MENCIONADAS NO TEXTO.....155

1. INTRODUÇÃO

A tuberculose (TB) é uma doença infectocontagiosa e uma das principais causas de morte em todo o mundo. Apesar de ser um problema de saúde pública milenar, o combate a TB ainda é um desafio a ser superado devido principalmente à emergência de estirpes droga-resistentes e sua disseminação na população (Bagcchi, 2023).

Após 40 anos desde a última introdução de um medicamento na terapia da TB, foram recentemente propostos quatro novos medicamentos: bedaquilina, delamanida, pretomanida e linezolida. O desenvolvimento destes novos medicamentos, bem como o uso de fármacos remanejados para tratar a TB, levaram à atualização das diretrizes para formulação de esquemas de tratamento pela Organização Mundial de Saúde (OMS) (OMS, 2022c).

No Brasil, bedaquilina foi aprovada e incorporada ao Sistema Único de Saúde (SUS) em novembro de 2021 como principal opção terapêutica para pacientes com TB resistente à rifampicina (RR), multirresistente (MDR) e extensivamente resistente a medicamentos (XDR) (Brasil, 2020, 2021).

Todavia, apesar de sua introdução dessa recente inovação em outros países, já tem sido reportado o desenvolvimento de resistência a tais medicamentos de modo, assim, a sua inserção em tratamentos clínicos deve ser acompanhada de testes de susceptibilidade antimicrobianos (TSA), uma vez que o conhecimento a respeito dos mecanismos de resistência permanece escasso (Nguyen *et al.*, 2018, 2020; Satapathy *et al.*, 2024).

Quanto a isso, uma das dificuldades a serem superadas é a ausência de testes padronizados e diretrizes claras sobre os fenômenos de aquisição de resistência para tais medicamentos (Mallick *et al.*, 2022). Alternativamente, à detecção fenotípica por TSA, tem sido proposto o uso do sequenciamento completo do genoma (SCG), uma vez que esta ferramenta tem sido popularizada e com custos diminuídos, além da proposta recente um painel de mutações pela OMS e um catálogo sugerindo interpretação para tais mutações (OMS, 2022a).

No entanto, ainda existem limitações quanto à seleção de variantes relacionadas à resistência à bedaquilina. Tendo sido classificadas como incertas ou não significantes, em decorrência da ausência de testes fenotípicos de alta qualidade que permitam a associação entre o SCG e a concentração inibitória mínima (CIM) dos isolados resistentes e a

necessidade um maior número amostral de cepas resistentes (Nimmo *et al.*, 2024; OMS, 2023).

Além desse cenário inovador e desafiador na área terapêutica, o contexto epidemiológico da TB no mundo, no Brasil e no estado do Pará são preocupantes. O Brasil está listado entre os 30 países com maiores cargas de TB no mundo e o Pará é foi o 6º estado com maior incidência da TB no ano de 2022 (Bagcchi, 2023; Brasil, 2024).

Considerando-se que a introdução da bedaquilina ocorreu em novembro de 2021 no Brasil, e a duração do esquema atual é de 1 ano e 6 meses, os primeiros pacientes já encerram seus tratamentos. Com base no atual cenário epidemiológico e sociodemográfico brasileiro quanto à TB, há grandes chances da ocorrência de infecção mista e/ou re-infecção, e emergência de resistência antimicrobiana aos novos medicamentos, seja por falha terapêutica ou abandono do tratamento. Diante disso, o presente estudo objetiva realizar a vigilância da TB-DR no contexto da incorporação do novo fármaco por meio da epidemiologia genômica no estado do Pará.

2. JUSTIFICATIVA

Tecnologias de vigilância genômica foram amplamente utilizadas para a durante a pandemia de Covid-19 para a busca ativa de mutações, investigações filogenéticas e rastreamento de casos com impacto imediato de políticas públicas a partir dos resultados encontrados (Robishaw *et al.*, 2021). Da mesma forma, tal abordagem pode ser aplicada à investigação para controle da TB-DR.

Com base no protocolo de pesquisa do estudo clínico a partir de medicina personalizada em andamento na África do Sul, o qual utiliza o Sequenciamento Completo de Genoma (SCG) para direcionamento do tratamento da TB-RR, sendo um importante exemplo para aplicação no contexto do Brasil, que também possui alta carga de TB. Dada a natureza pragmática, o estudo auxiliará os formuladores de políticas na tomada de decisões sobre a integração do SCG na rotina de TB-RR (Van Rie *et al.*, 2022).

Como resultado preliminar do estudo desenvolvido na África do sul, relatos de casos de TB-XDR com resistência à bedaquilina têm sido observados e o SCG tem sido importante ferramenta para apoio no resgate adequado e manejo terapêutico. Ressalta-se que a África do Sul foi o primeiro país a incluir a bedaquilina (em 2018) no sistema de saúde para tratamento da TB-MDR, todavia, em 2022 foram observados casos de recidiva ou abandono do

tratamento em indivíduos que utilizaram previamente a bedaquilina. Embora o SCG seja um apoio no diagnóstico, a falta de vigilância genômica inicial tem dificultado a compreensão do cenário epidemiológico atual de transmissão da resistência à bedaquilina, quatro anos após a sua introdução (Conceição *et al.*, 2022).

Esse exemplo revela a importância da implementação inicial da vigilância genômica aliada ao acompanhamento clínico como ferramenta para monitorar a emergência da resistência à bedaquilina. Além disso, a utilização dos dados individuais dos participantes, compartilhados por programas nacionais e grupos de pesquisa, tem sido fundamental para apoiar a geração de conhecimento e diretrizes baseadas em evidências (OMS, 2018b).

Apesar do SGC ser uma realidade como apoio diagnóstico, a interpretação de seus resultados e sua associação com à clínica e a tomada de decisão ainda deve ser explorada. Um recente estudo qualitativo que incluiu médicos do Brasil com experiência no manejo da bedaquilina demonstrou que a percepção e interpretação dos valores de probabilidade de resistência à bedaquilina variaram amplamente entre os médicos (Tu *et al.*, 2022).

Neste contexto, as hipóteses deste estudo quanto à implementação inicial da vigilância genômica da TB-DR integrada ao fluxo existente de diagnóstico laboratorial no Pará são:

1. O SCG é uma ferramenta importante para auxílio na tomada de decisão clínica local para o manejo de casos complexos de resistência antimicrobiana;
2. Na ausência de diretrizes para o TSA, o SCG é um método alternativo para o monitoramento da emergência de resistência à bedaquilina;
3. Há transmissão de estirpes TB-DR local;
4. Há grupos de risco para a transmissão de estirpes TB-DR local;
5. Existem mutações associadas a resistência à bedaquilina que ainda não foram propostas.

Assim, as predições são:

1. Auxiliar na tomada de decisão clínica-terapêutica local;
2. Viabilizar o monitoramento da emergência da resistência antimicrobiana - principalmente contra a bedaquilina;
3. Auxiliar o monitoramento epidemiológico da transmissão de estirpes TB-DR e da emergência de novos genótipos;
4. Contribuir para a detecção de grupos/locais de risco para aquisição ou desenvolvimento de TB-DR
5. Ampliar o conhecimento sobre os mecanismos de resistência antimicrobiana a nível genético dos novos medicamentos com impacto local e global.

3. REVISÃO DA LITERATURA

3.1 ASPECTOS GERAIS DA TUBERCULOSE

A TB é uma doença curável se o tratamento for administrado adequadamente. A maioria das pessoas expostas à TB não desenvolvem os sintomas, pois a bactéria pode viver sob a forma inativa dentro do organismo do indivíduo (forma latente) (OMS, 2022c).

Entretanto, variações no sistema imunológico do hospedeiro, como enfraquecimento causado por desnutrição, ou infecção pelo Vírus da Imunodeficiência Humana (HIV) por exemplo, a bactéria pode então sair do estado latente e passar para a forma de TB ativa e contagiosa. Demonstrando os seguintes sintomas clássicos: tosse persistente e prolongada (por mais de duas semanas), que pode apresentar-se ou não com sangue ou escarro; febre; perda de peso; fadiga, sudorese noturna; e dor no peito (OMS, 2022c).

A tuberculose pode se apresentar na forma pulmonar, com apresentação ou não de cavitações pulmonares. As manifestações externas ao pulmão como pleural, meningítica, óssea são classificadas como tuberculose extrapulmonar (OMS, 2022c).

3.2 CARACTERIZAÇÃO DO AGENTE ETIOLÓGICO

As Micobactérias (gênero: *Mycobacterium*) são bactérias gram positivas da classe dos actinomicetos, caracterizadas por um alto conteúdo guanina-citosina em seu genoma (>65%) e parede celular de rica em lipídeos e ácido micólico, o qual deu origem ao nome do gênero. São bactérias de crescimento lento (superior a 7 dias) de ampla dispersão no ambiente, possuindo representantes não patogênicos, patógenos obrigatórios ou facultativos, destacam-se por serem bacilos com coloração álcool-ácido resistente (Tortora; Case; Funke, 2016).

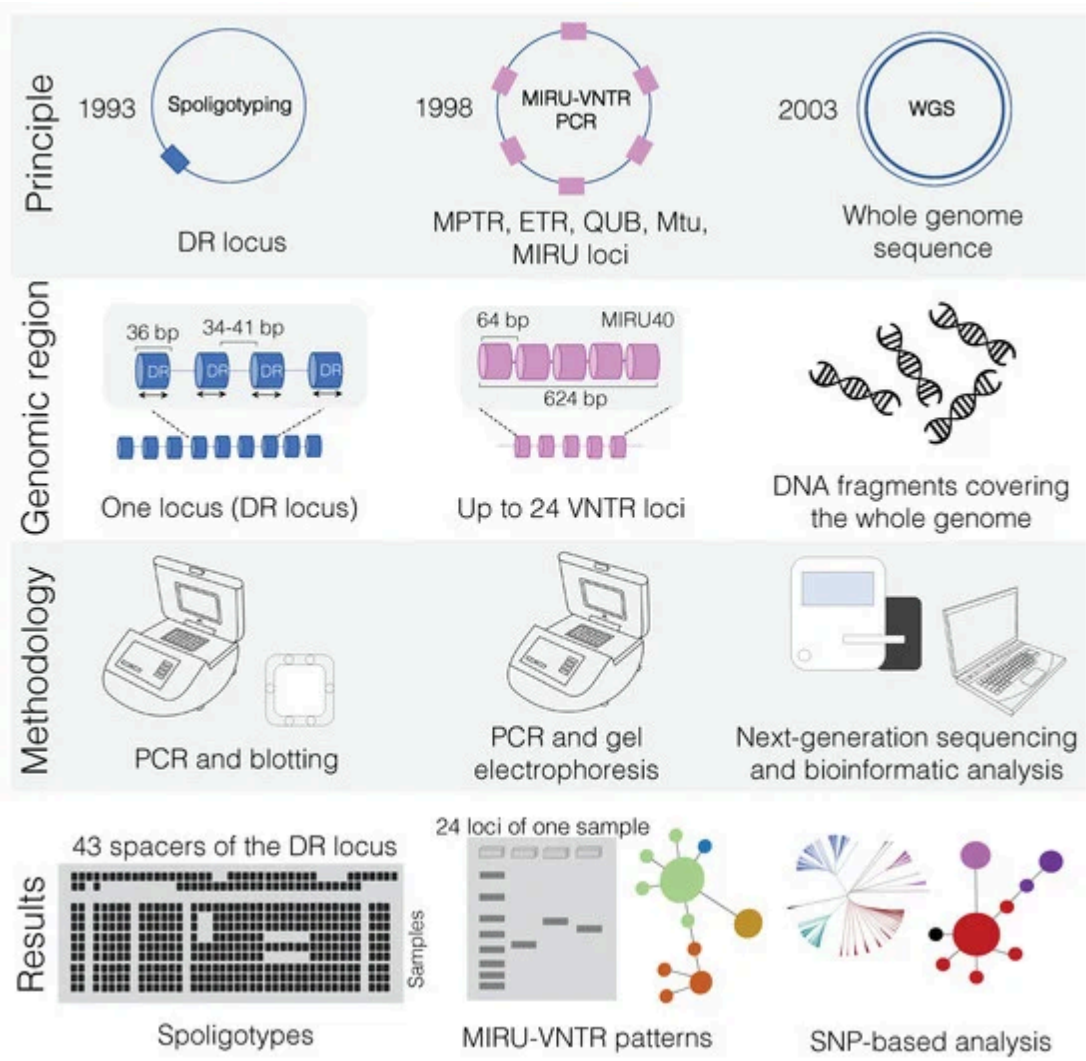
O principal agente etiológico da TB, também conhecido como “bacilo de Koch” (BK), é o *Mycobacterium tuberculosis* var. *tuberculosis* e faz parte do complexo *Mycobacterium tuberculosis* (CMTB). Embora taxonomicamente diferentes, os membros do CMTB pertencem à mesma espécie por possuírem cerca de 99% de similaridade genômica, eles são divididos quanto a sua adaptação aos hospedeiros . Dentre esses, os membros adaptados a humanos são: *M. tuberculosis* var. *tuberculosis, M. tuberculosis* var. *BCG* e *M. tuberculosis*

var. *africanum* e são categorizados em dez linhagens filogenéticas (L1 – L10) (Riojas *et al.*, 2018; Ngabonziza *et al.*, 2020; Guyeux *et al.*, 2024).

É importante diferenciar a tuberculose pulmonar, causada por *M. tuberculosis* de infecções pulmonares causadas por outras micobacterioses não tuberculosas (MNTs) causadas majoritariamente por organismos dos complexos *M. abscessus-chelonae* e *M. avium*. Ambas possuem apresentações radiológicas e sintomatológicas semelhantes, mas os agentes etiológicos e seu padrão de resistência comumente se diferem, sendo necessário a adequação do tratamento para cada caso. Um fenômeno importante, é a ocorrência de Infecções mistas, as quais são caracterizadas pela ocorrência de bactérias do complexo CMTB e MNT simultaneamente. Esse fenômeno apesar de relatado em diversas instâncias, é complexo e deve ser avaliado em cada um dos casos, sendo necessária a investigação através de diversas metodologias para sua confirmação (Yılmaz; Uçar; Sağlam, 2017; Mertaniasih *et al.*, 2017; Kim *et al.*, 2024).

Antes da descrição completa do genoma de MTB por Brosch e colaboradores (1998), no qual a cepa H37Rv foi completamente sequenciada revelando um genoma de aproximadamente 4.4Mb, diversas técnicas foram utilizadas para a diferenciação de membros do complexo MTB e da diferenciação de clones, como exemplificado na figura 1.

Figura 1. Esquemas das técnicas amplamente utilizadas para estudos de tipagem de Mycobacterium tuberculosis: Spoligotyping, MIRU-VNTR, WGS.



Fonte: (Guimaraes; Zimpel, 2020)

As técnicas mais utilizadas foram Spoligotyping e MIRU-VNTR. A primeira se baseia na hibridização de sondas com os locus *DR* (*Differentiation Region* - Região de Diferenciação) do genoma de MTB, e identifica o perfil de presença ou ausência de espaçadores. A segunda se baseia na identificação do número de repetições de cada loci *VNTR* (*Variable Number Tandem Repeat* - Repetição em Tandem de Número Variável) para a identificação da quantidade de unidades de repetições da região variável em tandem. Posteriormente, o uso de técnicas baseadas no sequenciamento genômico e na busca por SNPs permitiram identificar perfis clonais de maneira mais confiável e robusta, além de também permitir a realização do perfil de spoligotyping simultaneamente a outras ferramentas (Guimaraes; Zimpel, 2020).

3.3 ASPECTOS EPIDEMIOLÓGICOS DA TUBERCULOSE

Estima-se que 10,6 milhões de pessoas adoeceram com TB em 2021 no mundo, um aumento de 4,5% em relação aos 10,1 milhões em 2020. A taxa de incidência TB (novos casos por 100.000 habitantes por ano) aumentou 3,6% entre 2020 e 2021, revertendo declínios de cerca de 2% ao ano durante a maior parte das duas últimas décadas. A redução líquida de 2015 a 2021 foi de 10%, apenas meio caminho do primeiro marco do plano estratégico Fim da TB (End TB Strategy) até 2035 (Bagcchi, 2023).

A OMS sugere que em 2020 houve uma diminuição no diagnóstico de TB por conta da sobrecarga dos sistemas de saúde ao redor do mundo por causa da pandemia de COVID-19, com uma parcial recuperação no ano de 2021. O aumento da incidência da TB pode estar relacionado a diminuição no número de diagnósticos, uma vez que muitos casos não foram identificados adequadamente, e portanto, não foram tratados (Bagcchi, 2023).

No Brasil, a incidência de TB foi de 36,3 no ano de 2022, com acréscimo ao comparando-se com os anos de 2021 (34,9), 2020 (33,3), porém similarmente ao relatado pela OMS, o Ministério da Saúde também acredita que esses valores não refletem a totalidade da realidade devido à COVID-19. Os diagnósticos baseados em teste rápido molecular enfrentaram uma recuperação gradual após a pandemia e observa-se o aumento da quantidade dos testes realizados e consequentemente aumento no número total de casos (Garcez *et al.*, 2022).

O Norte do país é a região com maior incidência de TB no Brasil (52,7), que, semelhantemente ao estado do Pará (49,4), possui indicador de incidência maior que a média nacional (36,3). No ano de 2020 foram identificados 2.061 novos casos de TB na região metropolitana de Belém-Pará, destes, sendo 336 TB-DR, apesar da baixa testagem laboratorial. Nesta região, a incidência da TB foi de 94,1; valor aproximadamente três vezes maior que a média nacional para o mesmo ano (32,6). Além disso, um estudo epidemiológico recente em Ananindeua-PA (região metropolitana de Belém) demonstrou a situação de TB super-endêmica, sugerindo a necessidade do reforço na vigilância da TB neste local(SESPA, 2021; Brasil, M. da S., 2022; Garcez *et al.*, 2022).

3.4 O DIAGNÓSTICO DA TUBERCULOSE

A identificação de TB ativa pode ser feita em todas as esferas de atenção da saúde pública, sendo recomendada em todas as situações de tosse persistente, produtiva ou não e mediante a avaliação dos sintomas clínicos por um médico. Após a suspeita clínica, a

metodologia mais utilizada para diagnóstico é o exame microscópico direto (baciloscopia direta, também chamado de microscopia do esfregaço), nos locais onde há disponibilidade, o Teste Rápido Molecular (TRM) deve ser utilizado em razão da velocidade do resultado, no caso de positividade em ambos os casos deve ser solicitada cultura para realização de teste de susceptibilidade aos antimicrobianos de primeira linha (Brasil, 2019, 2022).

No Brasil o Teste Rápido Molecular (TRM) é realizado pelo teste Xpert Ultra MTB/RIF, o qual é capaz de detectar simultaneamente a TB e sua resistência, ou não, ao principal medicamento do esquema básico, a rifampicina. Todavia, este teste não é aplicável para todos os casos suspeitos de tuberculose, como por exemplo casos extrapulmonares (Brasil, 2022). Independentemente do uso de baciloscopia ou do TRM, o exame de cultura deve ser realizado. Em todos os resultados positivos, deve ser solicitado o TSA (Brasil, 2019).

3.5 A ESTRUTURA OPERACIONAL DO DIAGNÓSTICO NO ESTADO DO PARÁ

O Laboratório Central do Estado do Pará (LACEN-PA) é o laboratório de referência estadual, responsável pela realização dos testes de sensibilidade para drogas de primeira linha, distribuição dos *kits* de testes rápidos moleculares, testagem da tuberculose latente por meio do ensaio IGRA e envio de cepas para a referência nacional para realização de TS de segunda linha.

Recentemente o LACEN-PA realizou treinamento para incorporação do LPA de primeira e segunda linha, sendo esperada a oferta do serviço com a distribuição do material pelo Ministério da Saúde.

A rede de teste rápido molecular no Pará possui 8 máquinas geneXpert e é atualmente formada pelos municípios de Belém, Ananindeua, Marabá, Santarém, Altamira e Santa Izabel. A distribuição das máquinas pode ser observada na tabela 1.

Tabela 1 - Distribuição das máquinas de teste rápido molecular pelo estado do Pará

Município	**Localização**
Belém	UMS Guamá
Belém	HUJBB - Hospital Universitário João de Barros Barreto

Belém	UMS Marambaia
Ananindeua	Sem informação
Marabá	Laboratório Central de Marabá - LACEM
Santarém	URES – Santarém
Altamira	CTA Altamira
Santa Izabel	Unidade de Saúde do complexo penitenciário de Santa Izabel

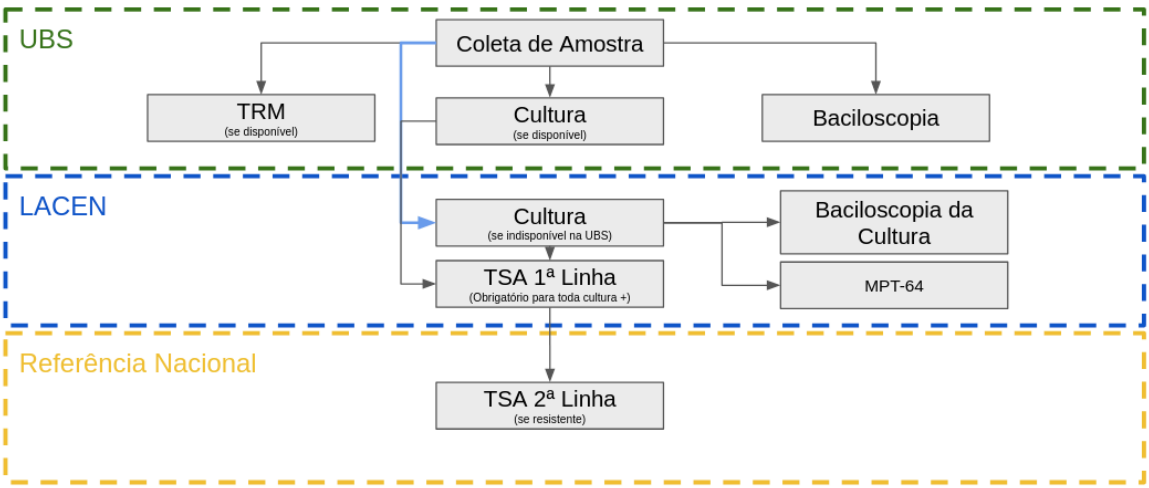
Fonte: Elaborado pelo autor

Uma característica particular do LACEN-PA é sua responsabilidade na rede de descentralização de exames de cultura utilizando a metodologia Ogawa-Kudoh, organizada pelo LACEN-PA para distribuição do material e apoio técnico na realização do diagnóstico base para tuberculose. No caso de positividade das culturas, o material é encaminhado ao lacen para realização de TS, fortalecendo o diagnóstico nos municípios. O LACEN-PA apoia todos os municípios do estado na realização de culturas e baciloscopia, na ausência de estrutura instalada para tal.

É importante ressaltar a importância do LACEN-PA na organização da rede de diagnóstico da tuberculose no Pará, sendo o único laboratório que realiza testes de sensibilidade. Também realizando o controle de qualidade dos diagnósticos baseados em baciloscopia e treinamento das equipes de unidades laboratoriais do estado, de acordo com as competências dos LACENS estabelecidas na portaria 2606 do MS (Ministério da Saúde, 2021, p. 5)

Conforme o representado na figura 3, o fluxo de trabalho atual do LACEN-PA consiste no recebimento de amostras na forma de biópsia, escarro ou isolado bacteriano, a depender da disponibilidade do procedimento de cultura na unidade onde o paciente foi atendido. A partir do isolado são realizadas baciloscopia confirmatória e teste MPT-64 (teste baseado em anticorpos do CMTB), para identificação preliminar de culturas sugestivas para micobactérias não tuberculosas (MNT) ou confirmação do diagnóstico de CMTB.

Figura 3: Diagrama representando o fluxo de Diagnóstico no Pará



Fonte: Elaborado pelo autor

A Cepa é então submetida a Teste de sensibilidade de primeira linha utilizando uma das metodologias disponíveis: MGIT, LPA ou proporções. No caso da detecção de resistência a rifampicina, isoniazida e outra droga, as cepas são enviadas por transporte aéreo para a referência nacional, atualmente o Centro de Referência Professor Hélio Fraga em Jacarepaguá no Rio de Janeiro. Onde será realizado o teste de sensibilidade para drogas de segunda linha (aminoglicosídeos e fluoroquinolonas) de acordo com as metodologias preconizadas pelo MS (LPA ou MGIT).

3.6 A ROTINA DE DETECÇÃO DE RESISTÊNCIA

O padrão de referência para o TSA são os métodos baseados em cultivo utilizando meios sólidos (método das proporções) ou meio líquido (MGIT - Método Automatizado) combinados com soluções padronizadas das drogas. Devido ao crescimento lento da espécie, esses testes podem levar de semanas a meses para produzirem resultado, afetando a escolha de regime de tratamento (OMS, 2018a).

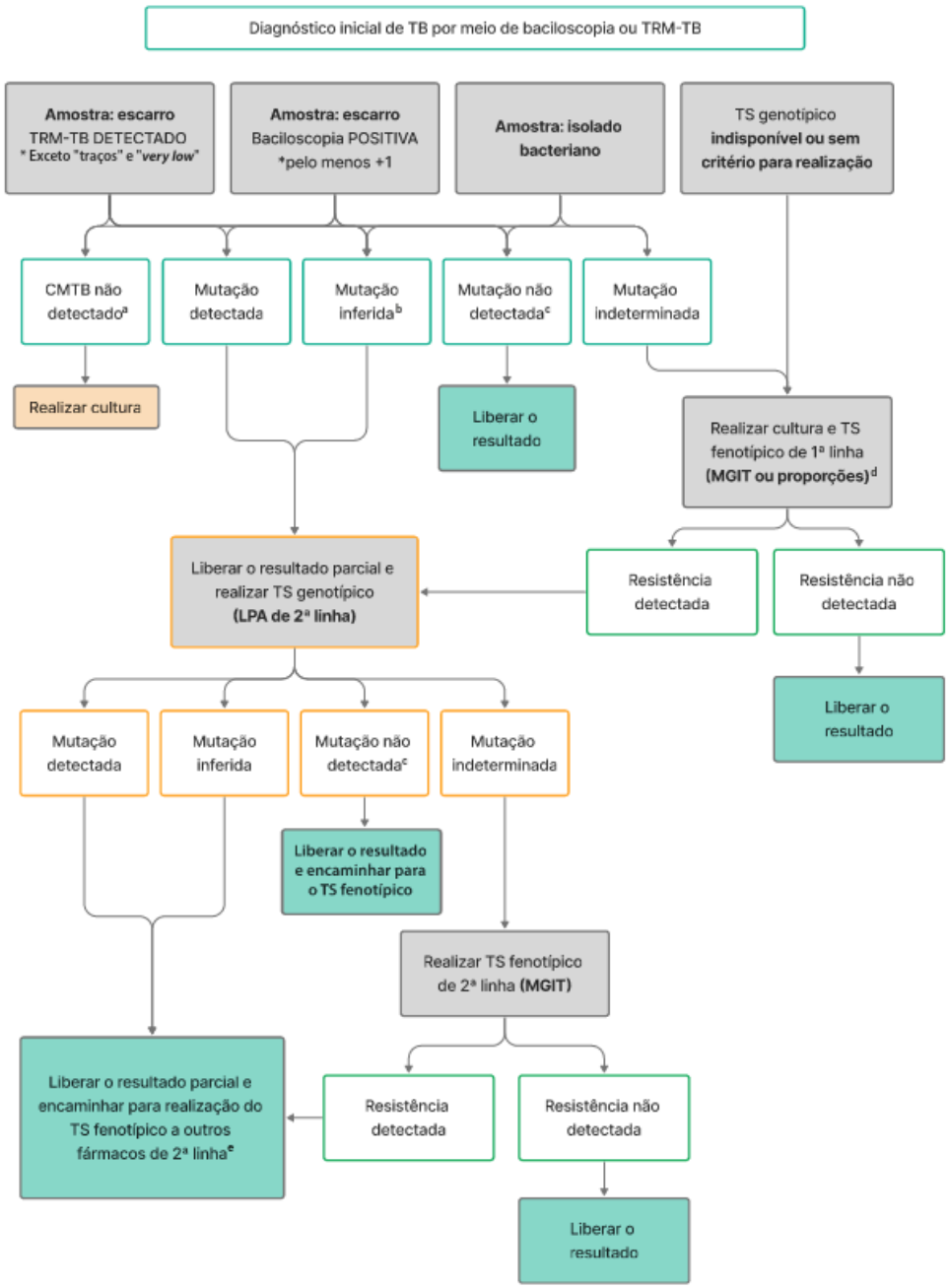
A detecção indireta da resistência antimicrobiana também pode ser feita por TRM: detecção da rifampicina (Xpert/MTB RIF), o *Line Probe Assay* (LPA) para rifampicina e isoniazida (Genotype MTBDRplus v.2.0) e para fluoroquinolonas e drogas de segunda linha (Genotype MTBDRsl v.2.0), sendo este último apenas disponível em laboratórios de referência regional e nacional, com expectativa da incorporação do teste na rede dos LACENs quando a distribuição do produto for normalizada pelo Ministério da Saúde.

Alternativamente, o sequenciamento parcial de genes relacionados à resistência (*rpoB*, *katG* e *inhA*) ou por meio do SCG permitem a predição da resistência utilizando como

interpretação para associação de resistência ou não o catálogo de mutações da OMS (Consortium, 2018, 2022; OMS, 2022b, 2023). Ressalta-se que a utilização do SCG para detecção do perfil de suscetibilidade antimicrobiana ainda não está disponível para fins de diagnóstico no Brasil e deve ser utilizado apenas para fins de pesquisa.

Após a incorporação do LPA pela portaria "PORTARIA SCTIE-MS Nº 43, DE 6 DE JULHO DE 2021", a estrutura básica da detecção de resistência foi alterada de acordo com o representado nas recomendações de técnicas para o LPA e na figura 2 (Brasil, M. da S., Secretaria de Vigilância em Saúde e Ambiente, Departamento de HIV/Aids, Tuberculose, Hepatites Virais e Infecções Sexualmente Transmissíveis., 2023).

Figura 2: Fluxo diagnóstico atualizado pelo Ministério da Saúde.



Fonte: (Brasil, M. da S., Secretaria de Vigilância em Saúde e Ambiente, Departamento de HIV/Aids, Tuberculose, Hepatites Virais e Infecções Sexualmente Transmissíveis., 2023)

Sendo o LPA tratado como principal forma de detecção de resistência, e o TS fenotípico tratado como confirmatório dos exames genotípicos, considerando que TS baseados em cultura devem continuar sendo realizados, independente do resultado do LPA[1,2].

[1] Durante o presente estudo, LPAs de segunda linha não estavam disponíveis na rede de LACENS
[2] Durante o presente estudo, o LACEN-PA realizou LPAs a partir da cultura

A norma sugere que o Teste de Sensibilidade baseado em LPA ou TRM seja realizado diretamente do escarro, diferenciando três casos: tuberculose droga-resistente, tuberculose sensível e negativo para tuberculose. Caso seja detectada uma mutação indeterminada, não haja disponibilidade para teste genotípico ou não seja detectado um membro do CMTB, a cultura deve ser realizada para determinação de teste fenotípico. Na detecção de mutação determinada, o teste genotípico de segunda linha deverá ser realizado e a amostra deverá ser encaminhada para teste de sensibilidade fenotípica.

3.7 A RESISTÊNCIA ÀS DROGAS ANTITUBERCULOSE

A classificação de resistência em *M. tuberculosis* var. *tuberculosis* foi recentemente atualizada pela OMS, podendo ser classificada em MDR, pre-XDR, XDR. Sendo pré-XDR uma nova classificação que abrange as estirpes classificadas como MDR que também são resistentes a isoniazida e qualquer outra fluoroquinolona e a reclassificação de XDR, sendo estirpes MDR com resistência fluoroquinolonas e à pelo menos uma droga do grupo A (bedaquilina, linezolida, levofloxacina, moxifloxacina) (OMS, 2021).

Majoritariamente a aquisição de resistência no CMTB ocorre aleatoriamente através de polimorfismos de um único nucleotídeo (SNPs), inserções e deleções (indels) não sendo reportados consistentemente a ocorrência de transferência horizontal (Allué-Guardia; García; Torrelles, 2021). O mecanismo de aquisição de resistência aos fármacos anti-TB no CMTB ocorre principalmente por mutações espontâneas em genes relacionados às proteínas-alvo dos fármacos durante o processo de multiplicação (Almeida Da Silva; Palomino, 2011).

Além disso, a resistência antimicrobiana pode ser classificada em inata ou intrínseca, a partir das estruturas existentes na célula micobacteriana que impedem a ação do fármaco, ou mecanismos que inativam ou modificam a molécula-alvo do fármaco. Os principais mecanismos de resistência inata do CMTB são: bombas de efluxo, modificação de alvos, modificação de antibióticos, degradação enzimática de antibióticos (Smith; Wolff; Nguyen, 2013).

3.8 ESQUEMAS DE TRATAMENTO DA TUBERCULOSE E DA TUBERCULOSE DROGA-RESISTENTE

Para o tratamento da TB no Brasil, o governo federal através da secretaria de vigilância em saúde via Ministério da Saúde, propôs uma série de recomendações quanto aos esquemas de tratamento, os quais são divididos em duas fases: intensiva (redução da população bacilar) e manutentiva (eliminação de bacilos persistentes) (Brasil, 2019).

O esquema básico para para tratamento de TB pulmonar em adultos e adolescentes é composto por dois meses de tratamento utilizando rifampicina, isoniazida, pirazinamida e etambutol (fase intensiva) e quatro meses utilizando rifampicina e isoniazida (2RHZE/4RH). Em casos onde a resposta clínica for insatisfatória a fase manutentiva pode ser estendida por até três meses (Brasil, 2019).

A falência no tratamento básico é caracterizada pela persistência de baciloscopia de escarro positiva ao final do tratamento; baciloscopia fortemente positiva (++ ou +++) com positividade até o quarto mês de tratamento e baciloscopia inicialmente positiva, seguida de negativação e nova positividade, por dois meses consecutivos, a partir do quarto mês de tratamento. No caso da suspeita de resistência, o paciente deve ser encaminhado para referência terciária (Brasil, 2019).

O acompanhamento do tratamento deve ser realizado na Unidade Básica de Saúde mais próxima da residência do indivíduo. Em caso de reações adversas, evolução desfavorável ou detecção de resistência, o caso deverá ser encaminhado para a referência de tratamento a tuberculose para ajuste do tratamento, o Ministério da saúde também recomenda a adoção do tratamento diretamente observado nestes casos (Brasil, 2019).

Quando detectado resistência a rifampicina no teste rápido molecular ou resistência aos fármacos de primeira linha no TS, deve ser administrado esquema para TB-DR. O qual deve ser composto por, pelo menos, três fármacos que devem ser selecionados baseados nos seguintes critérios: não terem sido usados pelo paciente em outro tratamento e efetividade testada através de teste de sensibilidade. Sendo pelo menos dois fármacos do grupo essencial e ao menos um do grupo dos acompanhantes e o esquema podendo ser revisto após a evolução clínica, laboratorial e radiológica ser considerada desfavorável pelo agente de saúde responsável (Brasil, 2019).

O esquema básico para tuberculose rifampicina resistente e multirresistente vigente no período do estudo (nov/2021-dez/2022) é composto por bedaquilina, levofloxacino, linezolida e terizidona (BLevLTzd). No caso de disponibilidade de TS para drogas de segunda linha ou presença fatores de risco ligados ao paciente (intolerância medicamentosa, comorbidades ou interações medicamentosas), outros fármacos podem ser incluídos como: moxifloxacino, clofazimina, imipenem-cilastatina, meropenem, amicacina, estreptomicina, etionamida e ácido paraminossalicílico. Compondo assim um esquema de tratamento individualizado (Brasil, 2021).

Quanto à duração do tratamento para TB-DR, a fase intensiva (com o uso da bedaquilina) deve durar de seis a oito meses, sendo considerada completa após duas culturas

negativas com intervalo de pelo menos 30 dias. Em demora de conversão da baciloscopia para negativa suspeita de ineficácia de fármacos durante fase manutenção ou de resistência a fluoroquinolonas a fase intensiva pode ser prolongada. Já a fase de manutenção deve durar de 18 a 24 meses, sendo o tratamento completo quando três culturas foram negativas (Brasil, 2019, 2021).

É Importante ressaltar que em Janeiro de 2025, o tratamento de RR-TB e MDR foi alterado para a incorporação do esquema oral encurtado, incluindo pretomanida após sua aprovação pela CONITEC. O novo esquema é baseado em bedaquilina, pretomanida e linezolida com adição do moxifloxacino em alguns casos (BPaL e BPaLM), de acordo com a nota informativa N º 1/2025 e o manual de recomendação de tratamentos da OMS (Brasil, 2023, 2025; OMS, 2022c).

3.9 BEDAQUILINA E SUA RESISTÊNCIA

A Bedaquilina (BDQ) é uma diarilquinolona comercializada sob nome Sirturo, cujo principal mecanismo de ação é a inibição da ATP sintase em micobactérias. Tendo um alto poder bactericida, mesmo contra bactérias resistentes a outras drogas, incluindo outras quinolonas. Os estudos clínicos iniciais sugeriram um valor de concentração mínima baixo para 99% das micobactérias testadas (Andries *et al.*, 2005).

A incorporação da BDQ ao tratamento da tuberculose foi vista como revolucionária pelo seu alto fator esterilizante e por obter sucesso em ensaios clínicos após 40 anos sem a incorporação de novos fármacos aos esquemas de tratamento de tuberculose. Apesar disto, a farmacovigilância torna-se necessária uma vez que diversos casos de falência terapêutica e desenvolvimento de resistência já são relatados na literatura. Portanto a adição ao SUS deve ser observada com cautela, mesmo sem relatos de desenvolvimento de resistência à bedaquilina no Brasil. O uso racional desse medicamento aliado às tecnologias de vigilância genômica são estratégias importantes no combate da tuberculose droga resistente, e são recomendadas pela OMS dentro da estratégia END TB.

Quanto a mutações relacionadas a resistência a bedaquilina o catálogo de mutações da OMS em sua segunda versão traz 6 mutações no Rv0678 classificadas como grupo 1 de confiabilidade e novas mutações nas regiões Rv0678, atpE e pepQ classificadas como grupo 2 de confiabilidade. Revisando os resultados da versão 1, que não possuía nenhuma variante classificada com alto grau de confiabilidade. Interessantemente, 42 variantes nas regiões Rv0678, mmpL5, Rv1979 e mmpS5 são classificadas como incertas ou de impacto

desconhecido, ressaltando a necessidade de estudos a respeito do desenvolvimento dos mecanismos de resistência à BDQ (OMS, 2023).

Recentemente uma revisão procurou buscar e elucidar os mecanismos de resistência à bedaquilina, sendo o gene mmpR5 o único estatisticamente relevante para a aquisição de resistência. Através da simulação computacional do gene mmpR5 (rv0678), foi predito que as principais mutações relacionadas à maior impacto para aquisição de resistência são aquelas que modificam o sítio de ligação da proteína codificada pelo gene com o DNA. Impactando a atividade do fármaco na proteína. Apesar da descrição importante, o artigo ressalta a importância da realização de novos estudos a respeito, em virtude do baixo número amostral de cepas resistentes à bedaquilina com testes fenotípicos de resistência de qualidade (Nimmo *et al.*, 2024).

O surgimento de resistência à bedaquilina é uma ameaça para o enfrentamento à tuberculose, existindo relatos na literatura de desenvolvimento de resistência após a incorporação desse fármaco em outros países. Demonstrando uma crescente necessidade no estudo aprofundado do desenvolvimento de resistência, bem como na adequação dos esquemas de tratamento frente a esses fenômenos (Satapathy *et al.*, 2024).

Como forma de explorar a resistência a bedaquilina na ausência de testes fenotípicos padronizados, podem ser utilizadas técnicas *in silico* a partir dos variantes detectados de genomas com o fenótipo resistente como o uso de modelagem computacional de proteínas, simulação de dinâmica molecular e estudos de *GWAS* que são considerados promissores no âmbito da pesquisa para a melhor compreensão dos mecanismos por trás do desenvolvimento dessas resistências (Cohen *et al.*, 2019; Sonnenkalb *et al.*, 2023).

Snobre e colabores (2024) exploram diversas mutações no gene rv0678 através de uma metodologia inovadora que combina o uso de ferramentas in silico para a predição do fenótipo baseada em sequenciamento de alta profundidade para identificação de variantes e a simulação de proteínas, buscando diminuir o vácuo entre a detecção de mutações relacionadas a resistência e seu real impacto no perfil fenotípico de resistência às drogas ao comparar os valores obtidos computacionalmente com os de concentração inibitória mínima.

3.10 VIGILÂNCIA GENÔMICA DA TUBERCULOSE

A vigilância genômica inclui a utilização do SCG na rotina diagnóstica, incorporando de ferramentas de bioinformática ao rastreamento e investigação de casos da TB (OMS, 2018b). Ao trazer tecnologias comumente utilizadas na pesquisa acadêmica para o diagnóstico clínico é possível integrar a atenção individualizada para detecção sensível de

variantes a atenção populacional através da observação e incorporação das informações genômicas nos estudos epidemiológicos.

No Brasil, os estudos utilizando o SCG estão sendo documentados banco de dados Genomas de Micobactérias do Brasil (GEMIBRA) <https://site.ioc.fiocruz.br/gemibra/> em parceria com o Laboratório de Biologia Molecular Aplicada às Micobactérias da Fundação Oswaldo Cruz (LABMAM-Fiocruz), o qual está em fase de desenvolvimento para tornar o banco de dados interativo e auxiliar na vigilância genômica das micobactérias como apoio a CGDR/DCCI/SVS/MS (Conceição *et al.*, 2021). A proposta reúne todos os genomas públicos de Micobactérias do Brasil, através de uma curadoria semi-automatizada, e está em fase de atualização e ampliação pelo grupo que coordena esta atual proposta. Até o momento (27 de Janeiro de 2025) foram publicados 2.580 genomas de CMTB no Brasil e poderão ser incluídos comparativamente em estudos de diversidade genética do CMTB no Brasil.

Nesse contexto ressalta-se a importância da incorporação de computadores de alto desempenho nos processos de vigilância, bem como o uso de fluxo de dados reprodutíveis baseados em linguagens voltadas para reprodutibilidade como Nextflow. Uma vez que a quantidade de dados gerados por projetos de vigilância genômica é substancialmente grande e precisa ser analisada de maneira efetiva e padronizada (Afolayan *et al.*, 2021).

Para a analisar esse tipo de conjunto de dados, a OMS sugere o uso de *pipelines*, apesar da OMS não reforçar o uso de uma ferramenta específica, mas sim, recomendar processo de chamada de variantes específicos internos aos softwares, podem ser destacados dois principais pipelines para análise de TB: MTBseq e MAGMA, recentemente, um estudo comparativo baseado em *benchmark*, sugeriu que o MAGMA seja uma ferramenta de melhor usabilidade por reforçar o uso específico do catálogo de mutações da OMS, enquanto MTBseq utiliza um banco de dados próprio. Além disto, o MAGMA realiza controle de qualidade das amostras interno, enquanto o MTBseq muitas vezes mantém amostras que não foram aprovadas no controle em suas análises finais (Heupink *et al.*, 2023; OMS, 2018b, 2022a, 2023).

4. OBJETIVOS

4.1 OBJETIVO GERAL

Realizar a epidemiologia genômica do *Mycobacterium tuberculosis* caracterizando transmissão, epidemiologia e resistência antimicrobiana durante a inserção da bedaquilina nos esquemas de tratamento da tuberculose droga-resistente.

4.2 OBJETIVOS ESPECÍFICOS

1. Implementar fluxo de análise acessível para usuários com pouca experiência para a análise de genomas do CMTB;

2. Propor um fluxo de sequenciamento de genoma completo a partir da cultura utilizada na rotina de diagnóstico e a análise de seus resultados;

3. Descrever o perfil epidemiológico e casos de transmissão de TB droga resistente em associação com dados clínicos e genômicos;

4. Identificar e Investigar padrões genômicos de susceptibilidade antimicrobiana das estirpes do CMTB por meio de WGS;

5. Investigar fenômenos de infecção mista e de heterorresistência.

6. Investigar o impacto de mutações relacionadas à resistência à bedaquilina e seu impacto em modificações estruturais de proteínas do genoma.

5 PRODUÇÃO TÉCNICA E CIENTÍFICA

5.1 ARTIGOS

5.1.1 Mtbseq-Nf: Enabling Scalable Tuberculosis Genomics “Big Data” Analysis Through A User-Friendly Nextflow Wrapper For Mtbseq Pipeline

Esta produção atende aos objetivos específicos: 1

Este trabalho compreende o objetivo específico 1 do trabalho, e busca a implementação reprodutível da análise genômica de MTB utilizando um wrapper (código de software que busca tornar outro software mais acessível para o usuário) escrito na linguagem nextflow, e representa a colaboração do autor com Abhinav Sharma, dividindo a primeira autoria no estudo e colaborando de maneira pública em código aberto (https://github.com/mycobactopia-org/MTBseq-nf).

MTBseq-nf: Enabling Scalable Tuberculosis Genomics "Big Data" Analysis through a User-Friendly Nextflow Wrapper for MTBseq pipeline

Abhinav Sharma[1], Davi Josué Marcon[2], Johannes Loubser[1], Karla Valéria Batista Lima[2,3], Gian van der Spuy[1], Emilyn Costa Conceição[1]

[1] SAMRC Centre for Tuberculosis Research; Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, Western Cape, South Africa
[2] Instituto de Ciências Biológicas e da Saúde, Universidade do Estado do Pará, Belém, Pará, Brazil
[3] Seção de Bacteriologia e Micologia, Instituto Evandro Chagas, Ananindeua, Pará, Brazil

Corresponding Author:
Abhinav Sharma[1]
Francie Van Zijl Drive, Tygerberg, Cape Town, Western Cape, 7505, South Africa
Email address: abhinavsharma@sun.ac.za

Abstract

Background. The MTBseq pipeline, published in 2018, was designed to address bioinformatics challenges in tuberculosis research using whole-genome sequencing data. It was the first publicly available pipeline to perform full analysis from quality control through mapping, variant calling for lineage classification, drug resistance prediction, and phylogenetic inference. However, the pipeline's design was not optimal for high-performance computing or cloud computing environments, which often require large datasets.

Methods. To optimize the software, a wrapper in the Nextflow language (MTBseq-nf) was created. This wrapper offers shorter execution times, automatic setup of conda environments and docker containers, a new parallel mode of execution, and allows addition of new functionality like a custom MultiQC report. The MTBseq-nf wrapper uses the modular nature of the standard MTBseq TBfull analysis step, which analyzes all input raw FASTQ files linearly. The MTBseq-nf wrapper, however, can execute multiple instances of the same step in parallel and therefore makes full use of the provided computational resources. For validation of results as well as execution time, we used 90 *Mycobacterium tuberculosis* genomes (ENA accession code PRJEB7727) for the benchmarking analysis on a dedicated computing server (32 vCPUs and 64GB RAM).

Results. The execution time of MTBseq-nf parallel analysis mode (8h 21s) is at least twice as fast as the standard MTBseq pipeline (20h 54m 46s). Furthermore, the MTBseq-nf wrapper facilitates reproducibility using the nf-core, bioconda, and biocontainers projects for platform independence. The proposed MTBseq-nf is a user-friendly pipeline optimized for hardware efficiency, scalability for larger datasets, and improved reproducibility compared to the original MTBseq.

Introduction

Next-generation sequencing (NGS) has revolutionized tuberculosis (TB) research, diagnosis, and surveillance, however analyzing this complex bioinformatics data presents computational challenges. In 2018, the MTBseq (Kohl et al., 2018) pipeline was published to address these challenges of data analysis and improve the reproducibility of whole-genome sequencing (WGS) analysis of *Mycobacterium tuberculosis* complex.

MTBseq was one of the first comprehensive end-to-end pipelines and has been widely used by researchers working with WGS data to generate and interpret the principal outputs namely (i) strain classification, (ii) phylogenetic tree, (iii) mapping and variant statistics, (iv) SNP distance matrix and (v) cluster groups. Traditionally, the researchers have relied upon in-house computing infrastructure such as university servers or job orchestrators such as the traditional high-performance computing (HPC) environments and modern cloud batch computing platforms such as AWS, Google, Microsoft Azure, Kubernetes and Hashicorp Nomad to analyze the increasing volume of genomic data (Stephens, 2015).

The application and utility of genomic data analysis is projected to only increase in precision public health and precision medicine for foreseeable future, the MTBseq pipeline in it current form performs suboptimally for these environments therefore we aimed to adapt the MTBseq pipeline for job orchestrators and large-scale genomic analysis which for keeping the costs under control and the analysis time predictable.

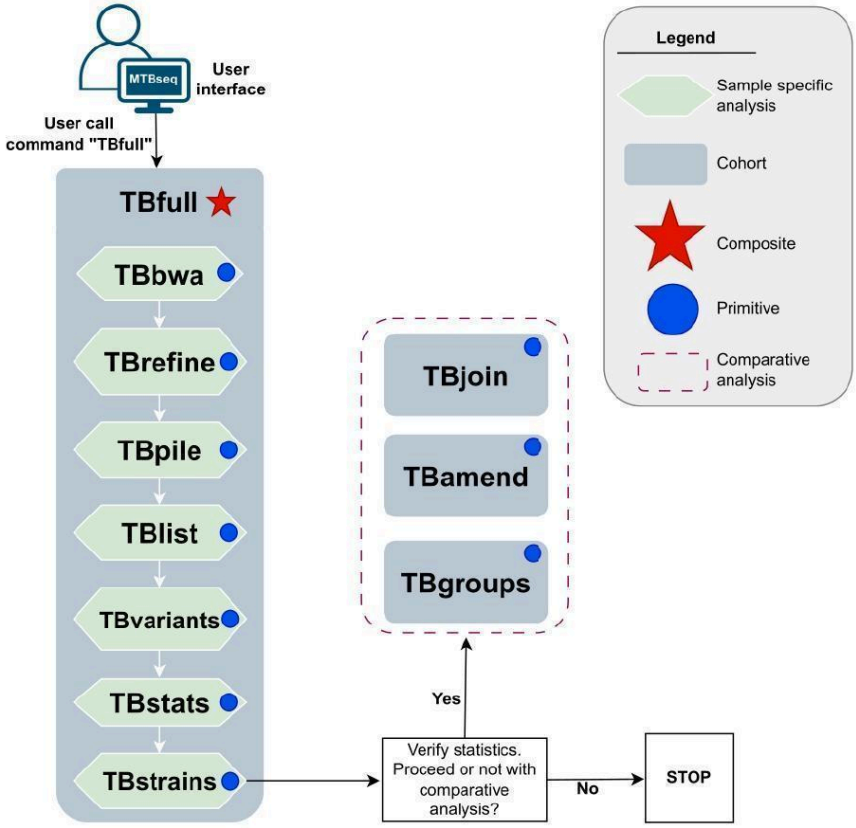
We developed MTBseq-nf, a wrapper built with the Nextflow language (Di Tommaso et al., 2017) _ to provide an alternative to current users of MTBseq with improved (i) user-friendliness, (ii) maintainability, (iii) reproducibility and (iv) scalability, for large-scale data analysis on modern job orchestrators.

Materials & Methods

The design of original MTBseq pipeline

The MTBseq pipeline by Kohl et al. (2018) relies upon underlying tools such as GATK3 (McKenna et al., 2010), PICARD (Picard Toolkit, 2019), BWA (Li & Durbin, 2009) and SAMTOOLS (Li et al., 2009); and glues these tools together using the perl5 programming language (Figure 1) (Wall et al., 1994). Its execution model relies upon core abstractions (called steps) such as reference mapping (TBbwa) and variant calling (TBvariants) which can be broadly

classified into two groups (i) Group-1, based on scope, with sample-specific steps (TBbwa, TBvariants) and cohort-specific steps (TBamend, TBgroups) and (ii) Group-2, based on composition, with primitive single-purpose steps (TBbwa, TBgroups), and composite multi-purpose step (TBfull) combining many of the primitive steps to facilitate routine usage of certain primitive steps. Table 1 offers further information regarding the groups in MTBseq.



Implementation of MTBseq-nf wrapper

We aimed to develop an MTBseq-nf wrapper, on top of the MTBseq pipeline, to implement numerous enhancements across four broad themes (i) user-friendliness (ii) scalability (iii) reproducibility and (iv) maintainability, summarized in Table-2. The implementation of the MTBseq-nf wrapper pipeline builds upon two core components (i) the unmodified MTBseq

pipeline as published by Kohl et al. (2018) and (ii) the use of Nextflow workflow manager as per nf-core best-practices template.

The standard MTBseq pipeline published by Kohl et al. (2018) exposes a --step parameter allowing users to control which step of the pipeline they wish to initiate as shown in Figure 1. The pipeline facilitates the usage of sequential sample-specific steps on a routine basis, through a composite TBfull step which automates the execution of sample-specific primitive steps from TBbwa until TBstrains. Upon completion of these steps, the users are expected to provide a TSV file, containing a list of samples and corresponding library names, that should be included in further comparative analysis steps such as TBjoin, TBamed and TBgroup generating the principal outputs of the pipeline as described in Table 2.

Report name	Description	Primitive step name	File location in MTBseq results
Classification	For lineage classification based on a set of phylogenetic SNPs. Within this tab-delimited file, the majority lineage is reported for each dataset. This entry also gives an indication of the data quality for the positions used to infer the phylogenetic classification. The column indicating the quality will contain the label good if all phylogenetic positions contained in the set are covered at least 10fold and show a frequency of at least 75%, bad if any phylogenetic position does not meet these, and ugly if any phylogenetic position does not have a clear base call.	TBstrains	Classification/Strain_Classification.tab

Mapping and variant statistics	For calculating an overview of mapping quality and detected variants for a dataset, using the SAMTOOLS flagstat program. The generated tab-delimited file stores all sample statistics for the analyzed datasets present in the working environment. This module employs the same thresholds to discern unambiguously covered positions as the TBvariants module (set by the –mincovf, –mincovr, –minphred20, –minfreq).	TBstats	Statistics/Mapping_and_Variant_Statistics.tab
SNP distance matrix	For inferring likely related isolates based on pairwise distance of distinct SNP positions. In an agglomerative process, samples are grouped together if they are within the threshold set by the –distance OPTION. The output files consist of a text file listing the detected groups and ungrouped isolates, and the calculated pairwise distance matrix.	TBgroups	Groups/[PROJECT]_joint_[mincovf]_[mincovr]_[minfreq]_[minphred20]_samples_amended_[unambig]_phylo_[window].matrix
Cluster groups	For inferring likely related isolates based on pairwise distance of distinct SNP positions. In	TBgroups	Groups/[PROJECT]_joint_[mincovf]_[mincovr]_[minfreq]_[minphred20]_samples_amended_[unambig]_phylo_[win

	an agglomerative process, samples are grouped together if they are within the threshold set by the –distance OPTION. The output files consist of a text file listing the detected groups and ungrouped isolates, and the calculated pairwise distance matrix.		dow]_[distance].groups
Phylogenetic tree	For post-processing of joint variant tables. This step produces a comprehensive variant table including calculated summary values for each position. In addition, the set of positions will be processed in consecutive filtering steps. In the first step, positions will be excluded if less than a minimum percentage of samples have unambiguous base calls at this position, with this threshold set by the –unambig OPTION. In addition, all samples need to have either a SNP or wild type base at the position, and positions within repetitive regions of the reference genome or within a resistance associated genes are excluded. The second filtering step	TBamend	Amend/[PROJECT]_joint_[mincovf]_[mincovr]_[minfreq]_[minphred20]_samples_amended_[unambig]_phylo.fasta

	removes positions that are located within a maximum distance to each other in the same sample, with this threshold set by the –window OPTION.		

Nextflow is a workflow management system designed to address the challenges of high-throughput NGS data at scale in a reproducible manner (Di Tommaso et al., 2017), allowing researchers to create complex workflows that integrate multiple bioinformatics tools into a single, cohesive workflow while being flexible and compatible with workstations, servers, HPC clusters, and cloud computing infrastructure. This portability is crucial for research groups and organizations with varying levels of expertise and access to computing resources. The nf-core community further enhances the Nextflow ecosystem by bringing together Nextflow users through a Slack group, hackathons, seminars, training, and other community initiatives that foster collaboration and knowledge sharing (P. A. Ewels et al., 2020; Langer et al., 2024). The community also develops and maintains a collection of Nextflow pipelines that are widely used in life science research and refined by experts worldwide. Furthermore, these pipelines are designed with strict guidelines, as per the best-practices template, to ensure robustness, reproducibility, and ease of use building upon other nf-core projects such as nf-core/configs and nf-core/modules.

Parallel mode in MTBseq-nf

For MTBseq pipeline, the input sequence files are required (i) to adhere to a particular naming convention, [SampleID]_[LibID]_[*]_[Direction].f(ast)q.gz and (ii) to be located within the directory of invocation where the relevant outputs, from the different processes in the pipeline, are generated and stored. These requirements concerning input file names and locations align very well with the implementation of different steps in the pipeline which rely upon the sequential analysis of input sequences using foreach loops and only proceeds towards the next step once all the samples have passed through an individual step.

The MTBseq-nf wrapper pipeline, on the other hand, offers two modes of execution (i) default, equivalent to the MTBseq pipeline's TBfull step and (ii) parallel, which can be enabled through the use of --parallel parameter. The key insight we relied upon, for the implementation of --parallel mode, was to combine the strengths of both (i) MTBseq's modular stepwise design and the exposed --step parameter which allows independent execution of the individual steps of the MTBseq pipeline (ii) the native task parallelization offered by the Nextflow workflow language. Therefore, when executed with the --parallel parameter, the MTBseq-nf wrapper pipeline, utilizes the very same primitive steps invoked by the TBfull step and optimises the flow of samples

through these primitive steps (Figure 2) with the benefit of a substantial reduction in the overall time required to execute the pipeline, especially when the number of samples increases in the cohort.

In essence, we rely upon the fact that when a sample-specific step is run with a single sample, the foreach loop executes only once since there is only one sample visible to that specific step which is analyzed quickly, facilitating the scalability of the overall analysis for larger number of samples and avoiding the trade-off of the linear increase in execution time with larger cohorts. Furthermore, the MTBseq-nf pipeline automates the creation of TSV file, once all the samples are analyzed through the primitive steps of TBfull step, and invoke the comparative analysis stage, mirroring the same behavior as the original MTBseq pipeline for the remaining steps.

Preprocessing of Validation Dataset

Original dataset

To evaluate the reproducibility and scalability of the MTBseq-nf wrapper in terms of (i) growth of execution time versus the cohort size as well as (ii) the validity of results produced, we relied upon the original dataset of 91 samples used by Kohl et al. (2018) as part of the original MTBseq publication. The dataset, as described by Schleusener et al. (2017) is publicly available under ENA accession code PRJEB7727. Furthermore, Kohl et al. (2018) analyzed 91 samples for resistance and lineage profiling in a comparative setting as part of supplementary 2 of the original publication. These IDs are shared in Supplementary 1.

Dataset cleanup

The data submitted to ENA contains a total of 133 paired-read samples, and upon close observation, we noticed that (i) some samples had multiple corresponding experiment_accession (Supplementary-FIXME provides a frequency table of secondary_sample_accession against experiment_accession.) (ii) the samples used by Kohl et al. (2018) had differences with the samples submitted under PRJEB7727 to ENA and (iii) one sample ERS457325 listed by Kohl et al. (2018) did not have any corresponding files under the indicated ENA submission.

Therefore, to reduce the confounding factors, we decided to deduplicate the samples rather than merging the samples with the same accession numbers by relying upon the first occurrence of specific secondary_sample_accession IDs from the files obtained from ENA project and to drop missing ERS457325 from our analysis, reducing the final dataset to 90 samples. The final sample sheet is shared in Supplementary-FIXME.

Design of validation datasets

We divided these samples into 6 cohorts with an incremental number of samples namely 5, 10, 20, 40, 80 and finally 90 samples. The largest dataset container 90 samples was used for

the validation of principal outputs generated by MTBseq and MTBseq-nf (default and parallel) pipelines.

Computational Resources for Validation

To conduct the computational experiments, we opted for a virtual server machine deployed on the Oracle Cloud Infrastructure, with the following hardware characteristics of 32 CPUs (16 CPUs), 64 GB RAM memory and a boot disk size of 2 TB. The necessary minimal setup for Java (Arnold et al., 2005), Nextflow (Di Tommaso et al., 2017)_ and Docker (Merkel, 2014)_ installations. The decision to use a server was key for our study of execution runtime since HPC queue systems introduce an unpredictable wait-time delay before a submitted job executes. The docker platform was selected due to its widespread usage in cloud batch computing platforms as well as modern container orchestrator systems such as Kubernetes and Hashicorp Nomad. The choice of docker containers was also motivated by the upstream biocontainer (Veiga Leprevost et al., 2017) assets which are created from the bioconda recipes and therefore, allow us to minimize the risk of non-reproducibility due to infrastructural differences and facilitate an isolated analysis of the principal results of the pipeline.

Triplicated Intra- and Inter-modal Evaluation of Scalability and Reproducibility

To ensure that there are no regressions in the results produced by MTBseq-nf (default and parallel modes) when compared with the MTBseq pipeline, we compared the principal results produced by MTBseq-nf through triplicated sets of experiments for (i) intra-modal analysis to evaluate reproducibility within each mode and (ii) inter-modal analysis to evaluate reproducibility between different modes.

As a result of this triplicated experiment design, we analyzed the output of 9 different experiments. The parameters we used for these runs are described in supplementary-FIXME while Table-FIXME summarizes the run names and notable parameters for a specific run. Furthermore, each individual experiment was done in isolation from others without utilizing Nextflow's resume functionality.

Scalability

For the growth of execution time for the aforementioned six cohorts, we conducted the analysis on these datasets in the MTBseq-nf default mode as well as parallel mode. For the baseline MTBseq pipeline, we relied upon the classic UNIX utility of "time" with the docker command for mounting the samples and then initiating the "TBFull" analysis as documented in Supplementary material-FIXME. The pipeline executions for MTBSeq-nf default and parallel modes were tracked and visualized through the Seqera Platform (formerly known as Nextflow Tower) (FIXME citation) which provides a central location for monitoring the pipeline execution metrics and storing the pipeline configuration.

Evaluation of Reproducibility

For the evaluation of reproducibility of inter- and intra-modal triplicated experiments, we compared the principal results through (i) 3-way diffs using Araxis Merge (FIXME citation) and (ii) range analysis of numerical data through R language (FIXME citation) and (iii) phylogenetic tree generated using IQTREE (FIXME citation).

For majority of principal results, which are qualitative in nature, a 3-way visual diff report was sufficient to evaluate the differences in results. However, for the mapping and variant statistics generated by TBstats (such as Mapped Reads, SNPs), we relied upon range statistics (min-max analysis) in R language for estimating the relative changes across different intra- and inter-modal experiments. The scripts used in the analysis and for visualizations have been provided in the following repository https://github.com/abhi18av-phd-projects/mtbseq-nf-publication-analysis.

Results

The MTBseq-nf wrapper, not only offers substantial enhancements on top of the standard MTBseq pipeline, but it also implements a parallel mode, which speeds up the overall execution time for large cohorts.

Improved scalability with MTBseq-nf

Figure 3 highlights the performance characteristic of MTBseq-nf default and parallel mode, which should hold as long as the computational resources in the servers/cluster are readily available.

TODO: Add the results of MTBseq analysis using the generated logs and time utility.

Improved reproducibility with MTBseq-nf

Intra-modal analysis of triplicated experiments

The triplicated experiments were conducted with the same (i) parameters, (ii) dataset and (iii) infrastructure as described in Table 3. The results of (i) Classification and (ii) SNP distance matrix demonstrated stability across experiments apart from the variable date of execution, which is expected since the successive runs could only be initiated after the previous run terminated. Similarly, the (iii) Phylogenetic trees were also stable across different experiments, owing to the stability of SNP distance matrix. TODO: Add the phylogenetic trees The results of (iv) Cluster groups were assigned labels, based on the SNP matrix, however the exact label fluctuates amongst different runs of the TBgroups step, due to the nature of computations done by an agglomerative algorithm and concerns itself with the accurate grouping of samples. FIXME SUPPLEMENTARY-1 contains the HTML diff report generated by Araxis merge software (FIXME cite) and demonstrates this label fluctuation across the three runs.

There were, however, some differences the (v) Statistics report generated by the TBstats process in triplicated runs of the MTBseq and MTBseq-nf default mode as shown in Figure-5.

It is worth highlighting that none of the numerical fields demonstrated any differences across triplicated runs of MTBseq-nf parallel mode.

Inter-modal analysis of MTBseq vs MTBseq-nf (default)

TODO Consider changing the table to show something more meaningful like fractional changes.

Inter-modal analysis of MTBseq-nf (parallel) vs MTBseq-nf (default)

Inter-modal analysis of MTBseq-nf (parallel) vs MTBseq

TODO: This is more LMIC friendly now. Add. angle about the versatility of the application in small hardware (BRICS). Rethink this statement as Emilyn mentioned Brazil and SA are in UMIC group now.

The design of MTBseq-nf wrapper pipeline

In the design of the MTBseq-nf wrapper pipeline, we relied upon the standardized template from nf-core community that offers a range of benefits such as (i) a directory structure of Nextflow pipelines following community best practices (ii) access to nf-core/configs project to facilitate portability across institutional infrastructures (iii) access to well-tested modules from the nf-core/modules project (iv) a graphical user interface for user who are less familiar with the command line, using the nf-schema project.

Scalability

Allocation of resources. The resource requirements as per the design decisions are proportional to the number of samples which compose a particular analysis cohort with the maximum threads being limited to 8 and recommended minimum memory between 20-25 GB RAM.

Dynamic retries

In the use of MTBseq in a shared computing environment, any execution-time failure due to a sample with bad quality or abrupt cancellation of the process means that the user would need to restart the analysis in a very hands-on manner either using the TBfull step or choosing the individual steps and running them manually.

User-friendliness

Ease of download. In the latest update of the pipeline v1.1.0 (https://github.com/ngs-fzb/MTBseq_source/tree/v1.1.0) the GATK3 jar has been made available directly through the conda channel due to a change in distribution licensing terms the GATK3 program from Broad institute, leading to a simpler setup for the MTBseq pipeline.

The access to the nf-core/modules project is of particular benefit to ease of installation offers integration with container systems (Di Tommaso et al., 2015; Kadri et al., 2022) such as docker or singularity (Kurtzer et al., 2017) and package managers such as conda or spack, facilitating the portability of the pipeline across different institutional infrastructures.

Design principles of using the exact same package/container as of MTBseq and GATK samplesheet. +Soft renaming

In terms of the data movement during the execution of the pipeline, all the samples are required to follow a specific naming convention [SampleID]_[LibID]_[*]_[Direction].f(ast)q.gz and the pipeline requires these samples to be in the directory of invocation, where the relevant outputs from the individual processes are stored, including the log files. This expectation of location of the samples aligns very well with the implementation logic of the pipeline which relies upon the sequential analysis using the "foreach" loops and only initiates the next stage once all the samples have been analyzed through an individual step within the sample-specific analysis.

Reproducibility

Specific software versions (like GATK) can introduce compatibility issues as a result, containerization is the primary focus for ensuring MTBseq-nf's reproducibility across different environments. The exact dependency tree of MTBseq are documented in the corresponding bioconda recipe (TODO add ref) and provided in supplementary material.

Declarative parameters

In bioinformatics analysis, it is often the case that multiple iterations are required to adjust the parameters to the nature of the sequenced samples and epidemiological context, therefore a combination of points-x,y facilitates the generation of results in different folders as opposed to the default behavior of the MTBseq pipeline which requires the pipeline execution directory to be cleaned before reinitiating the analysis with different parameters.

This facilitates easier troubleshooting since the previously generated files are stored within the publish directory. Furthermore, by relying upon the ability of the Nextflow engine to cache computations, it becomes easier for the users to iterate upon the parameters of the pipeline, since Nextflow would not re-execute the processes which have not been affected by the change in parameters or configuration.

Faster diagnosis (possibly) - after extensions Add figures for runtime analysis and also the overall reproducibility.

Maintainability and Extensibility

Extensibility of the pipeline (before/after) main mtbseq steps.

Addition of new modules such as for fastqc and multiqc, which would have required us to manually write and test the modules for these processes.

As a result, to add new steps for FastQC and MultiQC P. Ewels et al. (2016) in the MTBseq-nf pipeline, we downloaded the add-on modules and integrated them prior and posterior to the original MTBseq steps. This is in contrast to the MTBseq design, which requires these modules to be integrated via the "native" modules wrapped by the perl language, followed by extensive integration-level testing which the developer must implement for each module.

Data footprint

Add work directory size comparison

< EXPLAIN >

The MTBSeq-nf pipeline (ref) was built using the Nextflow workflow manager (ref), that facilitates the reproducibility and portable execution of pipeline tasks across heterogeneous computing environments such as laptops, servers, HPC clusters and cloud batch computing systems.

Discussion

(FIXME found a place for these. As a result, to add new steps for FastQC and MultiQC P. Ewels et al. (2016) in the MTBseq-nf pipeline, we downloaded the add-on modules and integrated them prior and posterior to the original MTBseq steps. This is in contrast to the MTBseq design, which requires these modules to be integrated via the "native" modules wrapped by the Perl language, followed by extensive integration-level testing, which the developer must implement for each module. The exact dependency tree of MTBseq is documented in the corresponding bioconda recipe (TODO add ref) and provided in the supplementary material.

TODO Establish equivalence of standard MTBseq and MTBseq-nf (default) mode.

TODO No differences in Statistical reports with MTBseq-nf (parallel)

We hypothesize that the differences are due to the memory-intensive nature of computations carried out by the TBvariants process The task-level isolation introduced by parallel mode ensures that each primitive step has sufficient memory (and CPU) for its computation and as soon as the step completes, the relevant docker container exits and the runtime ensures that the computational resources of the host system are released. Whereas in the TBfull step within MTBseq and MTBseq-nf (default), the consumed memory is not immediately released when the previous process completes as the computations are done sequentially within the same container, therefore the subsequent processes have slightly reduced access to memory, which has an impact upon the memory-intensive computations required by GATK.

NOTE: Explain the differences and highlight what Hanno mentioned "has shared an interesting insight that the differences in the Amend files are mostly in the regions of the genome which are not important for the phylogeny and therefore might not have an impact on the analysis.

The reliance on MTBseq and our decision to use it in an unchanged manner meant that we had to make the trade-off between unchanged MTBseq and further optimizations.

However, the removal of the Perl layer should offer benefits such as the ability to (i) update the (ii) update different versions of the underlying tools involved, this is straightforward due to the integration of nf-core/modules

We have not added iqtree as of now since researchers often generate trees as per the research question therefore, we have not implemented an IQTREE module/process within the MTBseq-nf wrapper.

Conclusions

The MTBseqs-nf wrapper optimizes the MTBseq pipeline for efficient usage of computational resources for genomic "Big Data" analysis for a rapidly increasing number of sequences in the field of Tuberculosis surveillance and diagnostics while improving the user-friendliness, reproducibility and customizability of the original MTBseq pipeline. These improvements imply that the MTBseq-nf is better suited for faster diagnostics and surveillance and is suitable for routine usage in resource-constraint LMIC settings.

This overall decrease in computation time and traceability of analysis implies that the pipeline is better suited for (i) faster diagnostics, and surveillance and (ii) LMIC settings the pipeline can be executed on less powerful hardware due to the fine-grained resource-level constraints.

Acknowledgements

We are grateful to Kohl et al. (2018) for the initiative of publishing MTBSeq pipeline with open source on GitHub.

References

Add

5.1.2 Comprehensive Genomic Surveillance Reveals Transmission Profiles Of Extensively Drug-Resistant Tuberculosis Cases In Pará, Brazil

Esta produção atende aos objetivos específicos: 2, 3 e 4

Este trabalho representa os principais resultados da dissertação após a implementação da extração a partir de amostras advindas de cultura e compreendendo a análise de 40 genomas do CMTB, investigando os 3 primeiros casos de TB-XDR no Pará.

frontiers | Frontiers in Microbiology

TYPE Original Research
PUBLISHED 22 January 2025
DOI 10.3389/fmicb.2024.1514862

Check for updates

OPEN ACCESS

EDITED BY
Axel Cloeckaert,
Institut National de Recherche pour l'Agriculture, l'Alimentation et l'Environnement (INRAE), France

REVIEWED BY
Padmani Sandhu,
Institute of Microbial Technology (CSIR), India
Simone Villa,
University of Milan, Italy
David S. Thaler,
University of Basel, Switzerland

*CORRESPONDENCE
Emilyn Costa Conceição
emilyncosta@gmail.com
Davi Josué Marcon
davijosuemarcon@gmail.com
Karla Valéria Batista Lima
karlavaleria2007@gmail.com

†These authors have contributed equally to this work

‡These authors have contributed equally to this work

RECEIVED 21 October 2024
ACCEPTED 23 December 2024
PUBLISHED 22 January 2025

CITATION
Marcon DJ, Sharma A, Souza AB, Barros RB, Andrade VdGD, Guimarães RJdPS, Lima LNG, Monteiro LHMT, Quaresma AJPG, Ribeiro LR, Suffys PN, Warren RM, Alberio CAA, Lima KVB and Conceição EC (2025) Comprehensive genomic surveillance reveals transmission profiles of extensively drug-resistant tuberculosis cases in Pará, Brazil.
Front. Microbiol. 15:1514862.
doi: 10.3389/fmicb.2024.1514862

COPYRIGHT
© 2025 Marcon, Sharma, Souza, Barros, Andrade, Guimarães, Lima, Monteiro, Quaresma, Ribeiro, Suffys, Warren, Alberio, Lima and Conceição. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Comprehensive genomic surveillance reveals transmission profiles of extensively drug-resistant tuberculosis cases in Pará, Brazil

Davi Josué Marcon[1,2*†], Abhinav Sharma[3†], Alex Brito Souza[1,2], Rafaella Bonfim Barros[4], Valnete das Graças Dantas Andrade[4], Ricardo José de Paula Souza Guimarães[5], Luana Nepomuceno Gondim Lima[1,2], Lúcia Helena Martins Tavares Monteiro[6], Ana Judith Pires Garcia Quaresma[1,2], Layana Rufino Ribeiro[2,7], Philip Noel Suffys[8], Robin Mark Warren[3], Carlos Augusto Abreu Alberio[9‡], Karla Valéria Batista Lima[1,2*‡] and Emilyn Costa Conceição[3*‡]

[1]Programa de Pós-Graduação em Biologia Parasitária na Amazônia, Belém, Brazil, [2]Seção de Bacteriologia do Instituto Evandro Chagas, Ananindeua, Pará, Brazil, [3]South African Medical Research Council Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, [4]Laboratório Central do Estado do Pará, Belém, Pará, Brazil, [5]Seção de Epidemiologia do Instituto Evandro Chagas, Ananindeua, Pará, Brazil, [6]Secretaria de Estado da Saúde do Pará, Belém, Pará, Brazil, [7]Programa de Pós-graduação em Epidemiologia e Vigilância em Saúde, Ananindeua, Brazil, [8]Laboratório de Biologia Molecular Aplicada a Micobactérias, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil, [9]Hospital Universitário João de Barros Barreto, Ambulatório de Tuberculose Multiressistente, Universidade Federal do Pará, Belém, Pará, Brazil

Bedaquiline, an antimicrobial used to treat drug-resistant tuberculosis (DR-TB), was introduced in Brazil in October 2021. Monitoring the emergence and transmission of DR-TB is crucial for implementing public health to control the spread of DR strains of *Mycobacterium tuberculosis*. To measure its impact on the multi-drug treatment scheme in the state of Pará, we aimed to conduct genomic surveillance of DR-TB after bedaquiline was introduced in Brazil. Individuals treated for DR-TB between October 2021 and December 2022, in the reference hospital to treat DR-TB cases from the state of Pará, were included in the study. Clinical and bacteriological information was obtained from the National Laboratory Management Environment and the Special TB Treatment Information System. Genomic DNA was extracted from bacterial cultures performed at the Pará Central Laboratory (LACEN-PA). Whole-genome sequencing (WGS) was obtained using Illumina Nextera-XT and NextSeq 550 and genomes were analyzed using the MAGMA and TB-Profiler pipelines interpreted according to the World Health Organization (WHO) mutations catalog 2nd edition. Geoprocessing was performed based on the patient's residences. Cutoffs of 5–12 single nucleotide polymorphisms (SNPs) were used for transmission analysis. From the 103 patients reported as DR-TB, viable cultures were obtained from 67. Forty isolates were selected randomly for WGS. Among these, a mixed infection of *M. tuberculosis* L1 and L4 and a co-infection of *M. tuberculosis* and *Mycobacterium chelonae* were observed. The genotypic drug susceptibility profile of TB stains (39/40) was as follows: sensitive (1/2, 5%), rifampicin mono-resistant (RR) (4/10%), isoniazid mono-resistant (1/2%),

multidrug-resistant (MDR) (21/52%), extensively drug-resistant (XDR) (3/7%), pre-XDR (8/20%), and other (1/2%). Among the 38 isolates of *M. tuberculosis* strains without mixed infection, using a cutoff of 12 SNPs and suggestive of recent TB transmission, 14 (37%) were grouped into five clusters (C1–C5) and included RR (C5), MDR (C3, C4, C5), pre-XDR, and XDR (C2) strains. We recommend greater attention from the regional public health authorities to detect and track resistance to new drugs, especially in areas with pre-XDR and XDR cases. This is the first report on the detection and transmission of XDR-TB in Pará, Brazil, after the recent re-definition of XDR-TB by the WHO in 2021.

KEYWORDS

bedaquiline, drug-resistant tuberculosis, extensively drug-resistant tuberculosis, genomic surveillance, whole-genome sequencing

1 Introduction

Whole genome sequencing (WGS) is rapidly transforming tuberculosis (TB) surveillance, particularly in the context of new drugs like bedaquiline. According to the latest World Health Organization (WHO) definition of drug-resistant tuberculosis (DR-TB), there are different levels of resistance with epidemiological importance: TB only resistant to isoniazid (HR), TB only resistant to rifampicin (RR-TB), multidrug-resistant TB (MDR-TB, resistant at least to isoniazid and rifampicin), pre-extensively drug-resistant TB (pre-XDR-TB, resistant to rifampicin, isoniazid, and any fluoroquinolone), and extensively drug-resistant TB (XDR-TB, resistant to rifampicin, isoniazid, and any fluoroquinolone, and at least one of the newer drugs bedaquiline and linezolid) (WHO, 2021). Introduced in Brazil in 2021 for the treatment of RR, MDR, and pre-XDR TB (Brasil, 2020), the addition of bedaquiline marked a significant advancement in therapy. However, cases of bedaquiline resistance have already emerged in countries such as China, South Africa, Moldova, and Indonesia, highlighting the need for close monitoring when using this and other new drugs (Chesov et al., 2021; Tong et al., 2023; Rukmana et al., 2024).

Traditional methods for diagnosing *Mycobacterium tuberculosis* drug resistance can be slow and limited, especially for new drugs. In this context, WGS offers a significant advantage by providing a comprehensive overview of the genetic profile of a particular TB strain (Cohen et al., 2019; Sonnenkalb et al., 2023). This allows researchers and public health authorities to identify existing and undescribed mutations associated with DR-TB (WHO, 2023).

Furthermore, WGS can reveal TB transmission patterns, along with surveillance of the development and spread of DR-TB strains (Satapathy et al., 2024). Combined with conventional epidemiologic and clinical data, this information can guide targeted public health interventions, ensuring more precise and effective approaches. Additionally, WGS can inform treatment rescue strategies in cases of observed resistance or treatment failure, especially in high-burden TB countries, such as Brazil, which is relevant considering a recent observation that XDR-TB strains are possibly circulating in the state of São Paulo already[1].

In 2023, Pará, a state in Brazil's Northern region, ranked sixth nationally for TB incidence (49.4 cases per 1,000 inhabitants), with a concerning rise in DR-TB in recent years (da Brasil, 2024). Given the context of the new drug introduction in November 2021 for MDR-TB and later for RR-TB, this study aimed to (i) conduct genomic surveillance using WGS on early cases of RR/MDR-TB from Pará treated with bedaquiline and (ii) compare WGS data with the current standard of care (SOC) approach in Brazil, which includes GeneXpert, phenotypic drug susceptibility test (pDST), and line-probe assays (LPA).

2 Materials and methods

2.1 Ethics aspect

This study was approved by the Ethics and Research Committee of the Evandro Chagas Institute (CEP/IEC) CAAE 0821119.4.0000.0019, approval number: 3.950.56.

2.2 Study design and inclusion/exclusion criteria

This retrospective, cross-sectional, and observational study included baseline sputum samples from individuals with pulmonary TB, broadly classified as DR-TB (encompassing RR-TB, isoniazid monoresistant [HR]-TB, and MDR-TB), according to the SOC routine diagnostics (Figure 1). All patients were treated with a bedaquiline-based regimen.

All cryopreserved culture samples, derived from sputum samples that were collected from November 2021 to December 2022, were included in this study following the inclusion criteria: (i) pulmonary TB; (ii) treated or undergoing treatment for RR-TB, HR-TB (if in retreatment or with commodities), or MDR-TB; (iii) residents of the state of Pará at the time of TB diagnosis. The exclusion criteria were as follows: (i) affected exclusively by non-tuberculous mycobacteria (NTM); (ii) *M. tuberculosis* did not remain viable for culture or subculture.

2.3 Drug susceptibility testing and treatment

When a sample tested positive for RR-TB using the Xpert MTB/RIF Ultra assay (Cepheid, Sunnyvale, United States), a second sample

1 http://www.ial.sp.gov.br/resources/insituto-adolfo-lutz/centros-tecnicos/centro-de-bacteriologia/report_resistencia_mtb.pdf

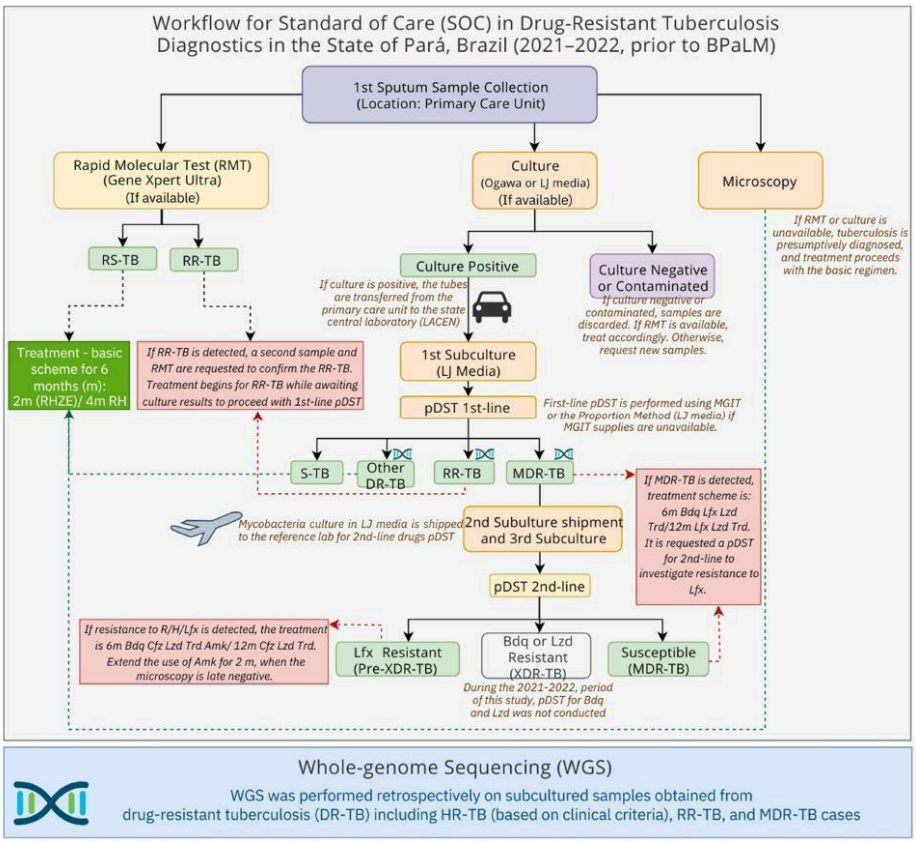


FIGURE 1
Schematic overview of the standard-of-care tuberculosis diagnostic and treatment workflow in Pará State, Brazil (November 2021–December 2022), and sample selection for whole-genome sequencing (WGS). Drug-resistant tuberculosis (DR-TB), Rifampicin-Resistant tuberculosis (RR-TB), Multidrug-resistant tuberculosis (MDR-TB), Phenotypic drug-susceptibility test (pDST), 1st (first)-line drugs: rifampicin, isoniazid, pyrazinamide and ethambutol, 2nd (second)-line drugs: levofloxacin, amikacin, bedaquiline, linezolid. Bedaquiline (Bdq), Linezolid (Lzd), Levofloxacin (Lfx), Susceptible tuberculosis (S-TB), months (m), Rifampicin (R), H (Isoniazid), Pyrazinamide (Z), Ethambutol (E).

was collected to confirm the resistance. This second sputum sample was processed and decontaminated using the *N*-acetyl-L-cysteine sodium hydroxide (NALC/NaOH) protocol, followed by culturing in Löwenstein–Jensen (L-J) medium at the Central Laboratory of Pará (LACEN-PA) (Brasil, 2022).

Further investigation was conducted through first-line pDST for isoniazid, rifampicin, ethambutol, and streptomycin, as performed at LACEN-PA (Brasil, 2022), using either automated mycobacteria growth indicator tubes (MGIT) (Becton, Dickinson and Company, Sparks, USA) or the proportion method. An aliquot of each sample was cryopreserved at −80°C for traceability.

For cases confirmed as MDR-TB, the samples were transported to the National Reference Mycobacteria Laboratory at the Professor Hélio Fraga Reference Center—Fiocruz (Jacarepaguá), Rio de Janeiro. Second-line pDST for fluoroquinolones and aminoglycosides was performed using either the MGIT system or the line probe assay MTBDR-sl (Hain Lifescience, Nehren, Germany). First-line pDST was also conducted on first sputum samples and compared to the results of the RMT when the test was available (Figure 1).

All individuals with samples classified as RR-TB by either Xpert MTB/RIF Ultra and/or pDST for first-line drugs were referred to the University Hospital João de Barros Barreto (HUJBB) at the Federal University of Pará (UFPA), a reference hospital for the treatment of DR-TB in the state of Pará. The treatment scheme followed the Brazilian Ministry of Health guidelines (Brasil, 2019, 2020).

2.4 Data acquisition

The clinical, epidemiological, sociodemographic, and bacteriological data were obtained through partnerships with the LACEN-PA and HUJBB/UFPA. These institutions have access to data from the national automated systems from the Laboratory Environment Manager (GAL)[2] and the Tuberculosis Special Treatment Information System (SITETB)[3]. Bacteriological data were obtained from GAL, including first and second-line pDST. The laboratory results were obtained from standard procedures established by the Brazilian Ministry of Health (Brasil, 2022) used to determine the drug susceptibility pattern in the SOC. The final data collection date was done on 25 July 2024.

2.5 DNA extraction and whole-genome sequencing

Cryopreserved bacteria were subcultured in L-J media, followed by a heat inactivation, and three loops of 10 µL bacterial biomass were resuspended in 200 µL of water, and incubated at 80°C for 1 h using a Vulcan thermoblock (Biancodent, Araucaria, Brazil). The DNA extraction was performed using the MagMAX™ Viral/Pathogen Ultra Nucleic Acid Isolation Kit (Thermo Fisher, Waltham, United States) and purified with Ampure XP magnetic beads (Beckman Coulter, Brea, United States). The DNA quality control (QC) was performed by visualization of the DNA integrity through electrophoresis in 1.5% agarose gel and quantification using the double-strand (ds) DNA High Sensitivity (HS) Qubit Kit (Thermo Fisher, Waltham, United States).

We randomly selected 40 samples for WGS using the online tool https://www.randomizer.org/. These selections were made independently of laboratory results or patient clinical data. The genomic library preparation was performed using Nextera XT Library Preparation Kit (Illumina, San Diego, United States) followed by quality-controlled using Agilent Bioanalyzer 2,100 High Sensitivity Double Strand DNA (Thermo Fisher, Waltham, United States) and quantification using Qubit ds-HS kit. The genomic libraries were pooled and sequenced using the Illumina NextSeq 550 system, Mid-Output Kit, 2 × 150 bp, paired-end, with an approximate 100 times coverage.

2.6 Genomic analysis and geoprocessing

The FASTQ files were processed using the MAGMA pipeline v.2.0.0, covering steps from quality control to drug resistance prediction and phylogenetic analysis. Detailed descriptions of the base call and mapping quality parameters can be found on the MAGMA GitHub page[4] and in the pipeline publications (Heupink et al., 2021, 2023). This pipeline includes representative samples from the 10 human-adapted lineages (L1–L10), as well as the outgroup *Mycobacterium canettii*, facilitating the positioning of study samples for phylogenetic inference analysis.

The phylogenetic analysis was performed as previously described (Heupink et al., 2021, 2023). Briefly, for the phylogenetic and cluster analyses, we concatenated filtered SNPs into a multi-FASTA file. We excluded sites represented by less than 95% of samples, common drug-resistance variants, and rRNA gene variants to avoid phylogenetic bias. SNP distances between sample pairs were calculated, both including and excluding complex regions of the *M. tuberculosis* genome. Cluster identification was performed using ClusterPicker with SNP cutoffs of 5 bp and 12 bp. IQtree was used to determine the most likely substitution model and compute the maximum-likelihood tree, which was visualized with iTOL v6[5] (Figure 2), incorporating drug resistance and lineage markers. The standard MAGMA parameters for establishing cutoffs for median coverage (10), breadth of coverage (0.9), NTM fraction (0.2), site representation (0.95), and strand bias (0.05) were maintained. To avoid the inclusion of a false-positive sublineage 4.9, the relative abundance cutoff was lowered from 0.8 to 0.6 before proceeding with further analysis.

In addition to the lineage prediction, TB-Profiler v6.3.0 (Verboven et al., 2022) was used for genotypic DST (gDST) profiling as part of the MAGMA pipeline, which was interpreted based on the WHO catalog of mutations in *M. tuberculosis* complex (MTBC) and their association with drug resistance, second edition (WHO, 2023). This catalog determines confidence of variants in grades: (1) associated with resistance, (2) associated with resistance—interim, (3) uncertain significance, (4) not associated with resistance—interim, and (5) not associated with resistance. Only variants detected by MAGMA with either high or low frequency (detection limit ≥3%) classified by WHO as associated with resistance, including the interim, were used to classify samples as *M. tuberculosis* drug-resistant. The high-frequency variants (major) were obtained from XBS variant calling (Heupink et al., 2021) and low-frequency variants (minor) using lofreq variant caller optimized with higher sensitivity to the detection of variants occurring in <0.05% of a population (Wilm et al., 2012).

The transmission network analysis was performed using a ≤5 and ≤12 single nucleotide polymorphisms (SNPs) cutoff, which was calculated by MAGMA and submitted to the transcluster R package (Stimson et al., 2019) for prediction based on the first TB symptoms date obtained from SITE-TB. The spoligotype profile was predicted using SpoTyping v2.0[6] and submitted to the SITVIT2 database[7] for Spoligotyping International Type (SIT) determination. The spatial distribution was performed using ArcGIS software[8] considering the patient's residence address.

3 Results

From November 2021 to December 2022, 103 individuals diagnosed with RR-TB initiated treatment that included bedaquiline. However, among these, 36 patients were excluded from this study due to the following reasons: 21 (20.39%) samples were not localized in the cryopreservation isolate collection, 14

2 http://gal.saude.pa.gov.br/
3 http://sitetb.saude.gov.br/
4 https://github.com/TORCH-Consortium/MAGMA/releases/tag/v2.0.0
5 https://itol.embl.de/
6 https://github.com/xiaeryu/SpoTyping-v2.0
7 http://www.pasteur-guadeloupe.fr:8081/SITVIT2
8 https://www.arcgis.com/

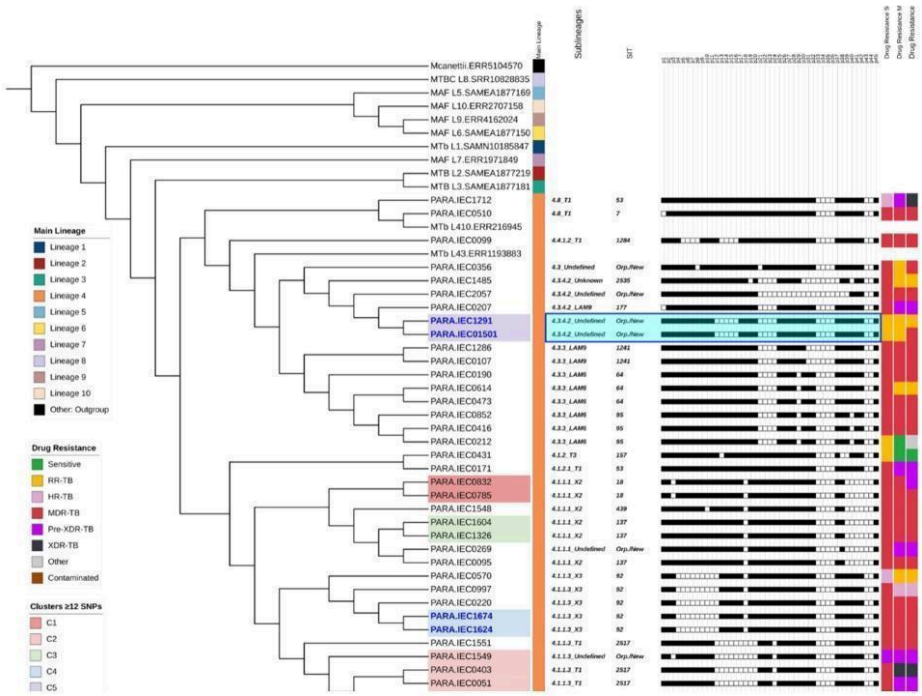


FIGURE 2
Phylogenetic tree based on Whole-Genome Sequencing (WGS), illustrating strain correlations and their respective profiles added as tree annotations after each sample name within the following order: (1) predicted lineages and sublineages, (2) Shared International Typing (SIT), and (3) spoligotyping profiles for the 43 spacers. The figure also presents drug susceptibility status based on (4) standard of care (SOC), and WGS (5) major variants, and (6) WGS (combined major and minor variants combined). The highlighted clades represent five clusters (C1–C5), identified using a ≤12 Single Nucleotide Polymorphism (SNP) threshold, while sample names highlighted in deep blue indicate strains with distances of ≤5 SNPs. The light blue highlight of C5 spoligotype pattern demonstrates a potential new SIT.

(13.59%) samples did not yield viable bacilli upon culturing and individuals had been treated based on Xpert MTB/RIF Ultra results, and one sample (0.97%) was negative for *M. tuberculosis* with a presumptive diagnosis of NTM. Of the remaining 67 samples, 40 were randomly selected for WGS. All patients included in the present study ended the treatments in 1 year and 6 months.

3.1 Lineage description

From 40 samples, 39 were predicted as members of MTBC, sample IEC1387 was identified as a mixed infection with 98% mapped reads to *M. chelonae-abscessus* complex species and 0.4% of reads aligned to the reference genome (*M. tuberculosis*, H37Rv), based on that, it was analyzed separately on downstream analysis. Sample IEC2061 was identified as a polyclonal infection with mixed lineage caused by lineages 4.3.3 and 1.2.2.2 with frequencies of 46 and 50%, respectively. The 38 remaining samples were classified as lineage 4, with clades 4.1.1.3 (12/38; 31%), 4.3.3 (8/38; 21%), 4.1.1.1 (7/38; 18%), 4.3.4.2 (5/38; 13%), 4.8 (2/38; 5%) and one representative from each of the clades 4.1.2, 4.1.2.1, 4.3, and 4.4.1.2.

Among the 38 samples, spoligotyping identified six sublineages: T1 (10/38; 26%), LAM6 (6/38; 16%), X2 (6/38; 16%), X3 (5/38; 13%), LAM9 (3/38; 8%), and T3 (1/38; 2.63%). The remaining samples were classified as "unknown" or "undefined" sublineages. The most frequent spoligotype SIT were 2,517 (6/38; 16%), 92 (5/38; 13%), 64 (3/38; 7%), 95 (3/38; 7%), and 137 (3/38; 7%) (Supplementary Table S1). Six samples (16%) were classified either as new or orphan spoligotypes due to the lack of matches with the SITVIT2 database. The two samples (IEC1291 and IEC01501) were predicted as lineage 4.3.4.2 and had a single SNP difference and spoligotype pattern suggestive of being of a new sublineage. The phylogenetic tree (Figure 2) provides a detailed overview of the lineages, sublineages, spoligotyping, clustering, and drug resistance results.

3.2 Drug resistance descriptive analysis

Figure 3 presents a comparison of the drug susceptibility profiles of the 39 *M. tuberculosis* strains, as determined by SOC and WGS

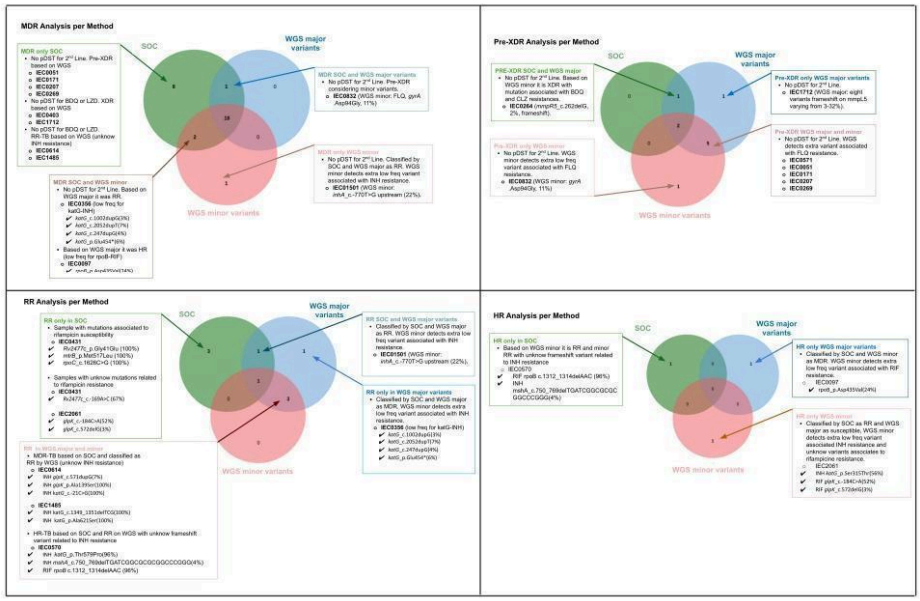


FIGURE 3
Comparative analysis of drug resistance profiles based on Standard of Care (SOC) and Whole-genome Sequencing (WGS), considering major and minor variants for each classification: multidrug-resistant (MDR), pre-extensively drug-resistant (Pre-XDR), rifampicin-resistant only (RR), and isoniazid-resistant only (HR). For samples with discordance across the three approaches, detailed explanations are provided for the discrepancies.

(including major and minor variants). The figure highlights the concordances and discordances between the results obtained using these two approaches for MDR, pre-XDR, RR, and HR classifications. Other information regarding major and minor variants associated with resistance is available in Supplementary Table S2.

Based on the SOC results obtained from the 39 *M. tuberculosis* strains, the susceptibility profile for first-line antibiotics was as follows: 31 were MDR, five were RR, two were HR, and one sample was inconclusive due to culture contamination in both MGIT (SIRE Kit) and LJ (Proportion Method) pDST.

Based on the genotypic WGS profile, rifampicin resistance was associated with several *rpoB* variants: p.Asp435Val (3/35 8%), p.Gln432Pro (2/35 6%), p.His445Leu (1/35 3%), p.Leu430Pro (1/35 3%), and p.Ser450Leu (23/35 66%). While isoniazid resistance was associated with the high-frequency mutations *inhA*_c.-777C > T (4/31 12, 90%) and *katG*_p.Ser315Thr (26/31 84%). Fluoroquinolone resistance was associated with *gyrA*_p.Ala90Val (1/11, 9%), *gyrA*_p.Asp94Ala (1/11, 9%), *gyrA*_p.Asp94Asn (6/11, 54%), *gyrA*_p.Asp94Gly (3/11, 27%), ethambutol resistance was associated with mutations *embB*_p.Met306Ile (9/14, 64%) and *embB*_p.Gly406Ser (5/14, 36%). Pyrazinamide resistance was linked to multiple pncA mutations, including *pncA*_c.464dupT (7/21, 33%), *pncA*_p.Gln122* (3/21, 14%), pncA c.-11A > G (2/21, 9%), and *pncA*_p.Val128Gly (2/21, 9%). Streptomycin resistance was identified with *gid*_c.311delT (3/7, 57%), *rpsL*_p.Lys88Arg (2/7, 43%), *gid*_c.101dupG (1/7, 14%) and *gid*_c.102delG (1/7, 14%). We did not observe resistance associated with amikacin, kanamycin, and capreomycin.

We detected mutation in genes associated with resistance to new drugs such as bedaquiline, linezolid, clofazimine, delamanid, and pretomanid. For linezolid, the *rplC* p.Cys154Arg (100%) mutation was observed in sample IEC0403. Delamanid and pretomanid resistance was associated with the low-frequency mutations on genes *fbiB* and *fbiC* (*fbiB* c.83_105delATCTGAGCGCCGCCGTCGCCGCG *fbiB* c.381delC and *fbiC* p.Glu670*) being all of them with 2% frequency in samples such as IEC1549, IEC0852, IEC1674, and IEC0264.

Bedaquiline and clofazimine resistance was associated with different low-frequency variants (<40%). For bedaquiline, the detected mutations were *mmpR5* c.262delG (3%) in sample IEC0264, *mmpR5* c.386delC (2%) in sample IEC1485, and multiple mutations in *mmpR5* in sample IEC1712, including c.144dupC (32%), c.141_142dupTC (17%), c.424dupC (7%), c.321dupC (10%), c.139dupG (6%), c.140dupA (3%), c.270dupC (3%), and c.320_321dupGC (3%). The detailed clinical and genomic information about the three cases with bedaquiline resistance detected based on WGS are described in Table 1. When comparing the SOC results to WGS data for the 39 strains, 22 (56.41%) showed agreement using both WGS high-frequency variant and low-frequency detection analysis. Considering only high-frequency variants, 24 (61.54%) samples were concordant. For concordance based solely on lower frequency analysis, 23 (58.97%) agreed (Figure 4).

TABLE 1 Clinical and genomic data from three *Mycobacterium tuberculosis* strains exhibiting low-frequency mutations related to possible bedaquiline resistance, as determined by Whole-Genome Sequencing (WGS).

Sample ID		IEC0264	IEC1485	IEC1712
Lineage		4.1.1.3 (T1, SIT* 2517)	4.3.4.2 (U, SIT* 2535)	4.8 (T1, SIT* 53)
Type of diagnosis		New case	Return after lost to follow-up	Return after lost to follow-up
Treatment comments		Bedaquiline was removed from treatment because of adverse effects	–	Irregularity
Outcome		Lost to follow-up	Cured	Dead
Current Standard of Diagnostic		Pre-XDR-TB	MDR-TB	MDR-TB
WGS (considering major variants)		Pre-XDR-TB	MDR-TB	Pre-XDR-TB
WGS (considering minor variants)		XDR-TB	MDR-TB	XDR-TB
WGS final interpretation		XDR-TB (bedaquiline resistance detection in 3% of variants detection)	MDR-TB (bedaquiline resistance detection in 2% of variants detection)	XDR-TB (bedaquiline resistance detection at least 32% of variants detection)
Major variants associated with the following anti-TB drugs	Rifampicin	*rpoB* p.Ser450Leu (100%)	*rpoB p.Ser450Leu* (100%)	*rpoB p.Leu430Pro* (100%)
	Isoniazid	*katG* p.Ser315Thr (99%)	Frameshift mutations *katG_c.1349_1351delTCG* (100%) and *katG_p.Ala621Ser* (100%) – unknown relevance[a]	*katG_c.-343G > T* (3%), *katG_c.-697 T > A* (6%), *katG_p.Gln295Pro* (100%) – unknown relevance[a]
	Ethambutol	*embB* p.Gly406Ser (100%)	–	–
	Pyrazinamide	*pncA* c.464dupT (100%)	–	*pncA* p.Val139Gly (100%)
	Fluoroquinolone	*gyrA* p.Asp94Asn (100%)	–	*gyrA* p.Asp94Ala (100%)
	Ethionamide	*ethA* c.935dupT (100%)	–	–
Minor variants associated with the following new anti-TB drugs	Bedaquiline and clofazimine	mmpR5 c.262delG (3%)	*mmpR5* c.386delC (2%)	*mmpR5* [c.139dupG (6%), c.140dupA (3%), c.141_142dupTC (17%), c.144dupC (32%), c.270dupC (3%), c.320_321dupGC (3%), c.321dupC (10%), c.424dupC (7%)]
	Delamanid and pretomanid	*fbiB* c.83_105delATCTGAGCGCCGCC GTCGCCGCG (2%)	–	–

*Spoligotyping International Type (SIT).
[a]Mutations with unknown resistance association by WHO Mutations Catalog, thus they are not considered for the status interpretation as isoniazid resistance.

A total of 17 (41.02%) discordances were identified between SOC and WGS results considering both high-frequency and low-frequency variants. Among these, five (31.25%) were attributed to the absence of second-line testing in the SOC and isolates classified as MDR by pDST were reclassified as pre-XDR or XDR by gDST. The remaining 11 (68.75%) discrepancies were linked to lower frequency variants involving rifampicin, isoniazid, or fluoroquinolones (n = 4; 25%). Additionally, pre-XDR strains were reclassified as XDR due to the detection of mutations associated with resistance to new drugs (n = 2; 5%). Other discrepancies included unknown mutations related to isoniazid resistance (n = 2; 5%), inconclusive phenotypic tests (n = 1; 2%), one sample (n = 1; 2%) with discordant results due to inconclusive SOC, and one sample (2.5%) showing phenotypic RR while presenting a genomic variant associated with rifampicin susceptibility.

The mutations with unknown status of resistance to isoniazid were observed in three strains phenotypically resistant to isoniazid (IEC1485, IEC0614, and IEC1712). The mutations and frequency were *glpK*_p.Ala139Ser (100%), *katG*_c.-21C > G (100%), *katG*_c.-683 T > A (7%), *glpK*_c.571dupG (7%) in strain IEC0614, *katG*_c.1349_1351delTCG (100%), *katG*_p.Ala621Ser (100%) in sample 1,485, and *Rv2752c*_p.Leu333Pro (100%), *Rv2752c*_p.Thr329Pro (100%), *ahpC*_c.-52C > T (100%), *katG*_c.-343G > T (3%), *katG*_c.-697 T > A (6%), *katG*_p.Gln295Pro (100%), *dnaA*_p.Pro124Leu (100%), and *Rv1258c*_c.148dupG (1%) in sample IEC1712.

3.3 Tuberculosis transmission analysis

Among the 39 *M. tuberculosis* strains, 14 (35.9%) belonged to five clusters (C1–C5) considering the ≤12 SNP difference cutoff, composed of one cluster with six isolates and four with two isolates each (Figure 2). When increasing stringency for cluster definition to the ≤5 SNP threshold level, six strains remained clustered within three clusters considering ≤5 SNPs cutoff, which were C3 (containing two pre-XDR strains), C4 (with two MDR strains), and C5 (with one RR and one MDR strain).

Based on the transmission network analysis, the *M. tuberculosis* strain IEC0403 was predicted as the initial case of the main transmission chain (C2). Additionally, it was predicted as associated

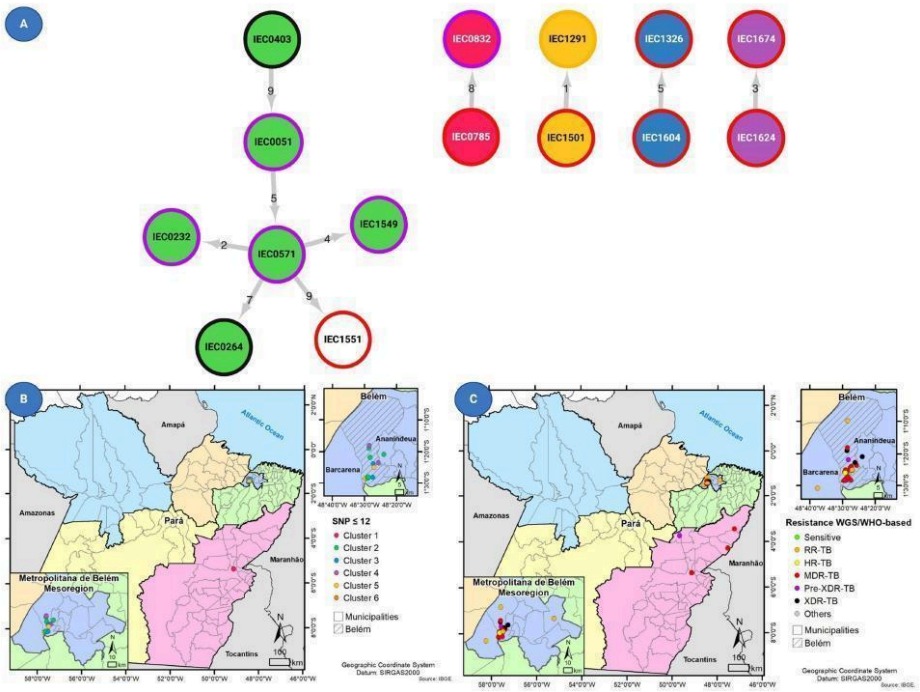


FIGURE 4
(A) Transmission network prediction with circle internal colors based on clusters and border color representing drug resistance status based on Whole-genome sequencing (WGS). In addition, the figure shows the geographical distribution of 14 patients with *Mycobacterium tuberculosis* isolates with **(B)** clustered genomes as defined using MAGMA and of 14 patients with isolates and **(C)** drug susceptibility profiling for all 39 tuberculosis cases.

with the transmission event to IEC0051 distant by five SNP from sample IEC0571. Sample IEC0571 is associated with at least four transmission events, being two with close contact (≤5 SNP) and two distant contacts (≤12 SNP); those events were associated with the transmission of a pre-XDR isolate containing the same resistance-associated variants (*rpoB* p.Ser450Leu, *katG* p.Ser315Thr, and *gyrA* p.Asp94Asn), all of them occurring in the metropolitan region of Belém and identified as part of SIT 2517.

Strains within C4 shared less than five SNPs of distance, with only one shared resistance variant (*rpoB* p.Ser450Leu) associated with RR. Both strains were identified as monoresistant by SOC and WGS (major variants analysis). Similarly, C4 contained two MDR strains sharing *rpoB* (p.Asp435Val) and *katG* (p.Ser315Thr). One distant transmission event (C1) was observed outside the metropolitan region of Pará, with two cases from Marabá, which were identified by the same medical facility, sharing *rpoB* (p.Gln432Pro), *katG* (p.Ser315Thr), and *embB* (p.Met306Ile) mutations.

4 Discussion

This study presents a comprehensive analysis of the first published XDR and pre-XDR-TB cases identified in Pará, Brazil, using WGS, and based on the latest WHO drug resistance definition (WHO, 2021). By examining baseline samples from an initial cohort of RR and/or MDR-TB patients treated with a bedaquiline-based regimen, we were able to compare SOC data with WGS findings, investigate transmission dynamics, and characterize the genetic diversity of circulating *M. tuberculosis* strains.

Our genomic analysis revealed multiple transmission clusters, including strains with mixed infections, one involving different *M. tuberculosis* lineages and co-infection with NTM. This is consistent with reports from high-burden TB countries such as China and South Africa (Tong et al., 2023; Rukmana et al., 2024), where WGS has similarly revealed emerging resistance patterns. However, the identification of XDR-TB transmission in Pará, Brazil, represents a novel finding, underscoring the critical need for enhanced surveillance in the region.

In this study, all *M. tuberculosis* strains belonged to lineage 4, the most generalist and predominant lineage in Latin America. Within lineage 4, the spoligotype T1 (SIT 2517) was reported in a larger cluster of five strains carrying mutations associated with MDR-TB. The SIT 2517 spoligotype is uncommon, having been previously reported only in Pará, Brazil, and linked to MDR-TB in a cohort of strains isolated between 2000 and 2010 (Conceição et al., 2017, 2024). Our genomic surveillance data further support the association between SIT 2517 and MDR-TB, suggesting that this genotype has been

circulating for at least 21 years, as observed in this study. Recent strains with additional mutations conferring fluoroquinolone resistance indicate the ongoing evolution of drug resistance toward pre-XDR status. In addition to the observed frequency of the rare sublineage T1 (SIT 2517), we identified potential new sublineages within clade lineage 4.2.4.2. These sublineages shared the octal spoligopattern "777,603,607,760,771" in two samples. This lineage is closely related to LAM9 SIT 177 (sample IEC0207), which differs in the presence of spacer 1 and the absence of spacers from 12 to 16. Future validation of this genotype signature and deposition of these novel spoligotypes in the SITVIT2 database will be crucial for their accurate classification and further understanding.

The state of Pará, Brazil, is the second largest state in the country in terms of area, and it covers approximately 1.2 million square kilometers (460,000 square miles) (IBGE, 2021). This vast territory is primarily made up of the Amazon rainforest. The majority of DR-TB cases in this study were found in Belém, but we observed evidence of transmission between Belém and Marabá, two cities located 440 km apart, as indicated by genetic similarities of less than 12 SNPs. In general, of the 39 *M. tuberculosis* strains, 20 strains were mainly divided into close clusters harboring 14 strains that formed five clusters with a ≤12 SNP difference and the stricter clusters harbored six strains of these being close clusters with a ≤5 SNP difference.

From a public health standpoint, priority should be given to closely monitoring individuals associated with the strain IEC0403, identified as the initial case in the main transmission chain. This strain was classified as XDR-TB with linezolid resistance conferred by the *rplC* p.Cys154Arg mutation at 100% frequency. Additionally, the strain IEC0571, linked to at least four transmission events, including two close contacts and two distant contacts, warrants increased surveillance. In particular, all patients infected by strains belonging to genotype SIT 2517 should be prioritized for contact tracing and further investigation. These measures align with the emerging field of Precision Public Health, which utilizes genomic, social, behavioral, and environmental data, and the application of artificial intelligence to TB control programs to enhance disease prevention and control (Roberts et al., 2024).

Strains IEC1387 and IEC2061 were identified as mixed infections through WGS, which are well documented in literature (Kargarpour Kamakoli et al., 2017; Pandey et al., 2020; Moreno-Molina et al., 2021) and can involve either multiple strains of *M. tuberculosis* (within the same or different lineages) or co-infection with NTM (Bjerrum et al., 2016; Huang et al., 2022). Distinguishing between these scenarios can be challenging, as standard diagnostic methods have limitations. Although the GenoType *Mycobacterium* CM (Hain LifeScience, Nehren, Germany) can sometimes identify mixed infections, WGS remains the most reliable method for precise differentiation, enabling the simultaneous detection and characterization of both NTM species and sublineages. Further strain analysis is essential to accurately determine its drug resistance profile.

During the treatment course of an individual from which sample IEC1387 was isolated, culture reversion after successive negative cultures was observed in treatment month 8, molecular identification-based Sanger sequencing of ribosomal RNA 16 s, genes *hsp65* and *rpoB* identified the culture isolate as part of *Mycobacterium abscessus–chelonae* complex. Based on GeneXpert initial result as rifampicin-resistant TB and the later isolation and identification of *M. abscessus–chelonae* complex, associated with the WGS analysis on which both *M. tuberculosis* and *M. chelonae* (part of *M. abscess–chelonae* complex) were identified, we were able to identify NTM-TB mixed infection. The responsible clinician team kept the ongoing treatment, and after the complete treatment with successive negative cultures and good clinical evolution, the final treatment outcome was successful. This success is aligned with MIC assays demonstrating the effectivity bedaquiline in *M. abscessus* strains (Brown-Elliott and Wallace, 2019), the effectivity of clofazimine *in vitro* (Hajikhani et al., 2021), and also with clinical study incorporating linezolid in *M. abscessus* treatment (Cheng et al., 2023). In this case, the standard regimen proposed for TB treatment composed of bedaquiline, linezolid, levofloxacin, and clofazimine, had good results.

When comparing the WGS gDST *versus* SOC we observed discrepancies, which were largely attributed to the absence of second-line drug testing in many cases for patients as part of SOC. Despite the recommendation from the Brazilian Health Ministry for the application of these tests should be conducted in all cases when MDR is detected in first-line tests (Brasil, 2021, 2022), there were operational challenges within the flow of the diagnostic algorithm protocol. For example, if the state-level laboratory (LACEN-PA) does not offer pDST for second-line drugs or the molecular gDST test LPA (MTBDRsl), specimens should be sent to the regional reference laboratory, located in another state, Amazonas (LACEN-AM). Additionally, these tests are not always available at the regional laboratory, and samples must then be forwarded to the National reference laboratory in Rio de Janeiro. This process can significantly increase the time required to obtain diagnostic results and decrease the chances of obtaining viable bacteria. Furthermore, the phenotypic drug pDST panel utilized between November 2021 and December 2022 did not encompass bedaquiline or linezolid.

The high rate of absence of second-line testing absence is expected to be reduced with the recent expansion of the second-line LPA (MTBDRsl) into the routine of LACEN laboratories (da Brasil, 2023), decreasing as such the need to send samples to the regional or national reference center and accelerating the diagnostic process. However, it is crucial to note that LPA has limitations, such as the lack of detection of mutations outside the regions targeted by the probes, and the impact of bacillary load on the test performance (Barnard et al., 2020; World Health Organization and Others, 2022). Moreover, studies have shown that LPA failed to detect isoniazid resistance in 33% of isolates due to the presence of low resistance mutations in *katG, inhA*, and the *inhA* promoter (Brossier et al., 2006), or to detect rifampicin resistance on sputum samples with negative smear compared to the Xpert MTB/RIF (Yadav et al., 2021).

In this context, it is important to perform minimum inhibitory concentration (MIC) for second-line and new drugs in at least some laboratory routines in each of the five main Brazilian regions, due to LPA limitations. The performance of MIC assay will allow the detection of resistance to the new drugs and also, to detect resistances, that uncovered by LPA providing important information for the description of and also unknown patterns of resistance, only patterns, normally detected by phenotypic testing methods that could be further investigated and elucidated by WGS (Nimmo et al., 2024).

Other discrepancies between WGS and SOC were mainly associated with low-frequency variants related to the development of resistance in subpopulations of the main strain and the evolution of the resistance pattern within the host. Detecting those variants is an important step in the MAGMA pipeline, allowing us to detect sub-populations within the

main strain (Heupink et al., 2021, 2023). This phenomenon is often associated with happening during the emergence of resistance; those small populations and such subpopulations might occur in a frequency even below the detection threshold of culture-based methods (Cohen et al., 2019).

Accurate variant calling and critical interpretation are essential for understanding the link between mutations and drug resistance, particularly for low-frequency variants. These variants, often associated with small resistant populations or sequencing errors, may not currently correlate with phenotypic resistance. However, a recent study (de Vos et al., 2019) highlights the potential for low-frequency variants to become dominant under antibiotic pressure, as seen in patients with treatment lost to follow-up or relapse. For patients treated with bedaquiline-based therapies, especially those lost to follow-up (like IEC0264 with detection of low frequency of resistance-related mutation), WGS is crucial to assess the acquisition of additional variants and to formulate a treatment scheme. In our study, we proposed to disregard low-frequency mutations that occur in less than 3% of the population, such as the *mmpR5* c.386delC variant (0.02) that occurred in the strain IEC1485, where a person completed bedaquiline-based treatment with a positive treatment outcome of cure.

Finally, the third category of discrepancies was associated with mutations of unknown association to resistance by WHO, as observed in three strains: IEC1485, IEC0614, and IEC1712, with phenotypic resistance to isoniazid. Further studies are needed to investigate the impact of these variants on isoniazid resistance development. These variants were *katG*_c.1349_1351delTCG, *katG*_p.Ala621Ser, *katG*_c.-683 T > A, *glpK*_c.571dupG, *katG*_c.-343G > T, and *katG*_c.-697 T > A, which should be confirmed with Sanger sequencing and MIC and analyzed by protein modeling.

The results from WGS provide critical insights that extend beyond the scope of SOC diagnostics. While SOC primarily focuses on first-line drugs and selected second-line agents, WGS offers a more comprehensive drug resistance profile, including the detection of mutations associated with newer antimicrobials such as bedaquiline, linezolid, clofazimine, and delamanid. In this study, WGS identified low-frequency variants linked to resistance against these drugs, which were not detectable through SOC testing. For instance, bedaquiline resistance mutations were detected in multiple strains, underscoring the importance of WGS in identifying emerging resistance patterns that are critical for optimizing treatment regimens. Furthermore, WGS can detect resistance mutations for drugs not routinely tested in SOC workflows due to logistical or technical limitations, providing clinicians with a broader understanding of potential therapeutic options. This enhanced diagnostic capability of WGS ensures a more tailored treatment approach, ultimately improving patient outcomes and aiding in the containment of DR-TB transmission. These findings highlight the indispensable role of WGS in complementing SOC diagnostics and underscore its potential to revolutionize the management of DR-TB in high-burden settings.

Our findings highlight the urgent need for enhanced genomic surveillance and targeted public health interventions in Pará. The emergence of XDR-TB strains, particularly in urban centers like Belém, poses a significant challenge to TB control efforts. The use of WGS as a routine diagnostic tool could provide valuable insights into resistance mechanisms and transmission pathways, enabling more effective and tailored treatment regimens (Nimmo et al., 2024).

Real-time WGS enables the rapid detection of mutations linked to transmission chain reconstruction, aiding decision-makers in monitoring, understanding, and preventing the spread of bedaquiline-resistant and other new drug-resistant strains (Larsson et al., 2024). Investing in WGS-based surveillance, with a focus on cost reduction and effective long-term analysis, can help maintain robust drug regimens such as bedaquiline–pretomanid–linezolid–moxifloxacin (BPaLM) (WHO, 2022). The costs of WGS should be weighed against the significant investment required for new drug discovery, which ranges from $161 million to $4.54 billion (Schlander et al., 2021). Implementing WGS for diagnostics and surveillance in Brazil could particularly help reduce the lengthy delays in second-line testing, facilitating the prediction of resistance and transmission chains. Additionally, we suggest that future microbiome research should consider analyzing samples without NALC/NAOH decontamination to better understand the role of non-mycobacterial infections, including viruses other than HIV, in patient outcomes. Our current study used the standard diagnostic flow, and all patients were HIV-negative.

While this study offers valuable insights into XDR-TB emergence and transmission, its reliance on LPA, MGIT, and proportions methods, without MIC testing, especially for second-line drugs including bedaquiline and linezolid, limits the generalizability of the comparison to WGS analyzed with the WHO catalog. To gain a more comprehensive understanding of XDR-TB transmission dynamics in Brazil, future studies should be performed, increasing not only the sample size, including a broader geographic period but also the study area, and conducting longitudinal analyses to better assess the impact of bedaquiline and other new drugs on the resistance the emergence of DR in the country. In Brazil's current diagnostic workflow, sputum samples are decontaminated and cultured in Ogawa Kudoh media at frontier laboratories, with positive cultures sent to state LACENs for further analysis. Implementing real-time *M. tuberculosis* WGS at each of Brazil's 27 state LACENs could significantly enhance genomic surveillance, leveraging existing WGS infrastructure from the COVID-19 pandemic.

5 Conclusion

This study provides important insights into the emergence, transmission dynamics, genetic diversity, and resistance profiles of the first XDR-TB cases in Pará, Brazil, within the cohort of RR/MDR-TB patients treated with bedaquiline-based regimens. By integrating SOC and WGS analyses, we identified multiple transmission clusters, including mixed infections with different *M. tuberculosis* lineages and co-infections with NTM. We demonstrated the importance of genomic surveillance in detecting and monitoring emerging drug-resistant TB strains, especially in high-burden areas such as Belém. Identifying transmission chains associated, with an example being the evolution of the local prevalent strain with SIT 2517 to pre-XDR and XDR belonging to SIT 2517 (an endemic SIT isolated only in Pará) enforces the importance of continuous genomic surveillance to identify the emergence of resistance and its spread. Showcasing the use of WGS in the detection of resistance-associated variants and transmission chains, particularly in resource-limited settings like Pará, where phenotypic testing for new drugs remains limited. The discrepancies between SOC and WGS results, almost surely due to the absence of second-line testing, show critical gaps in current diagnostic protocols that could be reduced by the implementation of WGS in

laboratory routines. The implementation of WGS as a routine tool could mitigate these delays, providing timely insights into resistance mechanisms and enabling more effective regimen prescriptions. The interpretation of minor variants detection (low frequency) should be further investigated and carefully performed to guide patient care.

Data availability statement

The datasets presented in this study can be found in online repository NCBI under the Bioproject ID PRJNA1135144. The names of the accession number(s) can be found in the article/Supplementary material.

Ethics statement

The studies involving humans were approved by Ethics and Research Committee of the Evandro Chagas Institute (CEP/IEC) CAAE 0821119.4.0000.0019, approval number: 3.950.56. The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from a by-product of routine care or industry. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

Author contributions

DM: Data curation, Formal analysis, Investigation, Methodology, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing. ASh: Data curation, Funding acquisition, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing, Formal analysis. ASo: Investigation, Writing – review & editing, Methodology. RB: Investigation, Writing – review & editing. VA: Resources, Writing – original draft, Writing – review & editing. RG: Formal analysis, Investigation, Software, Visualization, Writing – review & editing. LL: Resources, Writing – review & editing. LM: Resources, Validation, Writing – original draft, Writing – review & editing. AQ: Resources, Writing – review & editing. LR: Investigation, Writing – review & editing. PS: Validation, Writing – review & editing. RW: Resources, Validation, Writing – original draft, Writing – review & editing. CA: Conceptualization, Data curation, Investigation, Resources, Supervision, Writing – review & editing. KL: Conceptualization, Data curation, Funding acquisition, Methodology, Resources, Supervision, Writing – review & editing. EC: Conceptualization, Data curation, Formal analysis, Methodology, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing, Funding acquisition, Investigation, Project administration, Resources.

Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. Coordination for the Improvement of Higher Education Personnel (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—CAPES), Research Productivity Grants—PQ, CNPq Call No. 09/202, Process 311165/2022–2. We also acknowledge the use of a virtual server for this analysis from Oracle Cloud to conduct this analysis as part of the Oracle for Research Cloud Grant (ID 3083687). This study was also funded by the Brazilian National Council for Scientific and Technological Development (Conselho Nacional de Desenvolvimento Científico e Tecnológico -CNPq) grant process number 445784/2023-7 under the REVIGET N/NE project.

Acknowledgments

We extend our gratitude to the HUJBB/UFPA and LACEN-PA laboratory and clinical teams for their invaluable contributions to daily diagnostic routines and inputs for this project. We also acknowledge the State Department of Health of Pará (SESPA) team for providing crucial insights into the intricacies of TB diagnostics and information flow in the state of Pará. Additionally, we wish to thank the Evandro Chagas Institute (IEC) for providing institutional support for the laboratory work. For the computational infrastructure, we would like to acknowledge the Oracle Cloud Infrastructure credits, which facilitated multiple iterations of the analysis using a hosted cloud server.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The authors declare that Gen AI was used in the creation of this manuscript. Generative AI (Grammarly and Gemini), was used to assist authors with grammar corrections to enhance the manuscript's readability.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmicb.2024.1514862/full#supplementary-material

References

Barnard, M., Gey van Pittius, N. C., van Helden, P. D., Bosman, M., Coetzee, G., and Warren, R. M. (2020). The diagnostic performance of the GenoType MTBDRplus version 2 line probe assay is equivalent to that of the Xpert MTB/RIF assay. *J. Clin. Microbiol.* 50, 3712–3716. doi: 10.1128/jcm.01958-12

Bjerrum, S., Oliver-Commey, J., Kenu, E., Lartey, M., Newman, M. J., Addo, K. K., et al. (2016). Tuberculosis and non-tuberculous mycobacteria among HIV-infected individuals in Ghana. *Trop. Med. Int. Health* 21, 1181–1190. doi: 10.1111/tmi.12749

Brasil (2019). Manual de recomendações para o controle da tuberculose no Brasil. *2nd* Edn. Brasília: Editora MS.

Brasil (2020). Bedaquilina para pacientes com tuberculose resistente à rifampicina, multirresistente ou extensivamente resistente. Brasília. Ministério da Saúde. Available at: https://www.gov.br/conitec/pt-br/midias/consultas/relatorios/2020/relatorio_bedaquilina_tb_rr_mdr_xdr_546_2020_final.pdf

Brasil (2021). NOTA INFORMATIVA Nº 9/2021-CGDR/.DCCI/SVS/MS. Brasília: Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Doenças de Condições Crônicas e Infecções Sexualmente Transmissíveis, Coordenação-Geral de Vigilância das Doenças de Transmissão Respiratória de Condições Crônicas.

Brasil (2022). Manual de Recomendações para o Diagnóstico Laboratorial de Tuberculose e Micobactérias não Tuberculosas de Interesse em Saúde Pública no Brasil. *1st* Edn. Brasília: Ministério da Saúde.

Brossier, F., Veziris, N., Truffot-Pernot, C., Jarlier, V., and Sougakoff, W. (2006). Performance of the genotype MTBDR line probe assay for detection of resistance to rifampin and isoniazid in strains of *Mycobacterium tuberculosis* with low-and high-level resistance. *J. Clin. Microbiol.* 44, 3659–3664. doi: 10.1128/jcm.01054-06

Brown-Elliott, B. A., and Wallace, R. J. (2019). In vitro susceptibility testing of Bedaquiline against *Mycobacterium abscessus* complex. *Antimicrob Agents.* 63:e01919-18. doi: 10.1128/aac.01919-18

Cheng, L. P., Zhang, Q., Lou, H., Shen, X. N., Qu, Q. R., Cao, J., et al. (2023). Effectiveness and safety of regimens containing linezolid for treatment of *Mycobacterium abscessus* pulmonary disease. *Ann. Clin. Microbiol. Antimicrob.* 22:106. doi: 10.1186/s12941-023-00655-2

Chesov, E., Chesov, D., Maurer, F. P., Andres, S., Utpatel, C., Barilar, I., et al. (2021). Emergence of bedaquiline-resistance in a high-burden country of tuberculosis. *Eur. Respir. J.* 57:2002544. doi: 10.1183/13993003.02544-2020

Cohen, K. A., Manson, A. L., Desjardins, C. A., Abeel, T., and Earl, A. M. (2019). Deciphering drug resistance in *Mycobacterium tuberculosis* using whole-genome sequencing: progress, promise, and challenges. *Genome Med.* 11:45. doi: 10.1186/s13073-019-0660-8

Conceição, E. C., Loubser, J., dos Guimarães, A. E. S., Sharma, A., Rutaihwa, L. K., Dippenaar, A., et al. (2024). A genome-focused investigation reveals the emergence of a *Mycobacterium tuberculosis* strain related to multidrug-resistant tuberculosis in the Amazon region of Brazil. *Microorganisms* 12:1817. doi: 10.3390/microorganisms12091817

Conceição, E. C., Rastogi, N., Couvin, D., Lopes, M. L., Furlaneto, I. P., Gomes, H. M., et al. (2017). Genetic diversity of *Mycobacterium tuberculosis* from Pará, Brazil, reveals a higher frequency of ancestral strains than previously reported in South America. *Infect. Genet. Evol.* 56, 62–72. doi: 10.1016/j.meegid.2017.10.021

da Brasil, M. S. (2023). Recomendações Técnicas para Laudo e Interpretação do Teste de Hibridização com Sonda em Linha (Line Probe Assay–LPA) para Tuberculose. Brasília: Ministério da Saúde.

da Brasil, M. S. (2024). Boletim Epidemiológico de Tuberculose-Número Especial. Brasília: Departamento de HIV/Aids, Tuberculose, Hepatites Virais e Infecções Sexualmente Transmissíveis Secretaria de Vigilância em Saúde e Ambiente.

de Vos, M., Ley, S. D., and Cox, H. (2019). Bedaquiline micro-heteroresistance after tuberculosis treatment cessation. *N. Engl. J. Med.* 380, 2178–2180. doi: 10.1056/NEJMc1815121

Hajikhani, B., Nasiri, M. J., Hosseini, S. S., Khalili, F., Karimi-Yazdi, M., Hematian, A., et al. (2021). Clofazimine susceptibility testing of *Mycobacterium avium* complex and *Mycobacterium abscessus*: a meta-analysis study. *J. Global Antimicrob. Resist.* 26, 188–193. doi: 10.1016/j.jgar.2021.06.002

Heupink, T. H., Verboven, L., Sharma, A., Rennie, V., de Fuertes, M. D., Warren, R. M., et al. (2023). The MAGMA pipeline for comprehensive genomic analyses of clinical *Mycobacterium tuberculosis* samples. *PLoS Comput. Biol.* 19:e1011648. doi: 10.1371/journal.pcbi.1011648

Heupink, T. H., Verboven, L., Warren, R. M., and Van Rie, A. (2021). Comprehensive and accurate genetic variant identification from contaminated and low-coverage *Mycobacterium tuberculosis* whole genome sequencing data. *Microb. Genomics* 7:000689. doi: 10.1099/mgen.0.000689

Huang, M., Tan, Y., Zhang, X., Wang, Y., Su, B., Xue, Z., et al. (2022). Effect of mixed infections with Mycobacterium tuberculosis and nontuberculous mycobacteria on diagnosis of multidrug-resistant tuberculosis: a retrospective multicentre study in China. *Infect Drug Resist.* 15, 157–166. doi: 10.2147/IDR.S341817

IBGE (2021). IBGE-Cidades. Available at: https://cidades.ibge.gov.br/brasil/pa/panorama (Accessed January 5, 2023).

Kargarpour Kamakoli, M., Sadegh, H., Farmanfarmaei, G., Masoumi, M., Fateh, A., Javadi, G.. (2017). Evaluation of the impact of polyclonal infection and heteroresistance on treatment of tuberculosis patients. *Sci. Rep.* 7:41410. doi: 10.1038/srep41410

Larsson, L., Corbett, C., Kalmambetova, G., Utpatel, C., Ahmedov, S., Antonenka, U., et al. (2024). Whole-genome sequencing drug susceptibility testing is associated with positive MDR-TB treatment response. *Int. J. Tuberc. Lung Dis.* 28, 494–499. doi: 10.5588/ijtld.24.0052

Moreno-Molina, M., Shubladze, N., Khurtsilava, I., Avaliani, Z., Bablishvili, N., Torres-Puente, M., et al. (2021). Genomic analyses of *Mycobacterium tuberculosis* from human lung resections reveal a high frequency of polyclonal infections. *Nat. Commun.* 12:2716. doi: 10.1038/s41467-021-22705-z

Nimmo, C., Bionghi, N., Cummings, M. J., Perumal, R., Hopson, M., Jubaer, S. A., et al. (2024). Opportunities and limitations of genomics for diagnosing bedaquiline-resistant tuberculosis: a systematic review and individual isolate meta-analysis. *Lancet Microbe* 5, e164–e172. doi: 10.1016/S2666-5247(23)00317-8

Pandey, P., Bhatnagar, A. K., Mohan, A., Sachdeva, K. S., Samantaray, J. C., Guleria, R., et al. (2020). *Mycobacterium tuberculosis* polyclonal infections through treatment and recurrence. *PLoS One* 15:e0237345. doi: 10.1371/journal.pone.0237345

Roberts, M. C., Holt, K. E., Del Fiol, G., Baccarelli, A. A., and Allen, C. G. (2024). Precision public health in the era of genomics and big data. *Nat. Med.* 30, 1865–1873. doi: 10.1038/s41591-024-03098-0

Rukmana, A., Gozali, C., and Erlina, L. (2024). *Mycobacterium tuberculosis* lineage distribution using whole-genome sequencing and Bedaquiline, Clofazimine, and linezolid phenotypic profiles among rifampicin-resistant isolates from West Java, Indonesia. *Int. J. Microbiol.* 2024:2037961. doi: 10.1155/2024/2037961

Satapathy, P., Itumalla, R., Neyazi, A., Nabizai Taraki, A. M., Khatib, M. N., Gaidhane, S., et al. (2024). Emerging bedaquiline resistance: a threat to the global fight against drug-resistant tuberculosis. *J. Biosaf. Biosecurity* 6, 13–15. doi: 10.1016/j.jobb.2024.01.001

Schlander, M., Hernandez-Villafuerte, K., Cheng, C.-Y., Mestre-Ferrandiz, J., and Baumann, M. (2021). How much does it cost to research and develop a new drug? A systematic review and assessment. *PharmacoEconomics* 39, 1243–1269. doi: 10.1007/s40273-021-01065-y

Sonnenkalb, L., Carter, J. J., Spitaleri, A., Iqbal, Z., Hunt, M., Malone, K. M., et al. (2023). Bedaquiline and clofazimine resistance in *Mycobacterium tuberculosis*: an in-vitro and in-silico data analysis. *Lancet Microbe* 4, e358–e368. doi: 10.1016/S2666-5247(23)00002-2

Stimson, J., Gardy, J., Mathema, B., Crudu, V., Cohen, T., and Colijn, C. (2019). Beyond the SNP threshold: identifying outbreak clusters using inferred transmissions. *Mol. Biol. Evol.* 36, 587–603. doi: 10.1093/molbev/msy242

Tong, E., Zhou, Y., Liu, Z., Zhu, Y., Zhang, M., Wu, K., et al. (2023). Bedaquiline resistance and molecular characterization of rifampicin-resistant *Mycobacterium Tuberculosis* isolates in Zhejiang, China. *Infect. Drug Resist.* 16, 6951–6963. doi: 10.2147/IDR.S429003

Verboven, L., Phelan, J., Heupink, T. H., and Van Rie, A. (2022). TBProfiler for automated calling of the association with drug resistance of variants in *Mycobacterium tuberculosis*. *PLoS One* 17:e0279644. doi: 10.1371/journal.pone.0279644

WHO (2021). WHO announces updated definitions of extensively drug-resistant tuberculosis. Available at: https://www.who.int/news/item/27-01-2021-who-announces-updated-definitions-of-extensively-drug-resistant-tuberculosis (Accessed December 24, 2022)

WHO (2022). WHO operational handbook on tuberculosis. Module 4: Treatment – drug-resistant tuberculosis treatment. Geneva: Organização Mundial da Saúde.

WHO (2023). Catalogue of mutations in *Mycobacterium tuberculosis* complex and their association with drug resistance. *2nd* Edn. Geneva: Organização Mundial da Saúde-WHO.

Wilm, A., Aw, P. P. K., Bertrand, D., Yeo, G. H. T., Ong, S. H., Wong, C. H., et al. (2012). LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res.* 40, 11189–11201. doi: 10.1093/nar/gks918

World Health Organization and Others (2022). Line probe assays for detection of drug-resistant tuberculosis: Interpretation and reporting manual for laboratory staff and clinicians: World Health Organization.

Yadav, R. N., Singh, B. K., Sharma, R., Chaubey, J., Sinha, S., and Jorwal, P. (2021). Comparative performance of line probe assay (version 2) and Xpert MTB/RIF assay for early diagnosis of rifampicin-resistant pulmonary tuberculosis. *Tuberc. Respir. Dis.* 84, 237–244. doi: 10.4046/trd.2020.0171

5.1.3 Relato De Caso: Vigilância Genômica Revela Casos De Coinfecção Tb-Mnt E Tb-Tb Em Hospital Terciário De Belém.

Esta produção atende ao objetivo específico 5.

O Relato de caso busca investigar profundamente dois casos de infecção mista detectados pelo estudo principal. Partindo da investigação baseada em WGS para o ponto de vista clínico, relatando a complementaridade do sequenciamento genômico para o manejo de casos de alta complexidade no Hospital Universitário Barros Barreto.

Autores: Davi Josué Marcon, Abhinav Sharma, Alex Brito Souza, Rafaella Bonfim Barros, Valnete das Graças Dantas Andrade, Lúcia Helena Martins Tavares, Ana Judith Pires Garcia Quaresma,Layana Rufino Ribeiro, Robin Mark Warren, Emilyn Costa Conceição, Carlos Augusto Abreu Alberio e Karla Valéria Batista Lima

Título: VIGILÂNCIA GENÔMICA REVELA CASOS DE COINFECÇÃO TB-MNT E TB-TB EM HOSPITAL TERCIÁRIO DE BELÉM.

Introdução:As cepas IEC1387 e IEC2061 isoladas a partir de dois pacientes tratados na referência terciária do estado do Pará, foram identificadas como infecções mistas por meio do sequenciamento completo do genoma, um fenômeno bem documentado na literatura. Essas infecções podem envolver múltiplas cepas de Mycobacterium tuberculosis — pertencentes ao mesmo ou a diferentes linhagens — ou ainda a coinfecção com micobactérias não tuberculosas (NTM). No entanto, diferenciar esses cenários pode ser desafiador, uma vez que os métodos diagnósticos convencionais apresentam limitações. Embora o teste GenoType Mycobacterium CM (Hain LifeScience, Nehren, Alemanha) possa, em alguns casos, detectar infecções mistas, o WGS continua sendo a abordagem mais precisa, permitindo a identificação simultânea e detalhada tanto das espécies de MNT quanto das sublinhagens de M. tuberculosis. Uma análise mais aprofundada dessas cepas é essencial para uma caracterização precisa do perfil de resistência aos fármacos.

Relato de Caso:

Uma paciente de 36 anos, identificada como caso novo de tuberculose em abril de 2022, com apresentação de cavitação unilateral no exame de raio-X.

Cultura realizada em outubro e o TSA é liberado em dezembro de 2022 com a detecção de resistência a estreptomicina e isoniazida e o tratamento prescrito é composto por seis meses de bedaquilina (Bdq), levofloxacino (Lfx), linezolida (Lzd) e terizidona (Trd), seguido de uma fase de manutenção com Lfx, Lzd e Trd. Após o início do tratamento em janeiro de 2023, houve o registro de perda de acompanhamento do caso, não havendo nenhum registro nos sistemas de uma tentativa de resgate de tratamento ou nova notificação.

A vigilância genômica reavaliou a amostra inicial de 2022 em maio de 2024 (Identificada neste estudo pelo ID: IEC2061), detectando a presença simultânea de duas linhagens na amostra, 1.2.2.2 e 4.3.3

Outro paciente do sexo masculino, 46 anos, diabético, apresentou sintomas sugestivos de tuberculose em março de 2022, sendo diagnosticado com tuberculose pulmonar e iniciado o tratamento padrão com rifampicina, isoniazida, pirazinamida e etambutol (RHZE). Dois testes Xpert MTB/RIF realizados em abril detectaram Mycobacterium tuberculosis com resistência à rifampicina. A radiografia de tórax evidenciou cavitação unilateral. Com base nesses achados, o regime terapêutico foi modificado para um esquema composto por Bdq-Lfx-Lzd-Trd suplementado com vitamina B6 por seis meses, seguido de uma fase de manutenção por 12 meses com Lfx, Lzd, Trd e B6.

No terceiro mês de tratamento, foi solicitado um teste de sensibilidade antimicrobiana (TSA), que confirmou resistência fenotípica à isoniazida, rifampicina e etambutol, mas manteve a sensibilidade à estreptomicina. No entanto, um TSA de segunda linha foi rejeitado em janeiro de 2023 devido a resultados inconsistentes. Identificações moleculares realizadas em dezembro de 2022 e maio de 2023 detectaram infecção por Mycobacterium abscessus. Durante todo o curso do tratamento, as baciloscopias permaneceram negativas, mas as culturas de acompanhamento apresentaram variação nos resultados, alternando entre negativas e positivas. Culturas negativas foram registradas do quinto ao oitavo mês de tratamento, tornando-se positiva no nono mês, seguidas novamente de resultados negativos no décimo e décimo primeiro mês, e retornando a positividade nos meses subsequentes até o décimo quinto mês.

Durante o tratamento, no oitavo mês, foi observada uma reversão de cultura após sucessivas culturas negativas. A identificação molecular por sequenciamento de Sanger dos genes do RNA ribossomal 16S, hsp65 e rpoB revelou a presença do complexo M. abscessus–chelonae. Considerando o resultado inicial do GeneXpert, que indicava resistência à rifampicina, e a posterior detecção do complexo M. abscessus–chelonae, associada à análise de sequenciamento genômico (WGS), foi possível confirmar uma infecção mista por micobactéria não tuberculosa (NTM) e M. tuberculosis.

A vigilância genômica reavaliou a amostra inicial de 2022 em maio de 2024 (Identificada neste estudo pelo ID: IEC1387, identificando a presença de M. chelonae, com mutações associadas à resistência a beta-lactâmicos, aminoglicosídeos e rifamicinas. Dessa forma, sugere-se que o paciente tenha inicialmente apresentado uma infecção mista por M. tuberculosis e M. abscessus-chelonae, evoluindo posteriormente para uma infecção por M. abscessus-chelonae. A equipe médica optou por manter o regime terapêutico baseado em bedaquilina, linezolida, levofloxacino e clofazimina.

Discussão:

A infecção policlonal de TB, é descrita pela literatura, em alguns casos pode ser descrita após a detecção de múltiplos clones de TB em sítios removidos cirurgicamente de lesões ativas de TB. Este fato é incomum, e é associado ao desenvolvimento de populações mistas de bacilos, que podem responder de maneira diferente ao tratamento prescrito(García de Viedma *et al.*, 2003, 2005; Moreno-Molina *et al.*, 2021).

A perda de acompanhamento do caso identificado como infecção policlonal deve ser vista com preocupação, e devem ser tomadas medidas para a retomada do tratamento. Uma vez que, a presença de múltiplos bacilos pode levar a ocorrência de heterorresistência e resultados falso-positivos de sensibilidade a antibiótico, e, a não formulação de esquemas mais adequados para o tratamento do caso.

O sucesso no tratamento do caso de TB-MNT demonstra a eficácia do esquema padrão para TB-MDR no controle da infecção. Esse sucesso está alinhado com estudos que demonstram a eficácia *in-vitro* da bedaquilina, clofazimina e linezolida no tratamento de M. abscessus (Brown-Elliott; Wallace, 2019; Hajikhani *et al.*, 2021; Cheng *et al.*, 2023).

5.1.4 Genomic Diversity Of The Rarely Observed Genotype Of The Mycobacterium Tuberculosis Central Asian (Cas) Lineage 3 From North Brazil

Esta produção atende ao objetivo específico 1.

Este trabalho foi publicado no início do mestrado do Discente, representando sua primeira colaboração com a Dra. Karla Lima, iniciada em estágio obrigatório e culminando no início do projeto de pesquisa.

Atende ao objetivo específico 1, pois foram testados os primeiros scripts para análise de genomas do CMTB utilizando a infraestrutura do Instituto Evandro Chagas, os resultados desse estudo mostraram a presença do primeiro caso da linhagem 3 (Centro asiática) no norte do Brasil.

microorganisms

MDPI

Article

Genomic Diversity of the Rarely Observed Genotype of the *Mycobacterium tuberculosis* Central Asian (CAS) Lineage 3 from North Brazil

Emilyn Costa Conceição [1,2,*], Marília Lima da Conceição [2], Davi Josué Marcon [2,3], Johannes Loubser [1], Gabrielly Leite Andrade [2,4], Sandro Patroca da Silva [5], Ana Cecília Ribeiro Cruz [5,6], Abhinav Sharma [7,8], Philip Suffys [9] and Karla Valéria Batista Lima [2,6]

[1] Department of Science and Innovation—National Research Foundation Centre of Excellence for Biomedical Tuberculosis Research, South African Medical Research Council Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Tygerberg, Cape Town P.O. Box 241, South Africa
[2] Seção de Bacteriologia e Micologia, Instituto Evandro Chagas, Ananindeua 67030-000, PA, Brazil
[3] Centro de Genômica e Biologia de Sistemas, Universidade Federal do Pará, Belém 66075-110, PA, Brazil
[4] Departamento de Medicina, Centro Universitário do Pará—CESUPA, Belém 66613-903, PA, Brazil
[5] Seção de Arbovírus e Febres Hemorrágicas, Instituto Evandro Chagas, Ananindeua 67030-000, PA, Brazil
[6] Programa de Pós-graduação Biologia Parasitária na Amazônia, Universidade do Estado do Pará, Belém 66087-670, PA, Brazil
[7] Faculty of Engineering and Technology, Liverpool John Moores University (LJMU), Liverpool L35UX, UK
[8] Seqera Labs, 08005 Bacelona, Spain
[9] Laboratório de Biologia Molecular Aplicada a Micobactéria, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro 21040-900, RJ, Brazil
* Correspondence: emilyncosta@gmail.com

check for updates

Citation: Conceição, E.C.; da Conceição, M.L.; Marcon, D.J.; Loubser, J.; Andrade, G.L.; Silva, S.P.d.; Cruz, A.C.R.; Sharma, A.; Suffys, P.; Lima, K.V.B. Genomic Diversity of the Rarely Observed Genotype of the *Mycobacterium tuberculosis* Central Asian (CAS) Lineage 3 from North Brazil. *Microorganisms* **2023**, *11*, 132. https://doi.org/10.3390/microorganisms11010132

Academic Editor: Laura Rindi

Received: 19 December 2022
Revised: 31 December 2022
Accepted: 3 January 2023
Published: 4 January 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

Abstract: *Mycobacterium tuberculosis* (*Mtb*) Central Asian Strain (CAS) Lineage 3 (L3) genotype is predominantly found in East-Africa, Central-Asia, Western-Asia, and South-Asia; however, a new spoligotyping CAS/SIT2545 was found in northern regions of Brazil. We aimed to characterize and describe the genetic diversity and perform a phylogenetic assessment of this novel genotype. We performed 24-MIRU-VNTR loci and Whole-genome sequencing (WGS) of six Brazilian isolates previously spoligotyped. The libraries were prepared using a Nextera-XT kit and sequenced in a NextSeq 550 Illumina instrument. We performed lineage assignment and genomic characterization. From publicly available genomes of *Mtb* L3 and other lineages, we created a robust dataset to run the MTBSeq pipeline and perform a phylogenetic analysis. MIRU-VNTR and WGS confirmed CAS/SIT2545 belongs to L3. Out of 1691 genomes, 1350 (79.83%) passed in quality control (genomic coverage > 95%). Strain 431 differed in 52 single nucleotide variants (SNV), confirming it does not belong to the same transmission chain. The eight genomes from a global dataset clustered closer to Brazilian strains differed in >52 SNVs. We hypothesized L3 and L1 were introduced in Brazilian Northern in the same historical event; however, there is a need for additional studies exploring the genetic diversity of *Mtb* Brazilian Northern.

Keywords: tuberculosis; *Mycobacterium tuberculosis*; Lineage 3; phylogenetic; whole-genome sequencing; Brazil

1. Introduction

Mycobacterium tuberculosis (*Mtb*) Lineage 3 (L3), associated with the Central Asian Strain (CAS) genotype, is predominantly found in East Africa, Central Asia, Western Asia, and South Asia. Strains with the CAS genotype are likely the ancestors of the Beijing family strains (Lineage 2) as they occur in the same geographical location [1,2].

Mtb L3 (CAS/Delhi) is classified as a "modern" lineage that is more virulent than the ancient lineages [3] and has the distinct characteristic of the TbD1 genome deletion. L3 is associated with major tuberculosis (TB) epidemics largely due to higher transmission rates compared to L1, L5, L6 and L7. Additionally, L3 is metabolically distinct from other lineages because it has an increased abundance of proteins involved in lipid metabolism, the ESX-3 secretion system, the TCA cycle and other metabolic pathways associated with increased fitness and survival capacity [4].

There is compelling evidence that CAS lineage may have spread across Africa and Eurasia through ancient African-Asian trade networks and the routes of the Silk Road early in the first millennium [1]. Molecular data and identifiers from the CAS strains deposited in SITVIT2 largely originate from countries along this route: Afghanistan (3), Belgium (1), Bangladesh (8), Brazil (2), China (3), Ethiopia (7), Great Britain (3), India (100), Iran (4), Iraque (24), Italy (1), Jamaica (1), Myanmar (6), Mozambique (1), Malawi (1), Marrocos (2), The Netherlands (1), Nepal (1), Oman (25), Pakistan (38), Saudi-Arabia (38), South Africa (5), Sudan (4), Somalia (9), Tanzania (7), Uganda (3), United States (15), Yemen (2), and Zambia (4). As L3 is underrepresented in this database, important numbers from medium and high-TB-income countries might be missing, since 189 entries are from an unknown origin.

A large study of *Mtb* based on 2030 L3 genomes from the endemic TB region, excluding strains from low-burden regions, demonstrated that L3 does not contain distinct sublineages. It was therefore proposed that L3 has a good temporal sign with a most recent common ancestor (MRCA) dating between the 2nd and the 13th century AD. Additionally, the authors suggested that *M. tuberculosis* complex (MTBC) studies describing Lineage 1 (L1) and L3 are of great importance since they greatly affect humans yet remain relatively understudied even on the global setting [2].

Northern Brazil has a high incidence of TB, but only a limited number of studies describing the genetic diversity of *Mtb* has come from this region [5]. Despite this, L1, rarely observed in South America, was described and genetically characterized in this region and postulated to have been introduced during with the trans-atlantic slave trade events in Brazil from East-African countries such as Mozambique and Malawi [6–8].

It is worth noting that CAS SIT 2545 (octal 703460000000331) was also found in the State of Pará, Brazil and according to the SITVIT2 database, TB infection with strains of this genotype is rare and has been observed thus far only in this Amazon region in the North of Brazil [6].

One of the earlier genotyping methods known as spoligotyping has been shown to poorly distinguish between closely related strains, resulting in incorrect assumptions regarding strain relatedness, origin and evolution [9]. Additionally, it has no resolution in terms of homoplasy, which leads to further misassumptions. Therefore, we aimed to describe the genetic variability based on 24-MIRU-VNTR and whole-genome sequences (WGS) of a selection of strains from the State of Pará, North Brazil, with SIT 2545 and their phylogenomic assessment in conjunction with publicly available genomes of strains of L3.

2. Materials and Methods

2.1. Study Population

Among 980 *Mtb* strains isolated between 1998 and 2010 in the state of Pará/Brazil that have undergone genotyping by spoligotyping, six identified as SIT 2545 were selected for this study. In vitro culturing, drug susceptibility testing (DST), Phenol-Chloroform DNA extraction and spoligotyping were performed according to Conceição et al. [6].

2.2. Genotyping Based on MIRU-VNTR

MIRU-VNTR typing was performed by amplifying 24-loci [10]. The PCR reaction was performed using a QIAGEN kit (HotStarTaq DNA polymerase, QIAGEN, Hilden, Germany), and the PCR products were analyzed on the ABI3130 sequencer. Alleles were assigned numerical values according to the number of repeats using GeneMapper Soft-

ware v3.7 (Applied Biosystems, Waltham, MA, USA). Presently observed genotypes were compared to those present in the MIRU-VNTRplus database (https://www.miru-vntrplus.org/MIRU/index.faces, accessed on 7 July 2022) [11] and used as the basis of a neighbour-joining tree.

2.3. Whole-Genome Sequencing

DNA quantity for each sample was determined using Qubit dsDNA BR (Broad Range) assay kits (Thermo Fisher Scientific, Waltham, MA, USA). The DNA integrity was verified by gel electrophoresis on 1% ultrapure agarose gel in 1xTBE and using a 1 kb Plus DNA ladder (Thermo Fisher Scientific).

The libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina, CA, USA) according to the manufacturer's instructions and quality controlled using the Agilent High Sensitivity DNA Kit (Agilent, CA, USA), followed by the sequencing on an Illumina NextSeq 550 instrument using a 2 × 150 paired-end chemistry. The mean read depth was >20×.

2.4. Genome Analysis and Phylogenetic Reconstruction

The six raw data reads belonging to SIT 2545 and the robust dataset totalling 1691 *Mtb* genomes were subjected to quality controlled using FastQC v0.11.7 [12] before and after trimming for removal of adapter sequences and low-quality reads using Trimmomatic v0.33 [13] (parameters: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:36). The trimmed reads were mapped against the *Mtb* H37Rv reference genome (GenBank accession number: NC_000962.3) through the MTBseq (v1.0.3) pipeline using BWA-MEM v0.7.16. Samtools v1.9 [14] was used to convert the SAM files to BAM format and sort mapped sequences and BAM file quality was checked using Qualimap [15]; sambamba v0.6.8 was used to mark read duplicates [16].

The MTBseq (v1.0.3) performed individual ran sample analysis individually in separate environments, removing samples with genome coverage < 95% and joining the remaining samples for comparative analysis. Software dependencies were installed using the Conda package manager through the Bioconda channel [17]. The infrastructure required to run the analysis was provided by Instituto Evandro Chagas (PBS cluster with 80 x86 CPUs with 1TB RAM).

As part of the MTBseq (v1.0.3) pipeline, single nucleotide variants (SNVs), which includes single nucleotide polymorphisms (SNPs) and small insertions and deletion (*indels*) were called and filtered based on the following default pipeline criteria: mapping quality $\geq$ 50, base alignment quality $\geq$ 23 and $\leq$ 2000 reads covering each site. Most variants detected in PE/PPE and other hard-to-map regions were automatically excluded based on these criteria and according to default MTBseq pipeline parameters. The resulting variant tables were then screened according to the following parameters: minimum read depth of 10 to call a variant and 2000 maximum read depth. Variant functional annotation was performed with SnpEff (v4.3) [18], while mutations were manually confirmed from BAM files.

The similarity matrix was constructed using the cohort analysis step of the MTBseq pipeline. High-confidence variable sites detected in all isolates, including coding and non-coding SNPs, were concatenated to generate a multi-FASTA file to construct a maximum likelihood phylogeny of the isolates included in this analysis with IQ-TREE (v2.2.0) with 1000 bootstrap pseudo-replicates using an ultrafast and automatic model selection method. We included *Mtb* genomes of representative human-adapted Lineages (L1–L9) and the outgroup *Mycobacterium canettii* (GenBank accession number: SAMEA3905803). The tree was annotated using the web tool iTOL (v6.6) (https://itol.embl.de/, accessed on 20 November 2022) [19].

The classification of the genotypes to the lineage and sublineage level and characterization of mutations associated with drug resistance was performed using TB-Profiler v4.4.0 [20], Spotyping v2.1 [21] and RD-analyzer v1.01 [22] in the command line version

using default settings. We also performed de novo assembly using SPAdes to assemble the reads. Thereafter, the genome annotation was established using Prokka v11.3.4 [23].

Flowchart summarizing the various aspects of this study can be found in Figure S1.

2.5. Ethical Statement

This study was approved by the Ethical Committee of the Evandro Chagas Institute, number 059750/2017, CAAE: 69248217.0.0000.0019.

3. Results

3.1. MIRU-VNTR Result

The characteristics of the six selected MTBC strains of the CAS lineage with SIT 2545 (octal code 703660000000331) are described in Table 1. The strains had been isolated from 50% of male and 50% of female patients between 2001 and 2010 and therefore span a 10-year period. Strain 431, isolated in 2001, presented a multidrug resistance (MDR) profile, while the others were drug-susceptible.

Table 1. Strain isolation year, drug-susceptibility test and sociodemographic characteristics associated with the mycobacterial interspersed repetitive unit—variable number of tandem repeats (MIRU-VNTR) and whole-genome sequencing (WGS) clustering analysis.

Strain	Year	DST [1]	Sex	Age	MIRU-VNTR [2]	WGS Cluster [3]
431	2001	MDR	F	34	24233554**2**2443413**8**443843	Non clustered
1906	2005	Susc	M	29	242335542124424178343843	Cluster 1
1918	2010	Susc	F	61	242335542124424178343843	Non clustered
2224	2006	Susc	M	26	24233**4**542124424178343843	Cluster 1
2248	2006	Susc	M	-	242335542124424178343843	Cluster 1
2537	2008	Susc	F	38	242335542124424178343843	Cluster 1

[1] Drug susceptibility test based on Proportion Method: multidrug-resistant (MDR) tuberculosis, simultaneous resistance to rifampicin and isoniazid, drug-susceptible (Susc). [2] Mycobacterial interspersed repetitive unit—variable number of tandem repeats (MIRU-VNTR) profile arranged in a crescent order of the 24-loci. [3] Whole-genome sequencing (WGS) cluster interpretation based on 12 SNPs cut-off. Bold/underline: it highlights the alleles differences from the majority of cohort result.

Upon comparison of the MIRU pattern(s) with the MIRU-VNTRplus database, the six isolates were grouped into a monophyletic branch related to the clade of Delhi/CAS. The 24-MIRU-VNTR loci pattern was identical for 19 loci, with a difference in two alleles in five loci (960_MIRU10, 2163b_QUB11b, 2461_ETRB, 2996_MIRU26, 3171_Mtub34) poorly discriminant with a discriminatory index of $h = 0.13$. The six strains were divided into two main groups: (1) orphan profile strain 431 and (2) a cluster harboring strains 1906, 1918, 2248, and 2537. Strain 2224, also clusters with these strains, but with a single locus variation (SLV) (Figure 1).

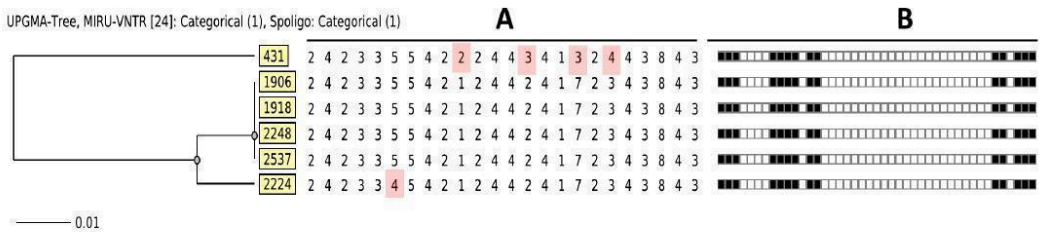


Figure 1. Phylogenetic tree (**A**) based on 24-MIRU-VNTR loci analysis on six *Mycobacterium tuberculosis* strains highlighting the quadruple-locus variation on strain 431 and a single-locus variation on strains 2224; identical spoligotyping profile (**B**) indicative of the Central-Asia (CAS) SIT 2545 genotype.

3.2. Whole Genome Sequencing Analysis and Phylogenomic Assessment

On average, the WGS of each of each of the six strains had of 16.18 million pared reads per sample before trimming and 6.5 million reads after quality trimming at an average genome coverage of 254.54×, and 65.51% GC content. Individual genomic characteristics are described in Table 2 and Table S1.

Table 2. Core genomic characteristics of the six *Mycobacterium tuberculosis* Lineage 3 strains from Northern Brazil based on the outputs of the reference-based MTBseq pipeline, Prokka annotation and de novo assembly analysis.

Strain	Mean Coverage	Mapping [1]	Contigs	Largest Contig	N_{50}	N_{75}	Insertion [1]	Deletion [1]	SNPs [1]
431	261.13	99.26%	165	173,577	61,250	39,904	109	162	1395
1906	277.38	99.23%	167	155,704	60,542	37,278	118	130	1381
1918	281.59	99.29%	162	161,736	62,665	37,278	139	120	1402
2224	267.62	99.32%	158	158,538	55,650	35,572	131	135	1387
2248	187.94	99.30%	178	214,707	52,142	33,124	135	121	1369
2537	251.62	99.33%	164	179,030	56,137	35,288	127	120	1386

[1] Against the *Mycobacterium tuberculosis* reference H37Rv (NC_000962.3).

For the pairwise genetic comparisons based on SNVs, strains 1906, 1918, 2224, 2248 and 2537 differed by between 4 and 21 SNVs. Strain 431 was the outlier with at least 52 SNVs between the closest strain, 1906 (Figure 2).

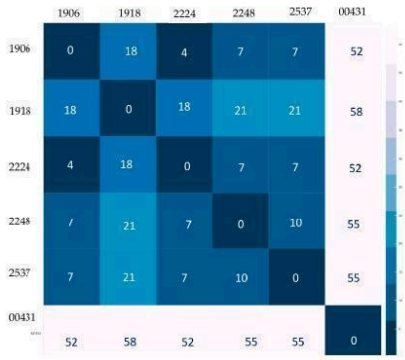


Figure 2. Heatmap based on single-nucleotide variant (SNV) distance among the six *Mycobacterium tuberculosis* Central-Asia (CAS) SIT 2545 genomes from Brazil. Darker blue is indicative of shorter genetic distance whereas light blue to white indicates a larger number of SNVs between isolates.

Out of the 1691 genomes, 1350 (79.83%) passed the quality control (genomic coverage > 95%). The six novel strains, clustering together, were positioned on a subclade which shares a common ancestor with the second subclade with seven known L3 strains at the tips of the nodes. These genomes are from the United Kingdom (n = 4; ERR072023; SRR2100697; SRR2101190; and SRR2100980), and an unknown origin but possibly from Azerbaijan, Bangladesh, Belarus, Pakistan, Philippines, South Africa, and Ukraine (= 3; SRR6964598; SRR6797470; and SRR6964614). This distinct L3 clade comprising of two subclades were positioned the closest to four other L3 clades (Figure 3).

commonly found in Asia. The lineage CAS-Delhi was reported using MIRU-VNTR 24 loci in four isolates from pulmonary TB patients in Minas Gerais, Brazil [25].

Due to spoligotyping not being able to accurately infer the true phylogenetic relationships owing to genetic convergence, we performed WGS to determine the true CAS lineage evolution in South America, in particular Northern Brazil. The 24-loci MIRU-VNTR was performed to provide a preliminary overview of the global population structure and spatial distribution of L3 strains found in the region. WGS further revealed the global phylogeny, geographically differentiated pattern, and related clusters.

We provide new evidence that L3 is found in Brazil, which therefore strongly suggests that an introductory event of the CAS-Delhi lineage might have occurred historically on the South American continent. This correlates with a previous study where two lineages CAS-Delhi (L3) were found in Ecuador, a region in which L4 is the most prevalent (98.4%). In this study, 24-MIRU-VNTR loci were used as the genotyping method in this first non-Brazilian report of CAS-Delhi in South America [26].

Here, based on MIRU-VNTR applied for CAS-Delhi strains, nine VNTR loci (4052-QUB26, 2996-MIRU26, 424-Mtub04, 960-MIRU10, 4156-QUB4156c, 3192-MIRU31, 2165-ETRA, 4348-MIRU39, and 2059-MIRU20) exhibited high discrimination [27], which is similar to the five most discriminatory loci in our study (960-MIRU10, 2163b-QUB11b, 2461-ETRB, 2996-MIRU26, 3171-Mtub34).

The *Mtb* L3 is not the only lineage with specific features likely to have adapted upon geographical introduction into Brazil. Similarly, L1, which commonly has a low correlation to transmissibility and virulence and is generally present in a restricted geographical distribution in Eastern Africa and the South of India, was also found in Pará and has shown different profiles as resistance to antibiotics being MDR [8]. This pattern of evolution was identified using the WGS method.

From our results of the six SIT 2545 strains, interesting observations were made when comparing 24-MIRU-VNTR loci and SNV-distance from WGS analysis. For example, strain 2224 differed by one MIRU locus from the other strains but had a mere a 4-SNV difference compared to strain 1906. In contrast, isolate 1918 appeared to be closely related to the other three strains based on MIRU-VNTR clustering analysis (Figure 1), but the SNV difference was at least 18 (Figure 2). This emphasizes the need to not only rely on more traditional genotyping methods, but also to utilize WGS analysis which provides a more accurate representation of strain relatedness.

Although the State of Pará has the fifth highest number of emerging TB cases in Brazil, molecular characterization of these strains is underutilized in this region. For comparison, MIRU-VNTR and WGS are used more widely in the Southeast, South, and Central West region [5]. The MDR profile of strain 431 (isolated in 2001) and the susceptible profiles of the other strains isolated after 2001 could imply that there are possibly several missed cases yet to be discovered, described and characterized with a similar SIT profile.

The divergence observed In MIRU-VNTR loci and WGS suggests that we are missing epidemiologically linked strains might be a limitation of our study, but it does suggest that there might be more clusters and outbreaks caused by this strain that was missed. Additionally, comparative analysis with the MTBseq pipeline requires excellent coverage of the genomes of interest, and 341 (20.16%) of the genomes downloaded from public databases.

Gold standard genotyping methods are often overlooked in regions with only a small number of laboratories, professionals, and appropriate supplies. Subsequently, the accurate identification and classification of *Mtb* is delayed, which leads to multiple health risk issues. These include increased community exposure to the bacillus, difficulties in outbreak investigations, transmission dynamics and ultimately, rapid access to an effective treatment. Large-scale genotyping is also needed to get better insights into the biological diversity and the evolution of the pathogen [28].

In a recent study it was demonstrated that the proportion of MDR and extensively drug-resistant (XDR) TB among CAS lineage is significantly higher than EAI lineage (from

the eastern hemisphere and found in the Americas) [1], which highlights the need for an association of DST profile and genotypes applied to TB surveillance.

To date, the state of Pará has used WGS successfully to investigate the transmission of TB applied to *Mtb* L1 [2,7] and *Mycobacterium bovis* [29], albeit retrospectively. There is a necessity for WGS to be utilized even more widely and to expand its usage TB strains. This will greatly contribute to our enhanced understanding of the dynamics of TB transmission in the local population and pathogen-host co-evolution in Northern Brazil.

Genotyping of MTBC strains is the basis of establishing the phylogeny of this complex and complements TB transmission studies. The latter provides important guidelines for health agencies responsible for screening and interventions to implement strategies for the prevention and treatment of the disease and identify areas of higher priority.

5. Conclusions

This is a pioneer study confirming the presence of *Mtb* L3 in Brazil through 24-MIRU-VNTR and WGS. The comparison against an international dataset including 1350 genomes confirmed SIT 2545 belongs to a separate cluster. The evaluation of the transmissibility and the search for the possible origins of this lineage in the country based on global phylogenomic have great relevance for the maintenance of continued surveillance and for the promotion of public health. Our study contributes to further understanding of the *Mtb* genetic diversity in a high-burden TB area and pinpoints the need to improve genetic/genomic surveillance of the *Mtb* population and unravel the evolution and diversity of L3 in Brazil and South America.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/microorganisms11010132/s1, Table S1: Statistics generated by MTBSeq pipeline of the *1350 M. tuberculosis* L3 genomes. Figure S1: Flowchart summarizing the various aspects of this study.

Author Contributions: Conceptualization, E.C.C., M.L.d.C. and K.V.B.L.; Data curation, E.C.C. and D.J.M.; Formal analysis, E.C.C., D.J.M. and J.L.; Funding acquisition, A.C.R.C. and K.V.B.L.; Investigation, M.L.d.C.; Methodology, M.L.d.C., D.J.M., J.L., G.L.A., S.P.d.S., A.C.R.C. and A.S.; Project administration, E.C.C. and K.V.B.L.; Resources, E.C.C., A.C.R.C. and K.V.B.L.; Software, E.C.C., D.J.M., J.L. and A.S.; Supervision, P.S. and K.V.B.L.; Validation, E.C.C., M.L.d.C. and A.S.; Visualization, E.C.C.; Writing—original draft, E.C.C., M.L.d.C., D.J.M. and J.L.; Writing—review & editing, E.C.C., M.L.d.C., D.J.M., J.L., G.L.A., S.P.d.S., A.C.R.C., A.S. and P.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by CNPQ, grant number 207,071/2014-4, and grant number Finance Code 001 and CAPES (Finance Code 001).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The six raw sequence reads were deposited at NCBI under the Bioproject accession ID PRJNA895862.

Acknowledgments: We acknowledge the Centro de Inovações Tecnológicas at Instituto Evandro Chagas and Maria Luiza Lopes for technical assistance.

Conflicts of Interest: A.S. works for Seqera Labs S.L. which is the driving company behind the open source Nextflow software. The remaining authors declare no competing interest.

References

1. Couvin, D.; Reynaud, Y.; Rastogi, N. Two tales: Worldwide distribution of Central Asian (CAS) versus ancestral East-African Indian (EAI) lineages of Mycobacterium tuberculosis underlines a remarkable cleavage for phylogeographical, epidemiological and demographical characteristics. *PLoS ONE* **2019**, *14*, e0219706. [CrossRef] [PubMed]
2. Menardo, F.; Rutaihwa, L.K.; Zwyer, M.; Borrell, S.; Comas, I.; Conceição, E.C.; Coscolla, M.; Cox, H.; Joloba, M.; Dou, H.-Y.; et al. Local adaptation in populations of Mycobacterium tuberculosis endemic to the Indian Ocean Rim. *F1000Research* **2021**, *10*, 60. [CrossRef] [PubMed]

3. Coscolla, M.; Gagneux, S. Consequences of genomic diversity in Mycobacterium tuberculosis. *Semin. Immunol.* **2014**, *26*, 431–444. [CrossRef] [PubMed]
4. Yimer, S.A.; Kalayou, S.; Homberset, H.; Birhanu, A.G.; Riaz, T.; Zegeye, E.D.; Lutter, T.; Abebe, M.; Holm-Hansen, C.; Aseffa, A.; et al. Lineage-Specific Proteomic Signatures in the Mycobacterium tuberculosis Complex Reveal Differential Abundance of Proteins Involved in Virulence, DNA Repair, CRISPR-Cas, Bioenergetics and Lipid Metabolism. *Front. Microbiol.* **2020**, *11*, 550760. [CrossRef] [PubMed]
5. Conceição, E.C.; Salvato, R.S.; Gomes, K.M.; Guimarães, A.E.D.S.; da Conceição, M.L.; Guimarães, R.J.D.P.S.E.; Sharma, A.; Furlaneto, I.P.; Barcellos, R.B.; Bollela, V.R.; et al. Molecular epidemiology of Mycobacterium tuberculosis in Brazil before the whole genome sequencing era: A literature review. *Memórias Do Inst. Oswaldo Cruz* **2021**, *116*, e200517. [CrossRef]
6. Conceição, E.C.; Rastogi, N.; Couvin, D.; Lopes, M.L.; Furlaneto, I.P.; Gomes, H.M.; Vasconcellos, S.E.G.; Suffys, P.N.; Schneider, M.P.C.; de Sousa, M.S.; et al. Genetic diversity of Mycobacterium tuberculosis from Pará, Brazil, reveals a higher frequency of ancestral strains than previously reported in South America. *Infect. Genet. Evol.* **2017**, *56*, 62–72. [CrossRef]
7. Conceição, E.C.; Guimarães, A.E.D.S.; Lopes, M.L.; Furlaneto, I.P.; Rodrigues, Y.C.; Da Conceição, M.L.; Barros, W.A.; Cardoso, N.C.; Sharma, A.; Lima, L.N.G.C.; et al. Analysis of potential household transmission events of tuberculosis in the city of Belem, Brazil. *Tuberculosis* **2018**, *113*, 125–129. [CrossRef]
8. Guimarães, A.E.D.S.; Sharma, A.; Furlaneto, I.P.; Rutaihwa, L.; Cardoso, J.F.; da Conceição, M.L.; Spinassé, L.B.; Machado, E.; Lopes, M.L.; Duarte, R.S.; et al. Evaluation of drug susceptibility profile of Mycobacterium tuberculosis Lineage 1 from Brazil based on whole genome sequencing and phenotypic methods. *Memórias Do Inst. Oswaldo Cruz* **2020**, *115*, e200520. [CrossRef]
9. Reyes, J.F.; Chan, C.H.; Tanaka, M.M. Impact of homoplasy on variable numbers of tandem repeats and spoligotypes in Mycobacterium tuberculosis. *Infect. Genet. Evol.* **2012**, *12*, 811–818. [CrossRef]
10. Yasmin, M.; Le Moullec, S.; Siddiqui, R.T.; De Beer, J.; Sola, C.; Refrégier, G. Quick and cheap MIRU-VNTR typing of Mycobacterium tuberculosis species complex using duplex PCR. *Tuberculosis* **2016**, *101*, 160–163. [CrossRef]
11. 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*, W326–W331. [CrossRef]
12. Wingett, S.W.; Andrews, S. FastQ Screen: A tool for multi-genome mapping and quality control. *F1000Res* **2018**, *7*, 1338. [CrossRef]
13. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]
14. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef]
15. García-Alcalde, F.; Okonechnikov, K.; Carbonell, J.; Cruz, L.; Götz, S.; Tarazona, S.; Dopazo, J.; Meyer, T.; Conesa, A. Qualimap: Evaluating Next-Generation Sequencing Alignment Data—PubMed. Available online: https://pubmed.ncbi.nlm.nih.gov/22914218/ (accessed on 30 November 2022).
16. Tarasov, A.; Vilella, A.J.; Cuppen, E.; Nijman, I.J.; Prins, P. Genome Analysis Sambamba: Fast Processing of NGS Alignment Formats. *Bioinformatics* **2015**, *31*, 2032–2034. [CrossRef]
17. Grüning, B.; Dale, R.; Sjödin, A.; Chapman, B.A.; Rowe, J.; Tomkins-Tinch, C.H.; Valieris, R.; Köster, J.; Bioconda Team. Bioconda: Sustainable and comprehensive software distribution for the life sciences. *Nat. Methods* **2018**, *15*, 475–476. [CrossRef]
18. Cingolani, P.; Platts, A.; Wang, L.L.; Coon, M.; Nguyen, T.; Wang, L.; Land, S.J.; Lu, X.; Ruden, D.M. 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* **2012**, *6*, 80–92. [CrossRef]
19. Letunic, I.; Bork, P. Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Res.* **2021**, *49*, W293–W296. [CrossRef]
20. Phelan, J.E.; O'Sullivan, D.M.; Machado, D.; Ramos, J.; Oppong, Y.E.A.; Campino, S.; O'Grady, J.; McNerney, R.; Hibberd, M.L.; Viveiros, M.; et al. Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. *Genome Med.* **2019**, *11*, 41. [CrossRef]
21. Xia, E.; Teo, Y.-Y.; Ong, R.T.-H. SpoTyping: Fast and accurate in silico Mycobacterium spoligotyping from sequence reads. *Genome Med.* **2016**, *8*, 19. [CrossRef]
22. Faksri, K.; Xia, E.; Tan, J.H.; Teo, Y.-Y.; Ong, R.T.-H. In silico region of difference (RD) analysis of Mycobacterium tuberculosis complex from sequence reads using RD-Analyzer. *BMC Genom.* **2016**, *17*, 847. [CrossRef]
23. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [CrossRef]
24. Gomes, H.M.; Elias, A.R.; Oelemann, M.A.C.; Pereira, M.A.D.S.; Montes, F.F.O.; Marsico, A.G.; Kritski, A.L.; Filho, L.D.A.; Caldas, P.C.; Possuelo, L.G.; et al. Spoligotypes of Mycobacterium tuberculosis complex isolates from patients residents of 11 states of Brazil. *Infect. Genet. Evol.* **2012**, *12*, 649–656. [CrossRef]
25. De Almeida, I.N.; Vasconcellos, S.E.G.; Figueredo, L.J.D.A.; Dantas, N.G.T.; Augusto, C.J.; Hadaad, J.P.A.; Suffys, P.N.; Carvalho, W.D.S.; De Miranda, S.S. Frequency of the Mycobacterium tuberculosis RDRio genotype and its association with multidrug-resistant tuberculosis. *BMC Infect. Dis.* **2019**, *19*, 556. [CrossRef]
26. Garzon-Chavez, D.; Garcia-Bereguiain, M.A.; Mora-Pinargote, C.; Granda-Pardo, J.C.; Leon-Benitez, M.; Franco-Sotomayor, G.; Trueba, G.; de Waard, J.H. Population structure and genetic diversity of Mycobacterium tuberculosis in Ecuador. *Sci. Rep.* **2020**, *10*, 6237. [CrossRef]

27. Hadifar, S.; Kamakoli, M.; Fateh, A.; Siadat, S.D.; Vaziri, F. Enhancing the Differentiation of Specific Genotypes in Mycobacterium Tuberculosis Population—PubMed. Available online: https://pubmed.ncbi.nlm.nih.gov/31784605/ (accessed on 30 November 2022).
28. Merker, M.; Kohl, T.A.; Niemann, S.; Supply, P. The Evolution of Strain Typing in the Mycobacterium tuberculosis Complex. *Strain Var. Mycobacterium Tuberc. Complex Its Role Biol. Epidemiol. Control* **2017**, *1019*, 43–78. [CrossRef]
29. da Conceição, M.L.; Conceição, E.C.; Furlaneto, I.P.; Da Silva, S.P.; Guimarães, A.E.D.S.; Gomes, P.; Boschiroli, M.L.; Michelet, L.; Kohl, T.A.; Kranzer, K.; et al. Phylogenomic Perspective on a Unique Mycobacterium bovis Clade Dominating Bovine Tuberculosis Infections among Cattle and Buffalos in Northern Brazil. *Sci. Rep.* **2020**, *10*, 1747. [CrossRef]

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

5.1.5 Simulação Computacional De Variantes Relacionadas À Resistência À Bedaquilina Em Mycobacterium Tuberculosis

Esta produção atende ao objetivo específico 6.

Este trabalho atende ao objetivo específico 6, por investigar através de metodologias da simulação computacional de proteínas, o impacto de mutações detectadas por WGS.

Artigo apresentando como Trabalho de Conclusão de Curso: **Análise estrutural in silico baseada em modelagem molecular: Explorando variantes de resistência à bedaquilina em Mycobacterium tuberculosis.** Apresentado pela Aluna Danna Karen Corrêa dos Santos para a obtenção do grau de bacharela em biomedicina na faculdade de biomedicina da Universidade do Estado do Pará. Este trabalho foi orientado pela Dra. Karla Lima e co-orientado pelo Discente: Davi Marcon durante a realização de seu projeto de mestrado.

Análise estrutural *in silico* baseada em modelagem molecular: Explorando variantes de resistência à bedaquilina em *Mycobacterium tuberculosis*

Danna Karen Corrêa dos Santos[1,2], Davi Josué Marcon[1,2], Karla Valéria Batista Lima[1,2]

[1]Universidade do Estado do Pará, Belém, Pará, Brasil.

[2]Seção de Bacteriologia do Instituto Evandro Chagas, Ananindeua, Pará, Brasil.

Autores correspondentes: Danna Karen Corrêa dos Santos (dannakcsantos@gmail.com), Davi Josué Marcon (davijosuemarcon@gmail.com), Karla Valéria Batista Lima (karlavaleria2007@gmail.com).

Resumo

A bedaquilina (BDQ) foi inserida no esquema de tratamento da TB multirresistente (TB-MDR) em 2012 pela *US Food and Drug Administration* (FDA). No entanto, pouco se conhece sobre os mecanismos de resistência à BDQ desenvolvido pelo complexo *Mycobacterium tuberculosis* (CMTB). Assim, é crucial esclarecer a correlação entre mutações específicas e a resistência fenotípica, visando otimizar os esquemas terapêuticos existentes e garantir maior eficácia no tratamento. O objetivo deste estudo é descrever modificações estruturais de proteínas do genoma de *M. tuberculosis* relacionadas a aquisição de resistência à BDQ por meio de simulação computacional baseada em amostras provenientes de banco de dados público. Foram analisados dados genômicos e fenotípicos de 41 isolados de *M. tuberculosis* com Concentração Inibitória Mínima (CIM) ≥1 e ≥2 obtidos a partir de 12.289 amostras extraídas do banco de dados público do Consórcio *Comprehensive Resistance Prediction for Tuberculosis: an International Consortium* (CRyPTIC). A pesquisa por variantes foi realizada com o *pipeline* MAGMA 1.1.1., a modelagem estrutural de macromoléculas, docagem, análise de interações moleculares e visualização de complexos foram feitas utilizando os programas ColabFold 1.5.5, AutoDock 4.2, LigPlot^{+} 2.2 e ChimeraX 1.8, respectivamente, destacando as mutações de maior impacto. Foram identificadas 19 variantes dos genes *mmpL5*, *mmpR5 (Rv0678)*, *atpE*, *mtrB*, *Rv1979c* e *lpqB* com mutações preditivas para potencial de resistência à BDQ. Nenhuma mutação associada a resistência à BDQ foi encontrada nos genes *mtrA* e *mmpS5* nas cepas avaliadas. A análise dos produtos estruturais revelou o silenciamento do gene *pepQ* e que as proteínas *mmpL5* e *mtrB* apresentaram interação não espontânea com a BDQ. A docagem gerou diversas conformações, selecionando as de menor energia livre de ligação juntamente às constantes de inibição obtidas. As interações entre a BDQ e os receptores foram analisadas com bom parâmetro de validação (pLDDT>80). Os valores adquiridos desses parâmetros para as proteínas *atpE, Rv0678, Rv1979c* e *lpqB*, foram correlacionados com os valores de CIM. Embora a BDQ interaja com o repressor *Rv0678* do sistema *mmpL5-mmpS5*, a droga não demonstra interação direta com o sistema dentre as amostras analisadas, destacando um achado relevante no mecanismo de resistência. Por meio dos resultados descritos, é revelada a importância de compreender a resistência à BDQ em *M. tuberculosis*, utilizando a bioinformática para a avaliação de variantes. Assim, propõe-se um método inovador, com uso de ferramentas de simulação para prever mutações e auxiliar na compreensão da resistência.

Palavras-chave: Tuberculose Multidroga Resistente; Mycobacterium tuberculosis; Modelagem Computacional; Docagem Molecular.

1 Introdução

A tuberculose (TB) é uma doença infectocontagiosa causada pelo complexo *Mycobacterium tuberculosis* (CMTB). Persistindo como um problema de saúde global significativo, em 2023 estima-se que 10,8 milhões de pessoas desenvolveram TB, resultando em 1,25 milhões de mortes, especialmente em países de baixo e médio desenvolvimento (WHO, 2023; WHO, 2024). Embora a TB seja prevenível e curável, seu impacto econômico é alto devido aos longos períodos de tratamento e surgimento de cepas resistentes como a TB multirresistente (TB-MDR) e a TB extensivamente resistente (TB-XDR) (WHO, 2023; Lange *et al*., 2018).

A emergência da resistência antimicrobiana, incluindo à bedaquilina (BDQ), é uma preocupação no combate contra a TB. Aprovada em 2012 para o tratamento da TB-MDR pela *US Food and Drug Administration* (FDA), estudos revelaram o rápido surgimento da resistência a droga em 2014, destacando a necessidade de novas estratégias terapêuticas (Barilar, et al., 2024; Makhado *et al*., 2018; Majahan, 2013; Perumal *et al*., 2023). A BDQ age ao interromper o mecanismo de síntese de adenosina 5'-trifosfato (ATP) celular, através da inibição da ATP sintetase, essencial para o metabolismo bacteriano. Mutações nesta e em outras proteínas-alvo, ainda pouco elucidadas, comprometem sua eficácia (Nguyen *et al*., 2018; Sachan *et al*., 2023).

As mutações associadas a resistência à BDQ têm sido catalogadas pelo Consórcio *Comprehensive Resistance Prediction for Tuberculosis: an International Consortium* (CRyPTIC), que compila dados genômicos e fenotípicos de 12.289 isolados de MTB. Ferramentas como *BashTheBug* e *Mycobacterial Growth Detection Algorithm* (AMyGDA) são usadas pelo CRyPTIC para determinar com precisão a concentração inibitória mínima (CIM) de cepas resistentes, auxiliando no esclarecimento dos fatores associados à resistência (CRyPTIC Consortium, 2022; Fowler *et al*., 2022). A BDQ mantém sua eficácia contra cepas droga-resistentes, desempenhando um papel fundamental no tratamento da TB-MDR e TB-XDR, com alta taxa de conversão de cultura negativa e sucesso terapêutico de aproximadamente 65,8% (Seelam *et al*., 2020; Hatami *et al*., 2022).

A resistência à BDQ envolve genes como *mtrA/B*, *lpqB*, *mmpR5* (*Rv0678*), *mmpS5*, *mmpL5*, *pepQ*, *atpE* e *Rv1979c*. O sistema *mtrA/B* regula a divisão celular, crescimento e resistência a medicamentos (Peterson *et al*., 2023; Plocinska *et al*., 2012). O gene *lpqB* é co-expresso com o operon *mtrA/B* e atua como um regulador dessa região por meio de sua porção estrutural, sugerindo a presença de uma lipoproteína envolvida na modulação do sistema (Nguyen *et al*., 2010; Li *et al*., 2022). O gene *mmpR5* (*Rv0678*) regula a bomba de efluxo *mmpL5-mmpS5*, e mutações em *Rv0678* resultam na superexpressão da bomba, reduzindo a suscetibilidade de MTB à BDQ e clofazimina (CFZ) (Arora *et al*., 2021; Snobre *et al*., 2023; Saeed *et al*., 2022).

O gene *pepQ* está envolvido na resistência à BDQ por codificar uma aminopeptidase citoplasmática que hidrolisa peptídeos antimicrobianos, enquanto que mutações no gene *atpE*, que codifica a subunidade C da ATP sintetase, afetam a permeabilidade da membrana bacteriana, diminuindo a absorção da droga (Almeida *et al*., 2016; Karmakar *et al*., 2019). O gene *Rv1979c* codifica um provável transportador de membrana com atividade de permease, também associada à resistência por alterar o perfil de absorção da droga (Kadura *et al*., 2020; Islam *et al*., 2024; Nimmo *et al*., 2024; Ismail *et al*., 2021; Derendinger *et al*., 2023).

A Organização Mundial da Saúde (OMS) identificou dezenove variantes associadas a resistência à BDQ e trinta e seis ainda permanecem com classificação incerta quanto a resistência, assim, sua extensão global e a base genética ainda precisam ser totalmente investigadas. A correlação entre mutações específicas e a resistência fenotípica é pouco compreendida, sendo essencial esclarecer essas lacunas para otimizar os esquemas terapêuticos selecionados, garantindo maior eficácia no tratamento. Nesse contexto, o uso de *pipelines*, *machine learning*, modelagem e docagem na pesquisa

de variantes à BDQ em MTB é uma alternativa confiável, menos morosa e de custo reduzido (Nimmo *et al*., 2024; WHO, 2023; Ejalonibu, *et al*. 2021; Jumper *et al*., 2021; Nguyen *et al*., 2024).

Com o aumento dos casos de TB-MDR e TB-XDR, novas metodologias de pesquisa são essenciais. A análise *in silico* avançou na compreensão da resistência à BDQ e na identificação de mutações associadas. Integrada a abordagens experimentais, pode transformar significativamente a compreensão dos mecanismos de resistência, fornecendo informações inéditas sobre mudanças estruturais e impulsionando o desenvolvimento de novos tratamentos e ferramentas de diagnóstico, fortalecendo a luta contra a TB-XDR (Bai *et al*., 2018).

Nosso estudo teve como foco descrever modificações estruturais em proteínas do genoma de *Mycobacterium tuberculosis* relacionadas a aquisição de resistência à BDQ, empregando simulação computacional baseada em dados provenientes de um banco público. Para isso, foi desenvolvido um banco local de mutações associadas à BDQ, possibilitando a classificação dessas mutações conforme seu impacto na estrutura final das proteínas traduzidas a partir dos genes mutados. Além disso, investigamos a relação entre as alterações estruturais nas proteínas e a aquisição de resistência ao medicamento, contribuindo para o entendimento dos mecanismos moleculares envolvidos.

2 Material e Métodos

2.1 Aspectos éticos

Este estudo *in silico* não necessitou coleta de dados pessoais sensíveis ou amostras biológicas de pacientes. É referente ao projeto de Pesquisa "VIGILÂNCIA DA TUBERCULOSE E NOVO TESTE NÃO INVASIVO DE DIAGNÓSTICO E PROGNÓSTICO", aprovado pelo CEP/IEC em 2 de abril de 2020, CAAE: 0821119.4.0000.0019, cadastrado no SISGEN em 11 de fevereiro de 2022, no A0BE3A4, registrado no ESPRO n° 2107004, SEI 25209.005626/2022-11.

2.2 Delineamento do estudo

Trata-se de um estudo exploratório e descritivo com abordagem quali-quantitativa, cujo tratamento e análise dos dados são fundamentados nos pressupostos de pesquisa *in silico*.

2.3 Análises *in silico* utilizando ferramentas de bioinformática

2.3.1 Ferramentas utilizadas

Para assegurar o controle de qualidade dos genomas, foram utilizados os programas Trimmomatic 0.39 (Bolger, *et. al*., 2014), FastQC 0.12.1 (https://github.com/s-andrews/FastQC) e MultiQC 1.23 (Ewels et al., 2016). Para alinhar leituras contra o genoma de referência H37Rv, permitindo a identificação de polimorfismos de nucleotídeo único (SNPs), foi utilizado o *pipeline Maximum Accessible Genome for Mtb Analysis* (MAGMA) 1.1.1, otimizado para analisar genomas de MTB detectando SNPs e *indels* com impacto na resistência (Heupink *et al*., 2023). Através da ferramenta EMBOSS Transeq (*European Molecular Biology Open Software Suite Transeq*) as variantes foram transformadas em aminoácidos (Madeira *et al.*, 2022), e para remoção de peptídeos sinal utilizou-se o SignalP 6.0 (Nielsen *et al*., 2024). Quanto ao estudo da modelagem, docagem, interação molecular e visualização de complexos, os algoritmos escolhidos foram, respectivamente, ColabFold 1.5.5 (Jumper *et al*., 2021), AutoDock 4.2 (Morris *et al*., 2009), LigPlot^{+} 2.2 (Laskowski e Swindells, 2011) e ChimeraX 1.8 (Meng *et al*. 2023).

2.3.2 Obtenção de leituras

Os dados utilizados para a execução desse estudo são leituras de sequenciamento de genomas completos de cepas de MTB, disponibilizadas pelo Consórcio CRyPTIC contendo informações de CIM e dados fenotípicos de alta confiabilidade em todas as cepas. As leituras adquiridas e os dados das cepas foram agregados em um banco de dados próprio suplementar, possibilitando consultas e análises durante a execução da pesquisa.

2.3.3 Busca por variantes relacionadas a resistência

O presente estudo empregou uma abordagem metodológica que envolveu a seleção de 41 cepas de MTB do Consórcio CRyPTIC (https://ftp.ebi.ac.uk/pub/databases/cryptic/release_june2022/) que apresentavam variantes com resistência à BDQ, seguindo os parâmetros estabelecidos para obtenção destas, que exigiam fenótipo de resistência à droga com CIM≥1 e com CIM≥2, onde esta última concentração representa um maior potencial para a investigação da pesquisa (CRyPTIC *Consortium*, 2022). A partir dessa triagem, as leituras brutas obtidas foram submetidas a ferramentas de bioinformática, iniciando pela realização de controle de qualidade, através do Trimmomatic 0.39 (Bolger, *et. al*., 2014), do FastQC 0.12.1 (https://github.com/s-andrews/FastQC) e do MultiQC 1.23 (Ewels et al., 2016).

A trimagem realizada via Trimmomatic 0.39 (Bolger, *et. al*., 2014) refere-se à etapa de filtragem das sequências genômicas, com o objetivo principal de eliminar leituras de baixa qualidade sequências de adaptadores. Ambos, FastQC 0.12.1 (https://github.com/s-andrews/FastQC) e MultiQC 1.23 (Ewels et al., 2016), oferecem parâmetros de controle de qualidade e geram relatórios interativos em HTML para visualização, com destaque para o MultiQC 1.23 (Ewels et al., 2016), que consolida análises do FastQC 0.12.1 (https://github.com/s-andrews/FastQC) em um único relatório e diretório de arquivos de dados para análises subsequentes.

O alinhamento das leituras foi realizado para cada uma das 41 amostras, utilizando a técnica de sobreposição das suas respectivas *reads* de sequenciamento genômico contra o genoma de referência H37Rv (NC_000962.3) pelo *pipeline* MAGMA 1.1.1 (Heupink *et al*., 2023). Este processo foi essencial para obtenção das 19 variantes que foram identificadas, permitindo uma análise comparativa entre as sequências amostradas e o genoma de referência, proporcionando uma base confiável para as etapas subsequentes de chamada de SNPs e análises estruturais.

Após o alinhamento, a chamada de variantes foi realizada também no ambiente do *pipeline* MAGMA 1.1.1 (Heupink *et al*., 2023) e então foi obtida a sequência consenso para cada cepa, por meio da submissão de cada uma ao *script bcftools consensus*, onde cada amostra alinhada em formato VCF teve suas variantes sobrepostas para análise de concordância entre elas e o genoma de referência. Nesta etapa, foi gerada a sequência correspondente à cada cepa, exportadas em formato FASTA, através do *software* BCF Tools 1.18 (Danecek *et al*., 2021).

Para a realização da transcrição e tradução das sequências contendo variantes, foi utilizada a ferramenta EMBOSS Transeq (*European Molecular Biology Open Software Suite Transeq*), que traduz as sequências nucleotídicas para as suas sequências peptídicas correspondentes e gera traduções em qualquer um dos três *frames* de leitura, no sentido direto ou reverso, ou em todos os seis *frames* a partir da detecção de códon de início e parada (Madeira *et al*., 2022). Após, utilizou-se a ferramenta SignalP 6.0 para a remoção de peptídeos sinal de cada uma das variantes estudadas, com o objetivo de garantir que a estrutura modelada representasse exclusivamente a porção funcional da proteína (Nielsen *et al*., 2024).

2.3.4 Modelagem e Docagem Molecular de Proteínas

Para estudo de simulação tridimensional e docagem molecular foi utilizado protocolo base para a determinação de interações atômicas e para simulação computacional proteica. As proteínas cujas estruturas modificadas obtiveram afinidade inferior de qualquer valor quando comparadas à ligação da proteína *wild type* à BDQ foram consideradas como potenciais causas de resistência e comparadas com os dados fenotípicos obtidos para a cepa correspondente (Verli, 2014; Morris *et al*., 2009).

As estruturas cristalográficas das proteínas selvagens, como *mmpR5* (*Rv0678*) do gênero *Mycobacterium* foram obtidas do *Protein Data Bank* (PDB) (PDB ID − 4NB5 cadeia A) (Radhakrishnan *et al*., 2014). Constatou-se que as estruturas *mtrB*, *pepQ*, *mmpL5*, *atpE* e *Rv1979c* não são disponíveis no PDB, portanto, as sequências foram derivadas, respectivamente, do banco de dados UniProt (UniProt ID − P9WGK9, A0A654ZTY1, A0A2A7N862, A0A7I7KUC3 e P9WQM4). O ligante BDQ foi obtido do banco de dados PubChem sob CID 5388906, no formato SMILES. Utilizou-se o site NovoPro Bioscience Inc. *server* (https://www.novoprolabs.com/tools/smiles2pdb) para converter a estrutura para o formato PDB.

A predição de estruturas de proteínas com mutações foi simulada por meio do *pipeline* nf-core/proteinfold com *subworkflow* ColabFold 1.5.5 (Jumper *et al*., 2021) conforme a **Figura 1,** que via *machine learning*, simula mutações de resistência com precisão atômica, e foi acelerado por GPU utilizando três reciclos mais o pacote AMBER para validação das energias para cada modelo. Os cinco melhores modelos para cada variante foram mantidos e validados utilizando pontuação pLDDT calculada para cada um junto a geração de gráfico Ramachandran. Assim, foram consideradas apenas as variantes com pLDDT superior a 80% e 97% das posições favoráveis no diagrama de Ramachandran (**Tabela Suplementar 1**), o que configura bons padrões de acordo com o sugerido pelos desenvolvedores do *pipeline*.

Após modelagem, com a obtenção dos arquivos em formato .PDB e .DLG *"Docking Log File"*, foi possível realizar a docagem através do AutoDock 4.2 (Morris *et al*., 2009), acelerado por GPU, utilizando o método de *blind docking*, que consiste na docagem da molécula sobre toda sua superfície sem conhecimento prévio do sítio alvo (Ghersi e Sanchez, 2009). Para identificar possíveis sítios de ligação, foram utilizadas 1.000 corridas e 2.500.000 avaliações de energia de ligação. Foi utilizado o software LigPlot^{+} 2.2 (Laskowski e Swindells, 2011) para a obtenção das interações dos aminoácidos entre ligante e macromolécula frente às diferentes variantes proteicas analisadas. A emulação gráfica da interação molecular entre ligante e receptor foi realizada por meio do programa ChimeraX 1.8 (Meng *et al*. 2023) que possui renderização de alto desempenho para visualização e análise interativa de estruturas moleculares.

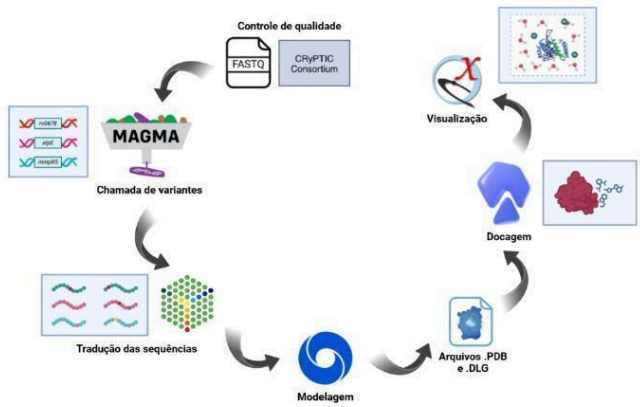


FIGURA 01 | Fluxograma para avaliar como mutações afetam a interação proteína-ligante. Imagem criada com BioRender.com

3 Resultados

3.1 Variantes detectadas

Nesta pesquisa buscou-se identificar mutações preditivas para potencial de resistência à BDQ. No estudo, 41 amostras com CIM≥1 e CIM≥2 foram cuidadosamente escolhidas a partir de uma coleção de 12.289 amostras, e destas, foram identificadas 19 variantes dos genes previamente selecionados − *mmpL5*, *mmpR5* (*Rv0678*), *atpE*, *mtrB*, *Rv1979c* e *lpqB*. Apesar de descritos no catálogo da OMS como associados a resistência, não foi encontrada nenhuma mutação associada à resistência à BDQ nos genes *mtrA* e *mmpS5* nas cepas avaliadas, além disso, a análise das proteínas que conferem resistência à BDQ revelou o silenciamento do gene *pepQ*, com alteração do 2º códon para um códon de parada.

3.2 Estruturas 3D de macromoléculas alvo

Todas as proteínas foram modeladas, contudo, somente os produtos estruturais dos genes *mmpL5*, *mmpR5* (*Rv0678*), *atpE*, *mtrB*, *Rv1979c* e *lpqB* foram submetidos à docagem. Diversas conformações do complexo proteína-ligante foram geradas, a seleção das melhores posições foi realizada com base nas que apresentaram a menor energia livre de ligação (ΔG), que representa a afinidade entre a proteína e o ligante. Apenas as proteínas *mmpL5* e *mtrB* apresentaram interação predita como não espontânea com a BDQ, por possuírem energia de ligação positiva (ΔG>0). Além disso, também foi obtida a constante de inibição (Ki), a qual avalia a quantidade de ligante necessária para alteração no equilíbrio da reação proteína-ligante. As interações entre a BDQ e os receptores-alvo foram visualizadas e analisadas com o uso do programa ChimeraX 1.8 (Meng *et al*., 2023). Acerca da confiança das predições, todas as proteínas modeladas apresentaram bom parâmetro de validação, com regiões de pLDDT > 80, logo, modeladas com precisão apropriada.

3.3 Análise dos parâmetros de atividade molecular das moléculas alvo com a bedaquilina

É essencial no contexto da docagem avaliar os parâmetros de atividade molecular, que são fundamentais para entender a variação na força de interação da droga com diferentes variantes proteicas. A **Tabela 01** a seguir sintetiza os valores obtidos para as proteínas de *atpE*, *Rv0678*, *Rv1979c* e *lpqB*, incluindo tanto as formas selvagens quanto as variantes mutantes, correlacionando esses parâmetros com os valores de CIM.

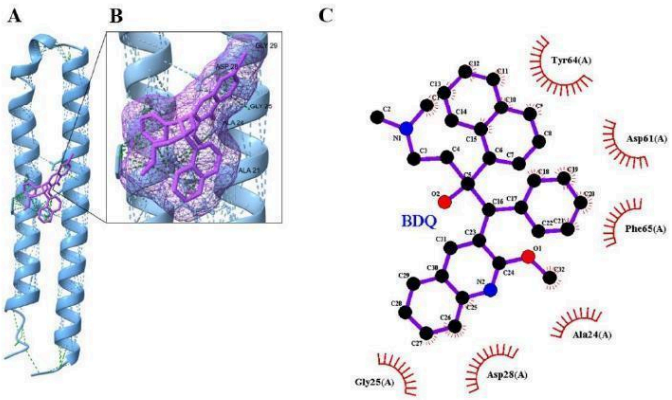
Complexo proteína-BDQ	Modificações de aminoácidos	ΔG (kcal/mol)	Ki (μM)	Alcance do CIM (μg/mL)	Interações por ponte de Hidrogênio	Interações por ligação hidrofóbica
***atpE*_wild**	-	-1,37	99,66	-	Asp28, Asp61	Phe65, Tyr64, Ala24, Ala21, Gly25
***atpE*_var1**	Ala63Pro	-3,4	3,22	>1	Glu61	Leu68, Tyr64, Phe65, Ala21, Asp28, Gly25
***atpE*_var2**	Glu61Asp	-0,25	653,32	2,0	-	Tyr64, Asp61, Phe65, Ala24, Asp28, Gly25
***Rv0678*_wild**	-	-3,92	1,33	-	-	Phe100, Tyr26, Met23, Gln22, Phe19
***Rv0678*_var1**	Met139Ile	-5,23	145,95	>2	-	Phe27, Tyr26, Gln22, Phe19, Leu44, Met23, Phe100, Phe79

***Rv0678*_var2**	MP[a]	-4,47	531,79	>2	-	Phe19, Met23, Leu44, Phe100, Phe27, Tyr26, Gln22
***Rv0678*_var3**	MP[b]	-2,06	31,06	1,0	-	Met17, Trp42, Leu39, Arg38
***Rv0678*_var4**	MP[c]	-3,03	5,99	1,0	-	Leu40, Met23, Leu44, Phe19, Tyr26, Phe27
***Rv0678*_var5**	Aln99Pro	-5,61	77,16	1,0	-	Phe19, Met23, Leu44, Phe100, Phe79, Phe27, Tyr26
***Rv0678*_var6**	MP[d]	-6,13	31,89	1,0	-	Trp111, Tyr26, Phe79, Leu44, Leu40, Phe19, Met23, Trp117
***Rv1979c*_wild**	-	-0,63	346,01	-	-	Asn240, Ala241, Leu22, Ala318, Asn163, Ser237, Ala319, Gly322, Ala323, Leu326, Arg 235
***Rv1979c*_var1**	Asp286Gly	-4,94	240,21	>2	-	Ile380, Tyr434, Phe438, Ile452, Leu208, Val430, Ala429, Phe456
***Rv1979c*_var2**	Arg409Gln	-5,2	154,49	1,0	-	Ile452, Leu377, Phe438, Val210, Leu208, Tyr434, Ala429, Val384, Phe456, Ile437
***lpqB*_wild**	-	-1,59	68,33	-	-	Pro134, Asp135, Pro136, Ile116, Leu92
***lpqB*_var1**	Asp142Gly	-2,73	9,92	>2	-	Pro512, Glu546, Val545, Ala543, Tyr541, Tyr501, Pro549

TABELA 01 | Parâmetros de atividade molecular da bedaquilina com as proteínas *atpE*, *Rv0678*, *Rv1979c* e *lpqB* do *Mycobacterium tuberculosis*. Mutação Precoce (MP). MP[a]= mutações a partir do sítio 110 e término precoce em 118 aminoácidos. MP[b]= mutações a partir do sítio 49 e término precoce em 80 aminoácidos. MP[c]= mutações a partir do sítio 67 e término precoce em 80 aminoácidos. MP[d]= mutações a partir do sítio 98 e término precoce em 121 aminoácidos.

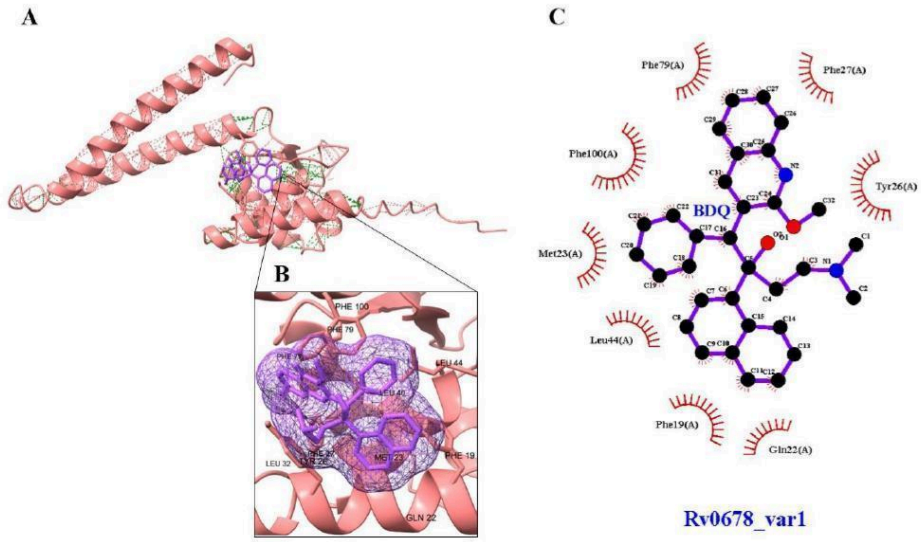
3.4 Docagem Molecular

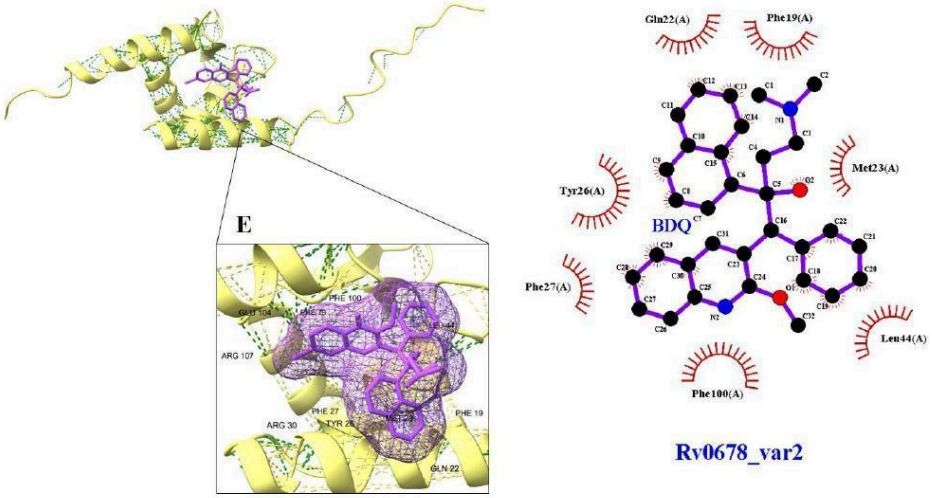
As **Figuras 2** e **3** apresentam as estruturas tridimensionais dos genes-alvo com resultados mais significativos: *atpE* e *Rv0678*. Especificamente, de suas variantes *atpE*_var2, *Rv0678*_var1, *Rv0678*_var2 e *Rv0678*_var6. Foram geradas diversas conformações para os complexos proteína-ligante através do *software* AutoDock 4.2 (Morris *et al*., 2009), com base na menor energia livre de ligação (ΔG), sendo selecionadas aquelas que apresentaram melhores conformações. As interações entre a BDQ e os genes-alvo foram analisadas e visualizadas utilizando os programas ChimeraX 1.8 (Meng *et al*., 2023) e LigPlot+ 2.2 (Laskowski e Swindells, 2011), que permitiram detalhar as interações a nível molecular.



atpE_var2

FIGURA 02 | Complexo estrutural *atpE*_var2-BDQ do gene *atpE*. Afinidade molecular em 3D (A, B). Interações em 2D (C) entre o sítio-alvo e a BDQ por ligações hidrofóbicas são representadas pelos semicírculos ciliados de cor vermelha. Os círculos pretos, azuis e vermelhos correspondem, respectivamente, a átomos de Carbono, Nitrogênio e Oxigênio. Imagem criada utilizando os programas ChimeraX 1.8 e LigPlot+ 2.2 (da esquerda para a direita).





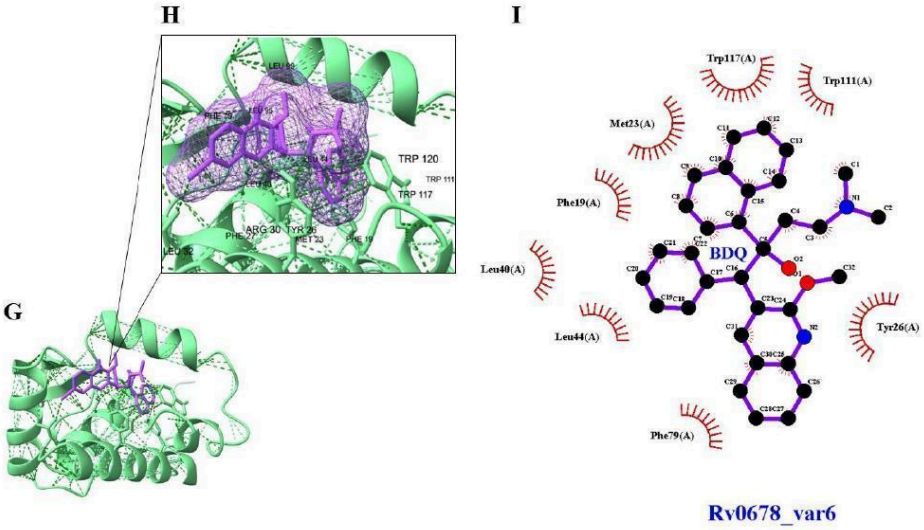


FIGURA 03 | Complexos estruturais *Rv0678*_var1, *Rv0678*_var2 e *Rv0678*_var6-BDQ do gene *Rv0678*. Afinidade molecular em 3D (A, B, D, E, G, H). Interações em 2D (C, F, I) entre os sítios-alvo e a BDQ por ligações hidrofóbicas são representadas pelos semicírculos ciliados de cor vermelha. Os círculos pretos, azuis e vermelhos correspondem, respectivamente, a átomos de Carbono, Nitrogênio e Oxigênio. Imagem criada utilizando os programas ChimeraX 1.8 e LigPlot⁺ 2.2 (da esquerda para a direita).

4 Discussão

Os resultados da análise de interação molecular entre a BDQ e as proteínas-alvo do *M. tuberculosis* revelaram variações significativas nos parâmetros de ΔG e Ki entre as formas selvagens (*wild*) e as variantes mutantes (var). Tais variações podem indicar tanto resistência quanto sensibilidade aumentada ao fármaco, a partir da análise das mutações não-sinônimas específicas presentes nas proteínas mutadas (Nehru *et al*., 2024).

Para *atpE*, observou-se que as variantes exibem comportamentos diferentes: ***atpE*_var1** sugere uma ligação mais forte com a BDQ, evidenciada pelos valores de ΔG e Ki bem como pela interação por ponte de hidrogênio com o aminoácido Glu e, portanto, possivelmente uma mutação que confere maior sensibilidade ao fármaco (apesar de um valor de CIM>1, o qual está relacionado ao fato de a resistência ser multifatorial, e não isolada). Em suma, a natureza da resistência à BDQ no gene *atpE* está relacionada a origem da variabilidade genética, vias metabólicas alternativas, influências ambientais, mecanismos de resistência concomitantes e implicações clínicas (Mushtaq *et al*., 2023; Karmakar *et al.*, 2019).

Já ***atpE*_var2** com CIM igual a 2,0 claramente apresenta uma redução significativa na afinidade com o fármaco, sendo necessária uma maior concentração de ligante para inibição do alvo, portanto, uma variante com potencial relação a resistência, refletida em seu valor de CIM. Nesta variante ocorreu também uma substituição de Glu por Asp na posição 61. Os dois aminoácidos têm cadeias laterais curtas carregadas negativamente, embora o Asp possua cadeia lateral ainda mais curta. Essa mudança pode impactar a conformação da proteína e a interação com a BDQ, alterando a dinâmica de ligação e possivelmente influenciando a resistência ao fármaco (Nehru *et al*., 2024; Mushtaq *et al*., 2023).

No caso da macromolécula *Rv0678* (gene com variantes mais associadas à resistência de acordo com catálogo de mutações no CMTB da OMS), ***Rv0678*_var1**, ***Rv0678*_var2** e ***Rv0678*_var6** apresentaram um padrão interessante. As variantes var1 e var2, com altos valores de CIM e Ki, indicaram resistência à BDQ, refletida na diminuição da eficácia do fármaco. No caso da ***Rv0678*_var1**, a substituição de Met por Ile na posição 139 troca dois aminoácidos hidrofóbicos semelhantes, mas a Met de cadeia lateral flexível difere da Ile que possui uma cadeia lateral mais rígida e ramificada. Isso pode alterar a flexibilidade da proteína, modificar a conformação do sítio de ligação, e afetar a interação com a droga, favorecendo a resistência (WHO, 2023).

Já ***Rv0678*_var2** e ***Rv0678*_var6** apresentaram proteínas truncadas com perdas significativas de aminoácidos, o que pode eliminar domínios críticos para a estabilização da proteína e comprometer a sua funcionalidade, contribuindo para a resistência. ***Rv0678*_var3**, ***Rv0678*_var4** e ***Rv0678*_var5** ainda mantêm uma boa afinidade, mas com CIM indicando resistência bacteriana, sugerindo que embora a afinidade seja alta, essas variantes são ainda presentes em bactérias resistentes. Tal fato pode ser explicado devido a resistência à BDQ no gene *Rv0678* ser complexa e apoiada pela presença de variantes pré-existentes a introdução da BDQ, presença de função regulatória e interações epistáticas (WHO, 2023; Nimmo *et al*., 2024; Saeed *et al*., 2022; Snobre *et al*., 2023).

A partir do exposto, em que mutações nos genes *atpE* e *Rv0678* desempenham papéis críticos na resistência à BDQ, interferindo em vias essenciais do CMTB, observou-se que as variantes em *atpE* e *Rv0678* ocorreram de forma isolada nas amostras em que foram relatadas, sem evidências de ocorrência simultânea. Isso sugere que a ausência de mutações acumulativas pode favorecer a adaptabilidade e a transmissão bacteriana, já que múltiplas mutações podem gerar um efeito deletério, comprometendo a aptidão da bactéria. Estudos anteriores indicam que a sinergia entre mutações pode ser prejudicial à viabilidade do organismo, reforçando que mutações isoladas são mais vantajosas ao balancear resistência e aptidão. Assim, as mutações isoladas em *atpE* e *Rv0678*, reforçam a hipótese de que a ausência de sinergia entre mutações oferece uma vantagem adaptativa na resistência à BDQ (Knopp e Andersson, 2018).

Finalmente para o gene *Rv1979c*, a variante ***Rv1979c*_var2** demonstra uma ligação mais forte à BDQ se comparada à forma selvagem, e seu Ki se comparado ao da proteína selvagem é menor, o que demonstra variante candidata a conferir sensibilidade à droga. Já em ***Rv1979c*_var1**, apesar do valor de CIM>2 caracterizar a resistência na bactéria, na proteína em específico o seu ΔG e Ki demonstram maior afinidade ao fármaco. Vale ressaltar que a CIM ainda aponta para a resistência nestas variantes, assim, é algo que deve ser melhor explorado, uma vez que o fenômeno da resistência é multifatorial (Kadura *et al*., 2020; Sonnenkalb *et al*., 2023).

Nos dados relacionados à interação entre a BDQ e *lpqB*, observa-se que a variante ***lpqB*_var1** exibe uma maior afinidade da BDQ pela variante, o que pode indicar uma alteração na sensibilidade ao fármaco. No entanto, o valor de CIM acima de 2 para a variante sugere que, apesar da maior afinidade, pode haver uma resistência emergente à BDQ, sugerindo que o mecanismo de resistência não é mediado exclusivamente por modificações em ***lpqB*_var1**. Essa discrepância entre a afinidade molecular e o perfil fenotípico de resistência pode estar relacionada a outros fatores moleculares que não apenas interações diretas com a droga, mas também mecanismos regulatórios transcricionais que modulam a expressão de genes associados ao metabolismo ou transporte do fármaco (Li *et al*., 2022; Nguyen *et al*., 2010).

Assim, a modulação da expressão gênica através da lipoproteína no sistema *mtrA/B* pode levar a mudanças na expressão de genes associados ao metabolismo e transporte de drogas, contribuindo para o fenótipo geral de resistência. Esses fatores, muitas vezes sistêmicos, podem reduzir a concentração intracelular efetiva do fármaco, anulando o benefício de uma ligação aparentemente estável com o alvo proteico. Esses resultados reforçam a necessidade de análises integrativas que combinem estudos de interação proteína-ligante com investigações de outros mecanismos de

resistência fenotípica, como o papel de reguladores globais e sistemas de efluxo (Li *et al*., 2022; Nguyen *et al*., 2010).

Foi também observado que as proteínas *mmpL5* e *mtrB* não exibiram interações significativas com a BDQ, apresentando energias de ligação positivas, indicando que as variantes encontradas nesses genes, na presente pesquisa, não foram alvos diretos do fármaco. Assim, infere-se que as mutações encontradas nas proteínas de *mmpL5* e *mtrB* não afetaram significativamente a interação com a BDQ (Hartkoorn *et al*., 2014). Acerca do gene *pepQ*, sua expressão foi completamente silenciada devido à presença de uma mutação de perda de sentido, logo após o códon de início, resultando no seu silenciamento gênico completo (Almeida *et al*., 2016).

Outrossim, um achado particularmente notável do estudo foi a interação da BDQ com o repressor *Rv0678* do sistema *mmpL5-mmpS5*. Embora a BDQ interaja com *Rv0678*, o sistema em si não demonstrou atuação direta com o fármaco, o que destaca uma nuance importante na compreensão dos mecanismos de resistência dentre as variantes analisadas. Esse dado sugere que, enquanto *Rv0678* desempenha um papel na resistência através do repressor, o sistema *mmpL5-mmpS5* não foi diretamente afetado pelo fármaco, o que abre novas perspectivas para estudos futuros sobre a função desse sistema na resistência à droga (Saeed *et al*., 2022; Hartkoorn *et al*., 2014).

Com base nos resultados da análise computacional, o estudo apresentou variações significativas em *Rv0678* e *atpE*, corroborando com a literatura e com dados do catálogo da OMS, onde essas variantes são as melhores descritas. Assim, dados dessas simulações *in silico* ajudam a buscar explicar como a resistência se desenvolve em cada um dos produtos estruturais desses genes e indicam que, apesar das interações específicas observadas em cada proteína, a resistência à BDQ é um fenômeno complexo e heterogêneo, exigindo uma análise aprofundada para sua compreensão (WHO, 2023; Nehru *et al*., 2024).

5 Conclusão

Conclui-se que as variantes ***atpE*_var2, *Rv0678*_var1, *Rv0678*_var2** e ***Rv0678*_var6** desempenham um papel como determinantes potenciais da resistência à BDQ, devido à redução de afinidade com o fármaco, demonstrada pela avaliação da energia de ligação e constante de inibição. Em contraste, a aquisição de resistência por mutações nos genes *mmpL5* e *mtrB* não demonstrou impacto significativo na interação com o fármaco, sugerindo que esses genes não desempenharam um papel direto na resistência mediada por BDQ. As demais variantes com CIM associados a resistência à BDQ podem atuar por meio de outros mecanismos, que não envolvem a interação direta entre a proteína e o fármaco. Além disso, o banco de dados local e o fluxo de análise de mutações e variantes desenvolvidos foram validados e representam ferramentas importantes para futuros experimentos de docagem.

Este estudo permitiu caracterizar variantes que demonstram influência significativa na eficácia da BDQ, sendo crucial para entender a resistência a droga em *M. tuberculosis*. Também forneceu dados relevantes para a identificação *in silico* de variantes e destacou a importância de investigações contínuas. Essas descobertas apresentadas transcendem o campo teórico, pois oferecem aplicação prática clara ao monitoramento da resistência, com potencial de integrar estratégias de saúde pública para o monitoramento contínuo da resistência e a contenção da TB resistente.

O procedimento bioinformático demonstrou eficiência na avaliação em massa de variantes, utilizando tecnologias de inteligência artificial acelerada por GPU, o que posiciona este projeto como inovador devido ao uso de tecnologias do estado da arte para a investigação de agravos endêmicos e negligenciados. Ele também sugere futuras modelagens de outras proteínas, estudos de dinâmica molecular e inclusão de novos fármacos. Tais avanços possibilitam o aprimoramento dos esquemas terapêuticos, como a adição de fármacos capazes de contornar a resistência no sítio alterado. A

metodologia também pode avaliar a eficácia de compostos naturais como possíveis fármacos anti-TB, fortalecendo as bases científicas para o combate à resistência da TB.

6 Declaração de Disponibilidade de Dados

Os dados de sequenciamento brutos (arquivos FASTQ) das variantes analisadas no estudo estão incluídas na Tabela Suplementar 2.

7 Financiamento

IEC/SVSA/MS, CNPq e CAPES. Bolsa PIBIC/IEC CNPq, chamada 01/2023; Bolsa CAPES, chamada PPGBPA 120/2022; Bolsa Pq/CNPq, chamada 09/2022.

8 Agradecimentos

Nós agradecemos ao IEC/SVSA/MS pela estrutura e, em especial, à Central de Serviços da Informática pela manutenção do serviço de computação de alto desempenho.

9 Material suplementar

Tabela Suplementar 1: Gráficos Ramachandran de proteínas selvagens e variantes de *Rv0678*, *atpE*, *mtrB*, *mmpL5*, *Rv1979c* e *lpqB*.

Tabela Suplementar 2: *Run Accession* de cada amostra para consulta de dados de sequenciamento brutos (arquivos FASTQ).

10 Referências

Almeida, D., Ioerger, T., Tyagi, S., Li, S. Y., Mdluli, K., Andries, K., Grosset, J., Sacchettini, J., & Nuermberger, E. (2016). Mutations in pepQ Confer Low-Level Resistance to Bedaquiline and Clofazimine in Mycobacterium tuberculosis. Antimicrobial agents and chemotherapy, 60(8), 4590–4599. https://doi.org/10.1128/AAC.00753-16.

Arora, G., Bothra, A., Prosser, G., Arora, K., & Sajid, A. (2021). Role of post-translational modifications in the acquisition of drug resistance in Mycobacterium tuberculosis. The FEBS journal, 288(11), 3375–3393. https://doi.org/10.1111/febs.15582.

Bai, Q., Li, L., Liu, S., Xiao, S., & Guo, Y. (2018). Drug design progress of in silico, in vitro and in vivo researches. In-vitro In-vivo In-silico Journal, 1(1), 16-37.

Barilar, I., Fernando, T., Utpatel, C., Abujate, C., Madeira, C. M., José, B., Mutaquiha, C., Kranzer, K., Niemann, T., Ismael, N., de Araujo, L., Wirth, T., Niemann, S., & Viegas, S. (2024). Emergence of bedaquiline-resistant tuberculosis and of multidrug-resistant and extensively drug-resistant Mycobacterium tuberculosis strains with rpoB Ile491Phe mutation not detected by Xpert MTB/RIF in Mozambique: a retrospective observational study. The Lancet. Infectious diseases, 24(3), 297–307. https://doi.org/10.1016/S1473-3099(23)00498-X.

410 Berman, H. M., Kleywegt, G. J., Nakamura, H., & Markley, J. L. (2014). The Protein Data Bank
411 archive as an open data resource. Journal of computer-aided molecular design, 28(10), 1009–1014.
412 https://doi.org/10.1007/s10822-014-9770-y.

413 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence
414 data. Bioinformatics (Oxford, England), 30(15), 2114–2120.
415 https://doi.org/10.1093/bioinformatics/btu170.

416 The CRyPTIC Consortium (2022). A data compendium associating the genomes of 12,289
417 Mycobacterium tuberculosis isolates with quantitative resistance phenotypes to 13 antibiotics. PLoS
418 biology, 20(8), e3001721. https://doi.org/10.1371/journal.pbio.3001721.

419 Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane,
420 T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools.
421 GigaScience, 10(2), giab008. https://doi.org/10.1093/gigascience/giab008.

422 Derendinger, B., Dippenaar, A., de Vos, M., Huo, S., Alberts, R., Tadokera, R., Limberis, J., Sirgel,
423 F., Dolby, T., Spies, C., Reuter, A., Folkerts, M., Allender, C., Lemmer, D., Van Rie, A., Gagneux, S.,
424 Rigouts, L., Te Riele, J., Dheda, K., Engelthaler, D. M., … Theron, G. (2023). Bedaquiline resistance
425 in patients with drug-resistant tuberculosis in Cape Town, South Africa: a retrospective longitudinal
426 cohort study. The Lancet. Microbe, 4(12), e972–e982. https://doi.org/10.1016/S2666-5247(23)00172-
427 6.

428 Ejalonibu, M. A., Ogundare, S. A., Elrashedy, A. A., Ejalonibu, M. A., Lawal, M. M., Mhlongo, N.
429 N., & Kumalo, H. M. (2021). Drug Discovery for Mycobacterium tuberculosis Using Structure-Based
430 Computer-Aided Drug Design Approach. International journal of molecular sciences, 22(24), 13259.
431 https://doi.org/10.3390/ijms222413259.

432 Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: summarize analysis results for
433 multiple tools and samples in a single report. Bioinformatics (Oxford, England), 32(19), 3047–3048.
434 https://doi.org/10.1093/bioinformatics/btw354.

435 Fowler, P. W., Wright, C., Spiers, H., Zhu, T., Baeten, E. M. L., Hoosdally, S. W., Gibertoni Cruz, A.
436 L., Roohi, A., Kouchaki, S., Walker, T. M., Peto, T. E. A., Miller, G., Lintott, C., Clifton, D., Crook,
437 D. W., Walker, A. S., Zooniverse Volunteer Community, & CRyPTIC Consortium (2022). A crowd
438 of BashTheBug volunteers reproducibly and accurately measure the minimum inhibitory
439 concentrations of 13 antitubercular drugs from photographs of 96-well broth microdilution plates.
440 eLife, 11, e75046. https://doi.org/10.7554/eLife.75046.

441 Ghersi, D., & Sanchez, R. (2009). Improving accuracy and efficiency of blind protein-ligand docking
442 by focusing on predicted binding sites. Proteins, 74(2), 417–424. https://doi.org/10.1002/prot.22154.

443 Hartkoorn, R. C., Uplekar, S., & Cole, S. T. (2014). Cross-resistance between clofazimine and
444 bedaquiline through upregulation of MmpL5 in Mycobacterium tuberculosis. Antimicrobial agents and
445 chemotherapy, 58(5), 2979–2981. https://doi.org/10.1128/AAC.00037-14.

446 Hatami, H., Sotgiu, G., Bostanghadiri, N., Abadi, S. S. D., Mesgarpour, B., Goudarzi, H., Migliori,
447 G. B., & Nasiri, M. J. (2022). Bedaquiline-containing regimens and multidrug-resistant tuberculosis:
448 a systematic review and meta-analysis. Jornal brasileiro de pneumologia : publicacao oficial da
449 Sociedade Brasileira de Pneumologia e Tisilogia, 48(2), e20210384. https://doi.org/10.36416/1806-
450 3756/e20210384.

451 Heupink, T. H., Verboven, L., Sharma, A., Rennie, V., de Diego Fuertes, M., Warren, R. M., & Van
452 Rie, A. (2023). The MAGMA pipeline for comprehensive genomic analyses of clinical Mycobacterium
453 tuberculosis samples. PLoS computational biology, 19(11), e1011648.
454 https://doi.org/10.1371/journal.pcbi.1011648.

455 Islam, M. M., Alam, M. S., Liu, Z., Khatun, M. S., Yusuf, B., Hameed, H. M. A., Tian, X., Chhotaray,
456 C., Basnet, R., Abraha, H., Zhang, X., Khan, S. A., Fang, C., Li, C., Hasan, S., Tan, S., Zhong, N., Hu,
457 J., & Zhang, T. (2024). Molecular mechanisms of resistance and treatment efficacy of clofazimine and
458 bedaquiline against Mycobacterium tuberculosis. Frontiers in medicine, 10, 1304857.
459 https://doi.org/10.3389/fmed.2023.1304857.

460 Ismail, N., Rivière, E., Limberis, J., Huo, S., Metcalfe, J. Z., Warren, R. M., & Van Rie, A. (2021).
461 Genetic variants and their association with phenotypic resistance to bedaquiline in Mycobacterium
462 tuberculosis: a systematic review and individual isolate data analysis. The Lancet. Microbe, 2(11),
463 e604–e616. https://doi.org/10.1016/s2666-5247(21)00175-0.

464 Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K.,
465 Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie,
466 A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., … Hassabis, D. (2021). Highly
467 accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583–589.
468 https://doi.org/10.1038/s41586-021-03819-2.

469 Kadura, S., King, N., Nakhoul, M., Zhu, H., Theron, G., Köser, C. U., & Farhat, M. (2020). Systematic
470 review of mutations associated with resistance to the new and repurposed Mycobacterium tuberculosis
471 drugs bedaquiline, clofazimine, linezolid, delamanid and pretomanid. The Journal of antimicrobial
472 chemotherapy, 75(8), 2031–2043. https://doi.org/10.1093/jac/dkaa136.

473 Karmakar, M., Rodrigues, C. H. M., Holt, K. E., Dunstan, S. J., Denholm, J., & Ascher, D. B. (2019).
474 Empirical ways to identify novel Bedaquiline resistance mutations in AtpE. PloS one, 14(5), e0217169.
475 https://doi.org/10.1371/journal.pone.0217169.

476 Knopp, M., & Andersson, D. I. (2018). Predictable Phenotypes of Antibiotic Resistance Mutations.
477 mBio, 9(3), e00770-18. https://doi.org/10.1128/mBio.00770-18.

478 Lange, C., Chesov, D., Heyckendorf, J., Leung, C. C., Udwadia, Z., & Dheda, K. (2018). Drug-
479 resistant tuberculosis: An update on disease burden, diagnosis and treatment. Respirology (Carlton,
480 Vic.), 23(7), 656–673. https://doi.org/10.1111/resp.13304.

Laskowski, R. A., & Swindells, M. B. (2011). LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. Journal of chemical information and modeling, 51(10), 2778–2786. https://doi.org/10.1021/ci200227u.

Li, S., Poulton, N. C., Chang, J. S., Azadian, Z. A., DeJesus, M. A., Ruecker, N., Zimmerman, M. D., Eckartt, K. A., Bosch, B., Engelhart, C. A., Sullivan, D. F., Gengenbacher, M., Dartois, V. A., Schnappinger, D., & Rock, J. M. (2022). CRISPRi chemical genetics and comparative genomics identify genes mediating drug potency in Mycobacterium tuberculosis. Nature microbiology, 7(6), 766–779. https://doi.org/10.1038/s41564-022-01130-y.

Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., Madhusoodanan, N., Kolesnikov, A., & Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. Nucleic acids research, 50(W1), W276–W279. https://doi.org/10.1093/nar/gkac240.

Mahajan R. (2013). Bedaquiline: First FDA-approved tuberculosis drug in 40 years. International journal of applied & basic medical research, 3(1), 1–2. https://doi.org/10.4103/2229-516X.112228.

Makhado, N. A., Matabane, E., Faccin, M., Pinçon, C., Jouet, A., Boutachkourt, F., Goeminne, L., Gaudin, C., Maphalala, G., Beckert, P., Niemann, S., Delvenne, J. C., Delmée, M., Razwiedani, L., Nchabeleng, M., Supply, P., de Jong, B. C., & André, E. (2018). Outbreak of multidrug-resistant tuberculosis in South Africa undetected by WHO-endorsed commercial tests: an observational study. The Lancet. Infectious diseases, 18(12), 1350–1359. https://doi.org/10.1016/S1473-3099(18)30496-1.

Meng, E. C., Goddard, T. D., Pettersen, E. F., Couch, G. S., Pearson, Z. J., Morris, J. H., & Ferrin, T. E. (2023). UCSF ChimeraX: Tools for structure building and analysis. Protein science : a publication of the Protein Society, 32(11), e4792. https://doi.org/10.1002/pro.4792.

Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. Journal of computational chemistry, 30(16), 2785–2791. https://doi.org/10.1002/jcc.21256.

Mushtaq, A., Wani., Devendra, K., Dhaked. (2023). An update on ATP synthase inhibitors: A unique target for drug development in M. tuberculosis.. Progress in Biophysics & Molecular Biology, 180-181:87-104. https://doi.org/10.1016/j.pbiomolbio.2023.04.009.

Nehru, V., Brammacharry, U., Ramkumar, S., Khusro, A., Vadakunnel, M. J., Gunasekaran, S., David, E., Priya, V., Aljowaie, R. M., & Almutairi, S. M. (2024). Molecular analysis of bedaquiline resistance and genetic variations in clinical isolates of multidrug and fluoroquinolones-resistant Mycobacterium tuberculosis. Journal of King Saud University - Science, 103226. https://doi.org/10.1016/j.jksus.2024.103226.

Nguyen, N. T., Pham, H. P., Cao, C. M., & Hoang, A. D. (2024). The Development of Research on AI and Cost Saving from 1990 to 2024: A Bibliometrics Review from the Web of Science Database. https://doi.org/10.31219/osf.io/sp4gq.

517 Nguyen, H. T., Wolff, K. A., Cartabuke, R. H., Ogwang, S., & Nguyen, L. (2010). A lipoprotein
518 modulates activity of the MtrAB two-component system to provide intrinsic multidrug resistance,
519 cytokinetic control and cell wall homeostasis in Mycobacterium. Molecular microbiology, *76*(2),
520 348-364.

521 Nguyen, T. V. A., Anthony, R. M., Bañuls, A. L., Nguyen, T. V. A., Vu, D. H., & Alffenaar, J. C.
522 (2018). Bedaquiline Resistance: Its Emergence, Mechanism, and Prevention. Clinical infectious
523 diseases : an official publication of the Infectious Diseases Society of America, 66(10), 1625–1630.
524 https://doi.org/10.1093/cid/cix992.

525 Nielsen, H., Teufel, F., Brunak, S., & von Heijne, G. (2024). SignalP: The Evolution of a Web Server.
526 Methods in molecular biology (Clifton, N.J.), 2836, 331–367. https://doi.org/10.1007/978-1-0716-
527 4007-4_17.

528 Nimmo, C., Ortiz, A. T., Tan, C. C. S., Pang, J., Acman, M., Millard, J., Padayatchi, N., Grant, A. D.,
529 O'Donnell, M., Pym, A., Brynildsrud, O. B., Eldholm, V., Grandjean, L., Didelot, X., Balloux, F., &
530 van Dorp, L. (2024). Detection of a historic reservoir of bedaquiline/clofazimine resistance-associated
531 variants in Mycobacterium tuberculosis. Genome medicine, 16(1), 34. https://doi.org/10.1186/s13073-
532 024-01289-5.

533 Nimmo, C., Bionghi, N., Cummings, M. J., Perumal, R., Hopson, M., Al Jubaer, S., Naidoo, K., Wolf,
534 A., Mathema, B., Larsen, M. H., & O'Donnell, M. (2024). Opportunities and limitations of genomics
535 for diagnosing bedaquiline-resistant tuberculosis: a systematic review and individual isolate meta-
536 analysis. The Lancet. Microbe, 5(2), e164–e172. https://doi.org/10.1016/S2666-5247(23)00317-8.

537 Perumal, R., Bionghi, N., Nimmo, C., Letsoalo, M., Cummings, M. J., Hopson, M., Wolf, A., Jubaer,
538 S. A., Padayatchi, N., Naidoo, K., Larsen, M. H., & O'Donnell, M. (2023). Baseline and treatment-
539 emergent bedaquiline resistance in drug-resistant tuberculosis: a systematic review and meta-
540 analysis. The European respiratory journal, 62(6), 2300639. https://doi.org/10.1183/13993003.00639-
541 2023.

542 Peterson, E. J. R., Brooks, A. N., Reiss, D. J., Kaur, A., Do, J., Pan, M., Wu, W. J., Morrison, R.,
543 Srinivas, V., Carter, W., Arrieta-Ortiz, M. L., Ruiz, R. A., Bhatt, A., & Baliga, N. S. (2023). MtrA
544 modulates Mycobacterium tuberculosis cell division in host microenvironments to mediate intrinsic
545 resistance and drug tolerance. Cell reports, 42(8), 112875.
546 https://doi.org/10.1016/j.celrep.2023.112875.

547 Plocinska, R., Purushotham, G., Sarva, K., Vadrevu, I. S., Pandeeti, E. V., Arora, N., Plocinski, P.,
548 Madiraju, M. V., & Rajagopalan, M. (2012). Septal localization of the Mycobacterium tuberculosis
549 MtrB sensor kinase promotes MtrA regulon expression. The Journal of biological chemistry,
550 287(28), 23887–23899. https://doi.org/10.1074/jbc.M112.346544.

551 Radhakrishnan, A., Kumar, N., Wright, C. C., Chou, T. H., Tringides, M. L., Bolla, J. R., Lei, H. T.,
552 Rajashankar, K. R., Su, C. C., Purdy, G. E., & Yu, E. W. (2014). Crystal structure of the

553 transcriptional regulator Rv0678 of Mycobacterium tuberculosis. The Journal of biological
554 chemistry, 289(23), 16526–16540. https://doi.org/10.1074/jbc.M113.538959.

555 Sachan, R. S. K., Mistry, V., Dholaria, M., Rana, A., Devgon, I., Ali, I., Iqbal, J., Eldin, S. M.,
556 Mohammad Said Al-Tawaha, A. R., Bawazeer, S., Dutta, J., & Karnwal, A. (2023). Overcoming
557 Mycobacterium tuberculosis Drug Resistance: Novel Medications and Repositioning Strategies. ACS
558 omega, 8(36), 32244–32257. https://doi.org/10.1021/acsomega.3c02563.

559 Saeed, D. K., Shakoor, S., Razzak, S. A., Hasan, Z., Sabzwari, S. F., Azizullah, Z., Kanji, A., Nasir,
560 A., Shafiq, S., Ghanchi, N. K., & Hasan, R. (2022). Variants associated with Bedaquiline (BDQ)
561 resistance identified in Rv0678 and efflux pump genes in Mycobacterium tuberculosis isolates from
562 BDQ naïve TB patients in Pakistan. BMC microbiology, 22(1), 62. https://doi.org/10.1186/s12866-
563 022-02475-4.

564 Seelam, B. K., Kumari, S. L., Sanjana, P. H., Sudheera, N., Jahnavi, D., Harika, A., & Thomas, R. T.
565 (2020). Outcomes of Bedaquiline Treatment in Patients with Multidrug-Resistant–Tuberculosis. Int J
566 Pharm Sci Invention, 9, 37-42.

567 Snobre, J., Villellas, M. C., Coeck, N., Mulders, W., Tzfadia, O., de Jong, B. C., Andries, K., &
568 Rigouts, L. (2023). Bedaquiline- and clofazimine- selected Mycobacterium tuberculosis mutants:
569 further insights on resistance driven largely by Rv0678. Scientific reports, 13(1), 10444.
570 https://doi.org/10.1038/s41598-023-36955-y.

571 Sonnenkalb, L., Carter, J. J., Spitaleri, A., Iqbal, Z., Hunt, M., Malone, K. M., Utpatel, C., Cirillo, D.
572 M., Rodrigues, C., Nilgiriwala, K. S., Fowler, P. W., Merker, M., Niemann, S., & Comprehensive
573 Resistance Prediction for Tuberculosis: an International Consortium (2023). Bedaquiline and
574 clofazimine resistance in Mycobacterium tuberculosis: an in-vitro and in-silico data analysis. The
575 Lancet. Microbe, 4(5), e358–e368. https://doi.org/10.1016/S2666-5247(23)00002-2.

576 Verli, H. (2014). Bioinformática: da biologia à flexibilidade molecular (1ª Edição). São Paulo:
577 Sociedade Brasileira de Bioquímica e Biologia Molecular – SBBq.

578 WHO (2023). Catalogue of mutations in Mycobacterium tuberculosis complex and their association
579 with drug resistance, second edition. Geneva: World Health Organization - WHO.

580 WHO (2023). Global Tuberculosis Report 2023. Geneva: World Health Organization - WHO.

581 WHO (2024). Global Tuberculosis Report 2024. Geneva: World Health Organization - WHO.

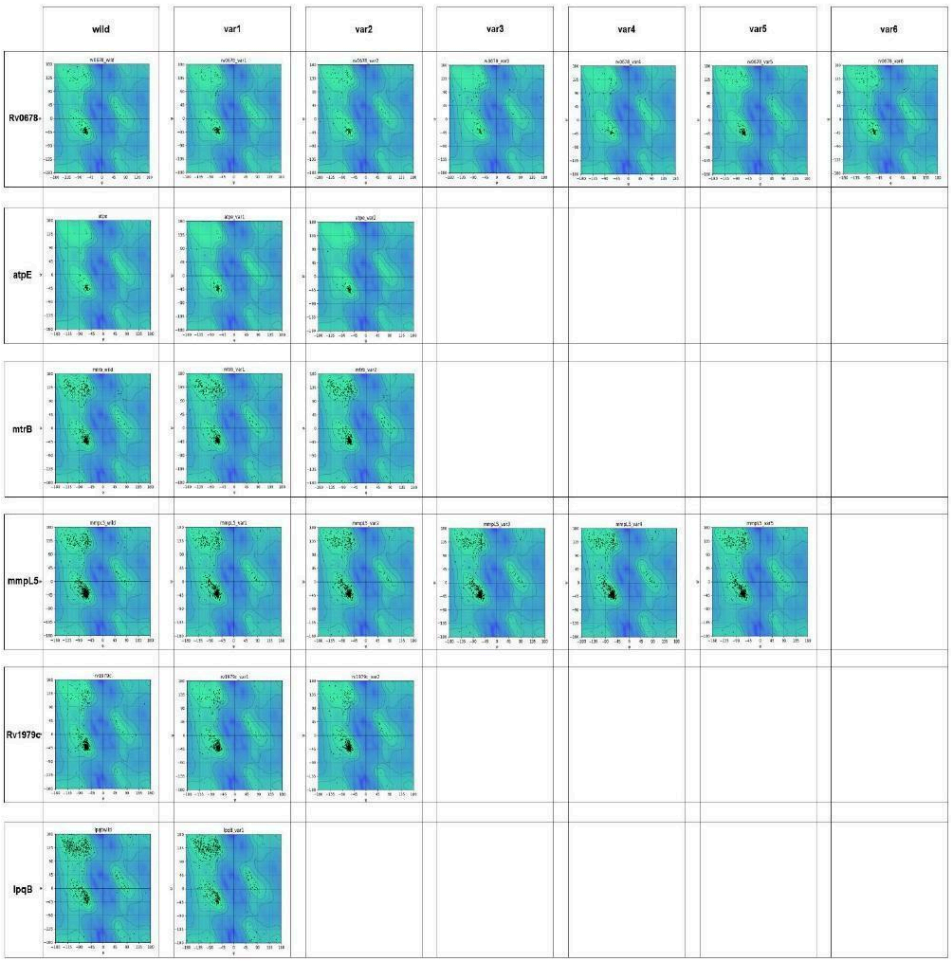
Material Suplementar

Análise estrutural *in silico* baseada em modelagem molecular: Explorando variantes de resistência à bedaquilina em *Mycobacterium tuberculosis*

Danna Karen Corrêa dos Santos[1,2], Davi Josué Marcon[1,2], Karla Valéria Batista Lima[1,2]

Correspondência: Danna Karen Corrêa dos Santos (dannakcsantos@gmail.com), Davi Marcon (davijosuemarcon@gmail.com), Karla Valéria Batista Lima (karlavaleria2007@gmail.com)

TABELA SUPLEMENTAR 1 | Gráficos Ramachandran de proteínas selvagens e variantes de *Rv0678*, *atpE*, *mtrB*, *mmpL5*, *Rv1979c* e *lpqB*.



Material Suplementar

Análise estrutural *in silico* baseada em modelagem molecular: Explorando variantes de resistência à bedaquilina em *Mycobacterium tuberculosis*

Danna Karen Corrêa dos Santos[1,2], Davi Josué Marcon[1,2], Karla Valéria Batista Lima[1,2]

Correspondência: Danna Karen Corrêa dos Santos (dannakcsantos@gmail.com), Davi Marcon (davijosuemarcon@gmail.com), Karla Valéria Batista Lima (karlavaleria2007@gmail.com)

TABELA SUPLEMENTAR 2 | *Run Accession* de cada amostra para consulta de dados de sequenciamento brutos (arquivos FASTQ).

Amostra	MIC	*atpE*	*pgB*	*mtrA*	*mtrB*	*mmpL5*	*mmpS5*	*Rv0678 - mmpR5*	*Rv1979c*	*pepQ*
CryPTIC_R_bdq_ERX4701151_ERR4831167	>1	var1	wild	wild	var1	var2	wild	wild	wild	wild
CryPTIC_R_bdq_ERX8324368_ERR8665937	1.0	wild	wild	wild	var2	var3	wild	wild	wild	wild
CryPTIC_R_bdq_ERX4681968_ERR4811978	>1	wild	wild	wild	var2	var3	wild	wild	wild	wild
CryPTIC_R_bdq_ERX8324398_ERR8665967	1.0	wild	wild	wild	var1	var2	wild	wild	wild	wild
CryPTIC_R_bdq_ERX4665671_ERR4796257	>2	wild	wild	wild	var1	var1	wild	wild	wild	wild
CryPTIC_R_bdq_ERX8324376_ERR8665945	1.0	wild	wild	wild	var1	var2	wild	wild	wild	wild
CryPTIC_R_bdq_ERX5544059_ERR8975831	1.0	wild	wild	wild	var1	var2	wild	wild	wild	wild
CryPTIC_R_bdq_ERX5544148_ERR8975920	1.0	wild	wild	wild	var2	var1	wild	var4	var2	wild
CryPTIC_R_bdq_ERX4666873_ERR4797457	1.0	wild	wild	wild	wild	var3	wild	wild	wild	wild
CryPTIC_R_bdq_ERX4700759_ERR4830775	1.0	wild	wild	wild	var2	var3	wild	wild	wild	wild
CryPTIC_R_bdq_ERX4701498_ERR4831514	1.0	wild	wild	wild	var2	var1	wild	wild	wild	wild
CryPTIC_R_bdq_ERX4699126_ERR4829142	1.0	wild	wild	wild	var1	var1	wild	wild	wild	wild
CryPTIC_R_bdq_ERX8324369_ERR8665938	1.0	wild	wild	wild	var2	var3	wild	wild	wild	wild
CryPTIC_R_bdq_ERX8324409_ERR8665978	1.0	wild	wild	wild	var1	var2	wild	wild	wild	wild
CryPTIC_R_bdq_ERX4678965_ERR4810519	>2	wild	wild	wild	var2	var3	wild	wild	wild	wild
CryPTIC_R_bdq_ERX4699151_ERR4829167	>2	wild	wild	wild	var2	var1	wild	var2	wild	wild
CryPTIC_R_bdq_ERX4684031_ERR4814041	>2	wild	wild	wild	var2	var3	wild	wild	wild	wild
CryPTIC_R_bdq_ERX8324397_ERR8665966	1.0	wild	wild	wild	var2	var4	wild	wild	wild	wild
CryPTIC_R_bdq_ERX5543946_ERR8975716	1.0	wild	wild	wild	var2	var1	wild	var3	wild	wild
CryPTIC_R_bdq_ERX5543987_ERR8975757	1.0	wild	wild	wild	var2	var1	wild	wild	wild	wild
CryPTIC_R_bdq_ERX8324414_ERR8665983	1.0	wild	wild	wild	var1	var2	wild	wild	wild	wild
CryPTIC_R_bdq_ERX5559197_ERR5918105	1.0	wild	wild	wild	var2	var1	wild	var6	wild	wild
CryPTIC_R_bdq_ERX4700740_ERR4830756	>1	wild	wild	wild	var2	var3	wild	wild	wild	wild
CryPTIC_R_bdq_ERX8324395_ERR8665964	1.0	wild	wild	wild	var1	var2	wild	wild	wild	wild
CryPTIC_R_bdq_ERX8324365_ERR8665934	>1	wild	wild	wild	var1	var2	wild	wild	wild	wild
CryPTIC_R_bdq_ERX4666159_ERR4796744	>2	var1	wild	wild	var2	var1	wild	var1	var1	wild
CryPTIC_R_bdq_ERX4679255_ERR4810809	>2	wild	wild	wild	var1	var5	wild	wild	wild	wild
CryPTIC_R_bdq_ERX8324373_ERR8665942	1.0	wild	wild	wild	var1	var2	wild	wild	wild	wild
CryPTIC_R_bdq_ERX4701983_ERR4832040	1.0	wild	wild	wild	var2	var1	wild	wild	wild	wild
CryPTIC_R_bdq_ERX4666139_ERR4796724	>2	wild	wild	wild	var1	var1	wild	wild	wild	wild
CryPTIC_R_bdq_ERX4699486_ERR4829502	1.0	wild	wild	wild	var2	var3	wild	wild	wild	wild
CryPTIC_R_bdq_ERX8324371_ERR8665940	1.0	wild	wild	wild	var2	var3	wild	wild	wild	wild
CryPTIC_R_bdq_ERX5544168_ERR8975940	1.0	wild	wild	wild	var2	var1	wild	var5	var2	wild
CryPTIC_R_bdq_ERX4665861_ERR4796447	2.0	var2	wild	wild	var1	var2	wild	wild	wild	wild
CryPTIC_R_bdq_ERX5544272_ERR8976046	1.0	wild	wild	wild	var2	var1	wild	wild	var2	silenciamento completo do gene
CryPTIC_R_bdq_ERX8324364_ERR8665933	1.0	wild	wild	wild	var1	var2	wild	wild	wild	wild
CryPTIC_R_bdq_ERX4678603_ERR4809063	2.0	wild	wild	wild	var2	var3	wild	wild	wild	wild
CryPTIC_R_bdq_ERX4679255_ERR12119124	>2	wild	wild	wild	var1	var5	wild	wild	wild	wild

5.2. PRODUÇÃO TÉCNICA

5.2.1. REVIGET_POP001: Extração de DNA utilizando InstaGene Matrix e Controle de Qualidade de DNA

A partir da análise da rotina laboratorial estabelecida no LACEN-PA, foi possível desenvolver, testar e implementar um Procedimento Operacional Padrão (POP) específico para a extração de DNA baseado no kit Instagene Matrix adaptando a metodologia descrita por Conceição e colaboradores (2024). O documento, intitulado **REVIGET_POP001: Extração de DNA utilizando InstaGene Matrix e Controle de Qualidade de DNA**, foi elaborado para padronizar o processo e orientar os integrantes da rede na realização do processo, com o objetivo de minimizar o impacto da introdução de uma nova técnica na rotina já existente. Atualmente, esse protocolo é adotado no projeto da Rede de Vigilância Norte e Nordeste da Tuberculose, sendo empregado nos diferentes laboratórios da rede que foram treinados para sua utilização no segundo semestre de 2024.

Observou-se que a extração de DNA deveria ser inserida utilizando a cultura realizada para o primeiro Teste de Sensibilidade no LACEN-PA, obtendo-se o tubo indicador do controle de crescimento, o qual seria descartado após a realização do teste. Este tubo deve ser utilizado sem a realização de subcultivos, para evitar o viés de seleção de bacilos gerado por subcultivos.

Procedimento operacional padrão (POP) nº REVIGET_POP001:
EXTRAÇÃO DE DNA UTILIZANDO INSTAGENE MATRIX E CONTROLE DE QUALIDADE DE DNA
Versão #: 1.0 beta, Página 1 de 9

SUMÁRIO

ESCOPO	**1**
RESPONSABILIDADES	**1**
1. GERENCIAMENTO DE RISCO	**2**
2. ANTECEDENTES	**2**
3. MATERIAL	**3**
3.1. CONSUMÍVEIS	3
3.2. EQUIPAMENTOS	4
4. PROCEDIMENTO	**4**
4.1. Inativação por calor	4
4.2. Pré-tratamento do material inativado (para meio líquido)	5
4.3. Extração utilizando Instagene Matrix	6
4.4. Controle de qualidade da extração	7
5. ABREVIATURAS	**9**
6. REFERÊNCIAS	**9**
7. HISTÓRICO DO DOCUMENTO	**9**

CÓPIA CONTROLADA

ESCOPO

Este procedimento operacional padrão (POP) descreve o fluxo de trabalho de extração de DNA de *Mycobacterium tuberculosis* de meio líquido ou sólido utilizando Instagene Matrix (Bio-Rad), contendo também os procedimentos recomendados para controle de qualidade do DNA extraído. O procedimento foi adaptado a partir do protocolo desenvolvido por Conceição et al. (2024), disponível em dx.doi.org/10.17504/protocols.io.6qpvr8rwzlmk/v1.

RESPONSABILIDADES

É responsabilidade de cada membro do Grupo de Pesquisa **(adicionar nome do grupo de pesquisa)**:

- Seguir esse POP conforme escrito
- Tomar as precauções de segurança necessárias ao manusear amostras durante este processo.

É responsabilidade do pesquisador principal garantir que este procedimento seja apropriado para o uso pretendido e que todos os alunos/bolsistas de pós-doutorado e pesquisadores envolvidos neste processo sejam devidamente treinados antes de processar quaisquer amostras.

CÓPIA CONTROLADA

1. GERENCIAMENTO DE RISCO

Os riscos envolvidos ao executar o procedimento descrito no POP incluem a manipulação de microorganismos infecciosos do complexo *Mycobacterium tuberculosis* os quais devem ser evitados ou reduzidos com o uso de instalações do nível de biossegurança adequado para a manipulação (NB-3), além do uso de equipamento de proteção individual (jaleco, máscara com purificador de ar, luvas e protetores de calçado). Antes de sair da sala NB3 as amostras serão inativadas através de calor a 80° por 1 hora para a cultura líquida e 2 horas para cultura sólido.

Para evitar/reduzir esses riscos: É essencial o uso de equipamentos de proteção individual (EPI), incluindo proteção ocular, máscara, luvas e jaleco de laboratório. Além disso, é importante manusear adequadamente os reagentes usados como resíduos químicos e descartá-los de acordo com os padrões de segurança vigentes.

2. ANTECEDENTES

Amostras de *M. tuberculosis*, cultivado em meio sólido (LJ, ou 7H11) ou meio líquido (7H9/MGIT) ou de amostras de escarro serão submetidas a extração de DNA a fim de ter material genético de alta qualidade para a realização do sequenciamento genômico. A matriz InstaGene é um composto formulado a partir de uma resina quelante formada de uma solução 6% peso/volume de chelex, que pode ser utilizado para obtenção de material nucleico de alta qualidade, livre de contaminantes e integro para sequenciamento genômico.

3. MATERIAL

Produto	Código do produto
Agarose para biologia molecular	-
Água livre de Nucleases ou Ultrapura	-
Alças de inoculação de 10 µL	-
Culturas de *M. tuberculosis* em meio sólido (7H11, LJ) ou liquido (7H9/MGIT)	-
DNAse I	
Etanol PA ou para Biologia Molecular	-
Esferas de vidro de 1 ou 2 mm (idealmente 1 mm)	-
InstaGene Matrix	Bio-Rad, 7326030
Lenço de papel Kim Wipe (para limpar o pedestal do Nanodrop)	-
Ladder 1kb plus	Invitrogen, 10787018
Pipeta Pasteur descartável de 2 ou 3 mL	-
Ponteiras (1000 µL, 200 µL, 100 µL e 1-10 µL)	-
Qubit dsDNA HS Assay Kit	Invitrogen, Q32851
Qubit Assay Tubes	Invitrogen, Q32856
Qubit Flex Assay Tube Strips	Invitrogen, Q33252
Qubit 1X dsDNA HS Assay Kit	Invitrogen, Q33230
SYBR Safe DNA Gel Stain	Invitrogen, S33102
Tampão DNAse	
Tampão TAE 10X	Invitrogen, AM9869
Tubos de 1,5 ou 2 mL (baixa retenção)	-
Tubos de rosca de 1,5 ou 2 mL (baixa retenção)	-
Tubos de polipropileno de 15 mL	-

COPIA CONTROLADA

3.1. CONSUMÍVEIS

- Água livre de nuclease
- Alças de 10 µL se utilizando meio sólido
- Caneta marcador permanente
- Ponteiras de 200µL para P200
- Ponteiras de 20µL pontas para P20
- Ponteiras de 1000µL para P1000

- Ponteiras de 10µL (pontas LoRetention) para P10
- Toalhas de papel (para limpeza da área de trabalho)
- Tubos de microcentrífuga Eppendorf
- Tubos de polipropileno 15 mL
- Tubos de 1.5 ou 2 mL com tampa rosqueável(low-binding).
- Tubos de 1.5 ou mL (low-binding).
- Beads de vidro de 1 mm.

3.2. EQUIPAMENTOS

- Cabine de segurança biológica
- Centrífuga para tubos de 15mL
- Microcentrifuga para tubos de 2mL
- Forno com temperatura de 80° (dentro da área NB3)
- Termobloco com temperatura de 65° a 100°
- Pipeta P1000
- Pipeta P200
- Pipeta P20
- Pipeta P10
- Qubit
- Nanodrop

COPIA CONTROLADA

4. PROCEDIMENTO

4.1. Inativação por calor

Para minimizar os riscos de manipulação de *M. tuberculosis* devemos inativar as células por tratamento de calor antes de iniciar a extração, esse processo deve ser realizado em ambiente de biossegurança compatível.

4.1.1. Inativação por calor (meio sólido)

1. Adicione 3 beads de 2mm a tubos de 1,5 ou 2mL e os autoclave
2. Identifique os tubos de 1,5 ou 2 mL com a identificação correspondente da amostra.
3. Adicione 200 µL de Água Livre de Nuclease nos tubos rotulados.
4. Em uma cabine de fluxo laminar (Nível II) dentro do laboratório NB3, raspe uma alça de 10 µL de cultura micobacteriana para o tubo rotulado, repita esse processo por 3 vezes.
5. Vortexe os tubos até que os grumos de células se desfaçam.
6. Incube as amostras a 80°C por 1 hora.
7. Limpe a superfície externa dos tubos e transfira as amostras inativadas para um laboratório NB2.

4.1.2. Inativação por calor (meio líquido)



1. Identifique cada tubo de 15 mL com a identificação da amostra.
2. Dentro do laboratório NB3, vortexe o tubo MGIT até que a biomassa bacteriana do fundo esteja homogeneizada.
3. Transfira 5 mL do conteúdo do MGIT para o tubo de 15 mL usando uma pipeta de Pasteur de 2 ou 3 mL e descarte a pipeta.
4. Coloque o tubo em um forno a 80°C e deixe inativar por 1 hora.



COPIA CONTROLADA

5. Limpe a superfície externa dos tubos e transfira as amostras inativadas para um laboratório NB2.

4.2. Pré-tratamento do material inativado (para meio líquido)

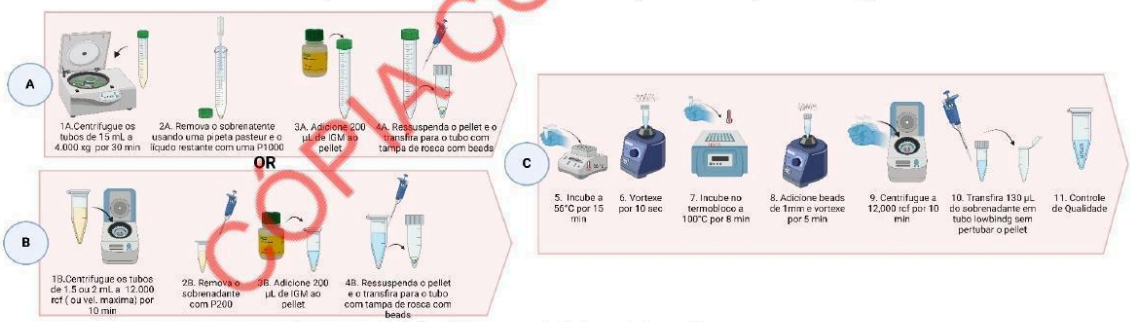
Ao trabalhar com CPC em meio líquido devemos considerar a possibilidade de contaminantes como DNA Humano ou de outras bactérias, mesmo após o tratamento por NALC e PANTA. A remoção do DNA extracelular pode ser feita através do tratamento com a enzima DNAse I e uma lavagem do pellet.

1. Em um laboratório NB2, centrifugue os tubos de 15mL a 4000rcf por 30 minutos a temperatura ambiente
2. Utilizando uma pipeta pasteur, aspire o sobrenadante com cuidado sem perturbar o pellet e descarte o aspirado.
3. Ressuspenda o pellet adicionando 5mL de tampão de DNAse
4. Homogenize com vortex por 10 segundos para dispersar as células
5. Centrifugue os tubos por 30 minutos a 4000 rcf a temperatura ambiente
6. Utilizando uma pipeta pasteur aspire todo o sobrenadante e descarte o aspirado
7. Remova o sobrenadante restante com uma pipeta P1000
8. Ressuspenda o pellet em 500 µL em tampão de DNAse e adicione 10 µL de enzima DNAse I (15 U/µL totalizando 150 U por amostra).
9. Incube a 37° por 30 minutos com agitação gentil para a remoção do DNA humano e de bactérias comensais
10. Incube a 65° por 10 minutos para inativação da enzima DNAse.
11. Centrifuge a 21000 rcf (ou velocidade máxima) por 20 minutos para a formação de pellet
12. Descarte o sobrenadante cuidadosamente utilizando P200 ou P100 sem perturbar o pellet.

CÓPIA CONTROLADA

4.3. Extração utilizando Instagene Matrix

1. Prepare tubos de rosca de 1,5 mL ou 2mL contendo 3 beads de 1mm e os autoclave.
2. Remova o IGM da geladeira e deixe-o homogeneizar a temperatura ambiente por 30 minutos, depois o homogenize com um agitador magnético de acordo com as instruções do fabricante.
3. Caso esteja trabalhando com cultura líquida: Centrifugue os tubos de 15mL a 4000 rcf por 30 minutos e depois remova o material sobrenadante com uma pipeta P1000.
4. Caso esteja trabalhando com cultura sólida: Centrifuge os tubos a 12000rcf por 10 minutos e remova o sobrenadante com uma pipeta P200.
5. Caso esteja trabalhando com cultura líquida, adicione 200 µL de IGM bem homogeneizado ao tubo de 15mL, ressuspendendo o pellet com uma pipeta P200, transfira todo o material para o tubo rosqueável contendo as beads.
6. Caso esteja trabalhando com cultura sólida, adicione 200 µL de IGM bem homogeneizado ao tubo com o material inativado, ressuspendendo o pellet com uma pipeta P200, transfira todo o material para o tubo rosqueável contendo as beads.
7. Incube o tubo rosqueável a 56 °C por 15 minutos
8. Homogenize utilizando vortex por 10 segundos para dispersar as células
9. Incube em um termobloco a 100° por 8 minutos
10. Homogenize as amostras em vortex por 5 minutos
11. Centrifugue os tubos a 12000 rcf por 15 minutos
12. Transfira 130 µL do sobrenadante com DNA para tubos (low binding) de 1,5 ou 2mL



sem encostar no pellet (matrix com debris celulares)

13. Prossiga com o controle de qualidade da extração.Controle de qualidade da Extração

4.4. Controle de qualidade da extração

Para garantir que o DNA extraído é adequado para o sequenciamento genômico devemos avaliar os seguintes fatores: Integridade, Concentração / Quantidade e Pureza.

COPIA CONTROLADA

4.4.1. Eletroforese em gel de agarose

A avaliação de Integridade do DNA pode ser feita através da eletroforese em gel de agarose, a partir do padrão de bandeamento observado podemos classificar o nível de fragmentação do DNA.

1. Pese 4 g de agarose e dissolver em 200 mL de tampão TAE 1X em um béquer resistente ao calor, para obter uma concentração final de 2% (m/v).
2. Aqueça a mistura em micro-ondas, com intervalos curtos de 10 a 15 segundos, até que toda a agarose esteja dissolvida e a solução esteja transparente.
3. Deixe a solução esfriar por 2-3 minutos e adicione 6 µL SYBR ao gel e o homogenize agitando o béquer.
4. Adicione o gel em uma forma para agarose e adicione o pente para formar poços, e aguarde a solidificação do gel.
5. Utilizando tubos de PCR misture 3 µL de cada amostra com 5 µL de Blue Juice (Corante de Carga).
6. Coloque o gel na cuba de eletroforese e encha a cuba com TAE 1X até que o gel esteja coberto.
7. Pipete a mistura de cada amostra em um poço, lembrando de deixar um poço livre para o marcador de peso molecular.
8. Pipete 5 µL de marcador de peso molecular (Ladder 1kb plus) em um dos poços, o marcador deve estar preparado de acordo com as instruções do fabricante.
9. Conecte os eletrodos a cuba de corrida e configure a fonte para 80V e corra por aproximadamente 1 hora.
10. Após a corrida desligue a fonte e visualize o gel no transiluminador UV.
11. Uma única banda no topo do gel indica que a amostra está íntegra, a visualização de rastros na amostra indica fragmentação do DNA.

CÓPIA CONTROLADA

4.4.2. Quantificação utilizando Qubit

Para estimar a concentração do DNA podemos utilizar o kit High Sensitivity Double Strand DNA com o equipamento Qubit. Este kit tem sensibilidade adequada para o posterior preparo de bibliotecas genômicas.

1. Retire o kit qubit DS HS da geladeira e o deixe a temperatura ambiente por pelo menos 30 minutos
2. Prepare a solução trabalho realizando uma diluição 1:200 do corante qubit no tampão qubit, para 20 amostras: adicione 22 µL de corante em 4378µL de tampão.
3. Em um tubo adequado para o equipamento qubit adicione 10 µL do tampão de calibração 1# e 190 µL de solução trabalho.

4. Em um tubo adequado para o equipamento qubit adicione 10 µL do tampão de calibração 2# e 190 µL de solução trabalho.
5. Aguarde 3 minutos
6. Realize o procedimento de calibração no equipamento clicando em "Run Standard", inserindo os tubos contendo #1 e #2 de acordo com as instruções na tela do equipamento.
7. Após a calibração, adicione 2 µL da amostra em um tubo identificado e 198 µL de solução trabalho, repita o processo para cada amostra.
8. Aguarde 3 minutos
9. Coloque o tubo contendo a primeira amostra e Clique em "Run sample" , ajuste o volume da amostra para 2 µL no menu lateral. Repita o processo para todas as amostras.
10. Anote o valor registrado na tela do equipamento e descarte o tubo.
11. É necessária uma concentração mínima de DNA de 1,0 ng/µL para o sequenciamento na plataforma Illumina.

4.4.3. Avaliação de Pureza utilizando Nanodrop

O Equipamento Nanodrop permite que determinemos a pureza do DNA através da determinação da absorbância em comprimentos de onda dos contaminantes comumente encontrados em extrações de DNA. Isso pode ser feito observando as razões 260/280 e 260/230. Para a razão 260/280, o valor ideal é 1,8. Valores entre 1,7 e 2,0 indicam algum grau de contaminação. A razão 260/230 deve estar entre 1,8 e 2,2. Em geral, se a razão estiver significativamente fora desse intervalo, pode indicar a presença de fenol residual, guanidina, esferas magnéticas, carboidratos ou proteínas. Uma razão 260/230 baixa pode indicar um problema com o tampão usado na medida em branco

1. Ligue o Nanodrop e abra o software correspondente no computador ou na tela do equipamento.
2. Escolha o tipo de amostra dsDNA (DNA de fita dupla) para que o Nanodrop ajuste os parâmetros de leitura.
3. Aplique 2 µL de água ultrapura na base de leitura do Nanodrop, feche o braço e selecione "Blank" para calibrar o equipamento. Em seguida, limpe com um lenço sem fiapos.
4. Aplique 2 µL da amostra de DNA, RNA ou proteína na base de leitura. Fechar o braço e clicar em "Measure". O software exibirá a concentração e os valores de pureza (260/280 e 260/230 para ácidos nucleicos). Repita o processo para todas as amostras.
5. Anote os dados de concentração e pureza fornecidos pelo software.
6. Após todas as leituras, limpar a base de leitura com um lenço sem fiapos para evitar resíduos.

CÓPIA CONTROLADA

5. ABREVIATURAS

CPC	Cultura Primária Clínica
DNA	Ácido Desoxirribonucleico
IGM	Instagene Matrix
NB3	Nível de biossegurança 3

6. REFERÊNCIAS

Conceição et al. 2024. *Mycobacterium tuberculosis* DNA Extraction Using InstaGene Matrix and High-Speed Homogenizer from Clinical Primary Culture. Protocols.io https://dx.doi.org/10.17504/protocols.io.6qpvr8rwzlmk/v1

7. HISTÓRICO DO DOCUMENTO

Versão nº.	Data	Localização do histórico de alterações	Autor/ Revisor	Aprovando Oficial	Data de aprovação	Próxima data de revisão
1.0			Davi Marcon			

CÓPIA CONTROLADA

5.2.2. REVIGET_POP003: Sequenciamento Genômico Utilizando O Kit Dna Prep E O Sequenciador Miseq (Illumina)”

Este POP foi utilizado para a implementação do uso de WGS durante os treinamentos do projeto REVIGET. E baseou-se nos equipamentos disponibilizados em cada um dos laboratórios centrais, voltado para o preparo de bibliotecas utilizando o kit illumina DNAprep e o DNA extraído no POP_001 para posterior sequenciamento com o equipamento Illumina MiSeq.

Procedimento operacional padrão (POP) nº REVIGET_POP003:
SEQUENCIAMENTO GENÔMICO UTILIZANDO O KIT DNA PREP E O SEQUENCIADOR MISEQ (ILLUMINA)
Versão #: 1.0, Página 1 de 29

SUMÁRIO

1. ESCOPO **2**
2. RESPONSABILIDADES **2**
3. GERENCIAMENTO DE RISCO **2**
4. ANTECEDENTES **3**
5. MATERIAL **3**
5.1 REAGENTES 3
5.2 CONSUMÍVEIS 4
5.3 EQUIPAMENTOS 5
6. PROCEDIMENTO **5**
6.1 NORMALIZAÇÃO E TAGMENTAÇÃO 5
6.1.1 Preparativos (Indexação) 5
6.1.2 Métodos (Indexação) 6
6.2 LIMPEZA DE PÓS-TAGMENTAÇÃO 7
6.2.1 Preparativos (Limpeza Pós-Tagmentação) 7
6.2.2 Métodos (Limpeza Pós-Tagmentação) 8
6.3 AMPLIFICAÇÃO DO DNA INDEXADO 9
6.3.1 Preparativos (Amplificação) 9
6.3.2 Método (Amplificação) 10
6.4 LIMPEZA DE BIBLIOTECAS 11
6.4.1 Preparo das beads 11
6.4.2 Preparo de reagentes (Limpeza Bibliotecas) 12
6.4.3 Método (Limpeza Bibliotecas) 12
6.6 DESNATURAÇÃO E DILUIÇÃO DAS BIBLIOTECAS E PHIX 15
6.6.1 Preparativos (Desnaturação e Diluição) 15
6.6.2 Método (Desnaturação e Diluição) 15
6.6.3 Desnaturação da Biblioteca 16
6.7 CONFIGURAÇÃO DE CORRIDA 17
6.7.1 O Sequenciador MiSeq 18
6.7.2 Configuração do Local Run Manager 18
6.7.3 Efetue login no Local Run Manager 18
6.7.4 Criar 'Nova corrida' 18
6.8 CARREGAMENTO DE CONSUMÍVEIS 19
6.8.1 Cartucho 19
6.8.2 Flowcell 20
6.8.3 Carregar o cartucho de reagente 22

CÓPIA CONTROLADA

6.8.4 Iniciar corrida de sequenciamento 23
6.8.5 Revisar Verificação Pré-corrida 24
6.8.6 Monitorar o progresso da corrida 24
6.8.7 Lavagem pós-corrida 26
7. ABREVIATURAS 27
8. REFERÊNCIAS 28
9. DOCUMENTOS DE APOIO 28
10. DISPONIBILIDADE 28
11. HISTÓRICO DO DOCUMENTO 28

1. ESCOPO

Este procedimento operacional padrão (POP) descreve o fluxo de trabalho de preparação e sequenciamento da biblioteca a ser seguido para o sequenciamento do genoma completo do DNA de *Mycobacterium tuberculosis* usando a plataforma **Illumina MiSeq.** O resumo do fluxo de trabalho dos procedimentos está apresentado no **Apêndice 1.**

2. RESPONSABILIDADES

É responsabilidade de cada membro do **(adicionar nome do laboratório)**: 1) Seguir esse POP conforme escrito, e 2) Tomar as precauções de segurança necessárias ao manusear amostras durante este processo. É responsabilidade do líder do laboratório (pesquisador ou gestor principal) garantir que este procedimento seja apropriado para o uso pretendido e que todos os integrantes envolvidos neste processo sejam devidamente treinados antes de processar quaisquer amostras.

3. GERENCIAMENTO DE RISCO

Os riscos envolvidos ao executar o procedimento descrito no POP incluem descartar os resíduos tóxicos da lavagem pós-corrida do Illumina **MiSeq**, pois os resíduos contêm formamida, uma amida alifática que é uma provável toxina reprodutiva. A porcentagem final de formamida na solução de resíduos é de 1,4%. Esse valor é calculado considerando que, após uma corrida, restam 2,1mL de formamida na posição nove do cartucho (de um total inicial de 5,47 mL), o que significa que 3,37 mL foram descartados junto com os resíduos. O volume total dos resíduos gerados é de 235 mL. Os resíduos contendo formamida são decantados em um frasco devidamente rotulado como "Resíduos contendo formamida". Quando os pods são esvaziados por meio de purga, eles podem ser descartados como resíduos de risco biológico no laboratório de controle de qualidade. Os cartuchos de sequenciamento são descartados no recipiente de resíduos de risco biológicos localizado no laboratório de sequenciamento do **[nome do local]**. A empresa **[nome da empresa]** realiza a coleta e remoção dos resíduos biológicos para incineração a cada **[frequência da coleta]**. Há riscos potenciais ao pessoal por meio de inalação, ingestão, contato com a pele e com os olhos. Além disso, existe o risco de contaminação do DNA nas amostras, o que deve ser evitado com medidas adequadas. Para evitar ou reduzir esses riscos é essencial o

CÓPIA CONTROLADA

uso de equipamentos de proteção individual (EPI), incluindo proteção ocular, máscara, luvas e jaleco de laboratório. Além disso, é importante manusear adequadamente os reagentes usados como resíduos químicos e descartá-los de acordo com os padrões de segurança vigentes.

4. ANTECEDENTES

O DNA genômico é extraído de *M. tuberculosis*, cultivado em placas 7H11, em meio líquido 7H9 (culturas MGIT) ou de amostras de escarro. Utiliza-se os protocolos "**REVIGET-POP001: Extração de DNA Micobacteriano Utilizando o Kit InstaGene**" para realizar o isolamento do DNA, e o "**REVIGET-POP002: Controle de Qualidade do DNA Genômico**" antes de iniciar o preparo de bibliotecas genômicas. Esse protocolo é adaptado para a plataforma Illumina MiSeq e descreve o preparo de até 96 amostras para sequenciamento utilizando o Kit Illumina DNA Library Prep. Todos os procedimentos experimentais serão realizados **[nome do laboratório].**

5. MATERIAL

5.1 REAGENTES

CÓPIA CONTROLADA

Produto	Temperatura de Armazenamento	Código do produto
Amostra de DNA genômico	Freezer (-25°C a -15°C)	-
Água livre de nucleases ou água Milli-Q	Ambiente (15°C a 30°C)	-
Bead-Linked Transposomes (BLT) - Transposons ligados a beads	Geladeira (2°C a 8°C)	20060059
Controle de sequenciamento PhiX v3	Freezer (-25°C a -15°C)	FC-110-3001
Enhancement PCR Mix (EPM) - Mistura de PCR aprimorada	Freezer (-25°C a -15°C)	20060059
Etanol 70% (para limpeza da área de trabalho e do cartucho após descongelamento)	Ambiente (15°C a 30°C	-
Etanol 80% (preparado diariamente antes do experimento)	Ambiente (15°C a 30°C	-
Hipoclorito de sódio - NaOCl		
Hybridization Buffer (HT1) - Tampão de hibridização	Freezer (-25°C a -15°C)	20060059
Índices de DNA CD Illumina (adaptador i7 e adaptador i5 para 96 amostras) ou placa de índice de 96 poços	Freezer (-25°C a -15°C)	20091654

NaOH 0,2N (preparado diariamente antes da sessão do experimento)	Ambiente (15°C a 30°C	-
MiSeq Reagent Kit v3 para 600 ciclos: Flowcell	Geladeira (2°C a 8°C)	MS-102-3003
MiSeq Reagent Kit v3 para 600 ciclos: Cartucho de reagentes	Freezer (-25°C a -15°C)	MS-102-3003
MiSeq Reagent Kit v3 para 600 ciclos: Wash buffer (PR2)	Geladeira (2°C a 8°C)	MS-102-3003
Reagentes Qubit (Tampão, marcador e padrões 1 e 2)	Geladeira (2°C a 8°C)	Q32851
Illumina Purification Beads (IPB) – Esferas para purificação das amostras da Illumina	Geladeira (2°C a 8°C)	20060059
Tween 20 para Manutenção pós corrida		
Resuspension Buffer (RSB) - Tampão de ressuspensão	Freezer (-25°C a -15°C)	20060059
Tagmentation Buffer 1 (TB1) - Tampão de Tagmentação 1	Freezer (-25°C a -15°C)	20060059
Tagment Stop Buffer (TSB) – Tampão de Parada de Tagmentação	Ambiente (15°C a 30°C)	20060059
Tagment Wash Buffer (TWB) - Tampão de lavagem de tagmentação	Ambiente (15°C a 30°C)	20060059

CÓPIA CONTROLADA

5.2 CONSUMÍVEIS

- Caneta marcador permanente
- Gelo (para as etapas de amplificação e manuseio de amostras de gDNA)
- Lenços umedecidos sem fiapos (lenços do tipo Kmix)
- Ponteiras de 200µL para P200
- Ponteiras de 20µL pontas para P20
- Ponteiras de 1000µL para P1000
- Ponteiras de 10µL (pontas LoRetention) para P10
- Tampas de adaptador de índice (inclusas no kit de índice de 24 amostras)
- Tiras de PCR de 200 µl com 8 tubos com tampas individuais anexas OU tubos de PCR individuais de 200µL
- Toalhas de papel (para limpeza da área de trabalho)
- Tubos de microcentrífuga Eppendorf

5.3 EQUIPAMENTOS

- Centrífuga para placas e Microcentrífuga para tubos de 2mL
- Microcentrífuga Eppendorf de baixa retenção (LoBind) de tubos de 2mL
- Pipeta P1000
- Pipeta P200
- Pipeta P20
- Pipeta P10
- Qubit
- Sequenciador MiSeq
- Termociclador
- Vórtex

6. PROCEDIMENTO

6.1 NORMALIZAÇÃO E TAGMENTAÇÃO

O

6.1.1 Preparativos (Indexação)

Item	Armazenado	Instrução
BLT (transposons ligados a beads)	2°C a 8°C	Leve à temperatura ambiente. Agite no vórtex. Não centrifugue antes de pipetar.
TB1 (Buffer de marcação)	-25°C a -15°C	Descongele até a temperatura ambiente. Agite no vórtice.

CÓPIA CONTROLADA

***Observação: NÃO USE BLT QUE TENHA SIDO ARMAZENADO ABAIXO DE 2°C!!!**

Pré-programe o seguinte programa TAG no termociclador e ajuste a temperatura da tampa de pré-aquecimento em 100°C e o volume de reação em 50µL.

Programa TAG	
Temperatura (°C)	**Tempo**
55	15 minutos
10	Infinito

6.1.2 Métodos (Normalização e Tagmentação)

CQ1

1. Determine a integridade do gDNA utilizando um gel de agarose a 1,5%.
   Consulte o POP: TBG020_PCR e eletroforese em gel para o preparo do gel de agarose a 1,5%.
2. Avalie a pureza e possíveis contaminantes do DNA utilizando o Nanodrop, que podem interferir em aplicações subsequentes (como WGS).
3. Os resultados do Nanodrop geralmente são fornecidos no Formulário de Solicitação de Sequenciamento pelos responsáveis.

CQ2

1. Utilize o ensaio Qubit dsDNA ou o ensaio de Quantificação de DNA de Faixa Ampla/Alta Sensibilidade (ver POP_Qubit) para quantificar o DNA.
2. Registre as concentrações de DNA.
3. Se as amostras de gDNA apresentarem alta concentração, resultando em um volume abaixo de 1,9–2µL, realize uma diluição adequada para atingir a quantidade necessária de DNA para a entrada (exemplo, 50ng).
4. Realize diluições em duas etapas, diluindo o gDNA com água livre de nucleases (NFW), e prossiga para a etapa seguinte.
5. Em uma planilha Excel, registre a concentração de DNA determinada e faça a diluição para obter 100ng de DNA de entrada em cada tubo da tira, de acordo com o seguinte cálculo:

DNA de entrada (ng)/ = volume de DNA de entrada

$$\frac{}{\text{Concentração de DNA (ng/µl)}} = \text{volume de DNA de entrada}$$

6. Dilua a amostra novamente em água ultrapura, conforme calculado, para atingir um volume final de 30µL e uma concentração final entre 100 e 500ng.
7. Adicione 5µL de gDNA a cada poço de uma placa de 96 poços ou em tubos de tira de 0,2mL, visando uma entrada final de 100-500ng de DNA.
8. Se o volume de gDNA for inferior a 30µL, complete com água livre de nucleases até atingir o volume final de 30µL.
9. Agite bem o BLT no vortex para garantir a ressuspensão adequada. Repita o processo de agitação conforme necessário.
10. Vortexe o BLT entre os usos para manter a homogeneidade da solução
11. Agite bem o BLT no vórtice para ressuspender. Repita, conforme necessário.
12. Faça um vórtice entre os usos do BLT.
13. Prepare o Master Mix de tagmentação da seguinte forma:
    Multiplique o volume de cada componente pelo número de amostras que estão sendo processadas, adicionando 2µL extras por BLT e TB1.

Componente	Volume (µL) por amostra	Volume (µL) para 30 reações

CÓPIA CONTROLADA

BLT	10	300
TB1	10	300
Total	20	600

Observação: O excesso de reagente está incluído no volume para garantir uma pipetagem precisa.

14. Agite bem o Tagmentação MM (Master Mix) para garantir sua ressuspensão.
15. Adicione 20µL do MM de Tagmentação em cada tubo das duas tiras de oito tubos, contendo 100ng de DNA.
    Observação: Use uma nova ponteira para cada amostra.
16. Misture cuidadosamente as amostras, pipetando 10 vezes para ressuspender lentamente cada amostra, utilizando uma pipeta P20 ajustada para 15µL.

Observação: Sempre pipete lentamente para evitar a formação de bolhas. Se o Tagmentation MM estiver no interior da tampa do tubo de amostra, centrifugue rapidamente antes de pipetar lentamente para misturar.

17. Feche as tampas dos tubos.
18. Descarte do MM de Tagmentação.
19. Execute as amostras no termociclador utilizando o programa TAG.

6.2 LIMPEZA DE PÓS-TAGMENTAÇÃO

6.2.1 Preparativos (Limpeza Pós-Tagmentação)

Item	Armazenado	Instrução
TSB (buffer de parada de marcação)	15°C a 30°C	Se forem observados precipitados, aqueça a 37°C por 10 minutos e agite no vórtex até que os precipitados sejam dissolvidos. Use em temperatura ambiente.
TWB (tampão de lavagem de marcação)	15°C a 30°C	Use em temperatura ambiente. Agite em vórtex.
EPM (mistura de PCR aprimorada)	-25°C a -15°C	Descongele no gelo. Inverta para misturar e centrifugue brevemente.
Adaptadores de índice (em placa)	-25°C a -15°C	Descongele até a temperatura ambiente. Agite no vórtex. Gire brevemente antes de usar.

CÓPIA CONTROLADA

Pré-programe o seguinte programa **PTC** no termociclador e ajuste a temperatura da tampa de pré-aquecimento para 100°C e o volume de reação para 60µL.

Programa PTC	
Temperatura (°C)	**Tempo**
37	15 minutos
10	Infinito

6.2.2 Métodos (Limpeza Pós-Tagmentação)

1. Adicione 10µL de TSB a cada reação de Tagmentação, utilizando uma pipeta P10.
2. Pipete lentamente cada tubo com uma pipeta P100 ajustada para 40µL, repetindo 10 vezes para ressuspender. Utilize uma nova ponteira para cada reação.
   Observação: Se houver bolhas visíveis, estoure-as com uma ponta.
3. Feche os tubos.
4. Execute as amostras no termociclador usando o programa PTC.
5. Enquanto aguarda, prepare o *EPM Master Mix* em um tubo Eppendorf de 1,5mL, conforme indicado:

Componente	**Volume (µL) por amostra**	**Volume (µL) para 30 reações**
EPM	20	600
Água ultrapura	20	600
Total	40	1200

Observação: O excesso de reagente foi incluído no volume para garantir pipetagem precisa.

6. Armazene a mistura principal de EPM no gelo até ser necessário.
7. Após a conclusão do programa PTC, remova os tubos e coloque-os na placa magnética. Aguarde até que o líquido fique claro (~3 min).
8. Use uma pipeta P100 ajustada para 100µL para remover e descartar o sobrenadante.
9. Realize duas lavagens da seguinte maneira:
   a. Remova os tubos com amostras da placa magnética e adicione lentamente 100µL de TWB diretamente sobre as beads.

COPIA CONTROLADA

Observação: A pipetagem lenta minimiza o risco de formação de espuma no TWB, o que pode causar aspiração incorreta e mistura incompleta.

b. Ajuste a pipeta para 80µL e pipete lentamente para cima e para baixo para ressuspender completamente as beads.
c. Coloque as tiras de tubos de volta na placa magnética e aguarde até que o líquido fique claro (~3 min).
d. Ajuste a pipeta para 100µL para remover e descartar o sobrenadante.

10. Remova as tiras de tubos da placa magnética e pipete lentamente 100µL de TWB diretamente sobre as beads.
11. Pipete lentamente cada tubo para ressuspender as beads.
12. Feche os tubos e coloque as tiras de volta na placa magnética até que o líquido fique claro (~3 min). Mantenha as tiras na placa magnética até a etapa 2 da seção de **Amplificação**.

Observação 1: O TWB deve permanecer nos tubos para evitar a secagem excessiva das beads.
Observação 2: Ao final, três lavagens devem ser concluídas.

6.3 AMPLIFICAÇÃO DO DNA INDEXADO

6.3.1 Preparativos (Amplificação)

1. **Abra somente um tubo de adaptador de índice por vez** para evitar o extravio das tampas e a contaminação cruzada. Se as tampas forem extraviadas ou potencialmente contaminadas, utilize novas tampas após abrir o tubo.
2. Pré-programe e execute o programa BLT PCR no termociclador. Defina a temperatura de pré-aquecimento da tampa para 100°C e o volume de reação para 50µL.

PROGRAMA BLT PCR		
Temperatura (°C)	**Tempo**	
68	3 minutos	
98	3 minutos	
98	45 segundos	**8 ciclos***
62	30 segundos	
68	2 minutos	
68	1 minuto	
10	2 minutos	

CÓPIA CONTROLADA

4	Infinito	

**** Observação:***

- *Use **8 ciclos** quando a entrada de DNA for de 100-500 ng.*
- *Use **10 ciclos** quando a entrada de DNA for de 10-24 ng.*
- *Use **14 ciclos** quando a entrada de DNA for de 1-9 ng.*

6.3.2 Método (Amplificação)

1. Agite o *PCR Master Mix* (PCR MM) em um vórtex e centrifugue a 280xg por 10 segundos.
2. Nas tiras de tubos que estão na placa magnética, use uma pipeta de 200µL para remover e descartar o sobrenadante.
   OBSERVAÇÃO: Sempre se certifique de remover o sobrenadante residual com uma pipeta de 10µL **antes de adicionar o PCR MM às beads**!
   A espuma que permanece nas paredes dos tubos não afeta a biblioteca.
3. Remova as tiras de tubos da placa magnética.
4. Adicione imediatamente 40µL de PCR MM diretamente nas beads de cada amostra.
5. Adicione 5µL de cada índice apropriado (adaptador i7 e i5) a cada tubo, conforme indicado na Tabela 1, quando o kit de reação para 24 amostras for utilizado
   OBSERVAÇÃO: Se estiver usando uma placa de índice de 96 poços, adicione 10µL de um único poço a uma única amostra.
6. Utilizando uma pipeta P200 ajustada para 40µL, pipete lentamente 10 vezes para misturar.
   OBSERVAÇÃO: Como alternativa, sele a placa de amostras e use um agitador de placas a 1600rpm por um minuto.
7. Feche as tiras de tubos.
8. Centrifugue a 280xg por 30 segundos.
9. Coloque os tubos com amostras no termociclador e execute o programa **BLT PCR**.



PONTO DE PARADA SEGURO
Se for parar, armazene entre 2°C e 8°C por até 3 dias.

CÓPIA CONTROLADA

Tabela 1: Combinações do adaptador de índice para 12 amostras.

Amostra #	I7	I5
1	H705	H503
2	H706	H505
3	H707	H506

CÓPIA CONTROLADA

6.4 PURIFICAÇÃO DAS BIBLIOTECAS

6.4.1 Preparo das beads

1. Calibrar os seguintes itens para a temperatura ambiente (IPB e RSB) por 30 min antes de usá-los.

Item	Armazenado	Instrução
IPB (Beads de purificação da Illumina)	2°C a 8°C	Deixe em temperatura ambiente por 30 minutos. Vórtex e inverta para misturar.
RSB (tampão de ressuspensão)	-25°C a -15°C	Descongele e deixe em temperatura ambiente. Agite no vórtex.

2. Prepare três conjuntos de tubos conforme o número de amostras (Tipo Eppendorf lo-binding de 1,5mL), identificando-os.
3. Vortex o IPB várias vezes para ressuspender antes de usar.
4. No primeiro conjunto de tubos Eppendorf Lo-Bind de 1,5mL, adicione o seguinte:
    a. 40 µL de água ultrapura
    b. 45 µl de IPB
    c. Misture bem pipetando ou sele a placa e use um agitador de placas a 1600rpm por 1 minuto.

5. No segundo conjunto de tubos Eppendorf Lo-Bind de 1,5 ml, adicione 15µL de IPB (estoque não diluído).
   Observação: Misture bem o IPB antes de aliquotar o IPB nos novos tubos de 1,5mL!

6.4.2 Preparo de reagentes (Limpeza Bibliotecas)

80% etanol (preparar solução fresca)
4 mL etanol absoluto
1 mL água ultrapura

6.4.3 Método (Limpeza Bibliotecas)

6. Centrifugue os tubos de tiras a 280xg por 1 minuto para coletar o conteúdo no fundo dos tubos.
7. Coloque os tubos na placa magnética e espere até que o líquido fique límpido (~5 minutos).
8. Transfira 45µL do líquido sobrenadante à mistura de IPB diluída (primeiro conjunto) e pipete **lentamente em cada** tubo 10 vezes para misturar.
9. Feche e incube em temperatura ambiente por 5 minutos, no mínimo.
   Observação 1: Inicie o tempo de incubação imediatamente após o último tubo ser completamente misturado.
   Observação 2: O tempo de incubação pode ser maior que 5 minutos, mas não deve ser menor, para garantir a purificação de amplicons do mesmo tamanho.
10. Coloque os tubos na placa magnética e espere até que o líquido fique límpido (~5 minutos).
11. Transfira 125µL de sobrenadante de cada um dos primeiros tubos para o segundo conjunto correspondente de tubos de 1,5mL (contendo 15µL de IPB não diluída).
    Observação: Pressione o fundo do tubo contra o ímã e remova o sobrenadante com cuidado para evitar a ressuspensão das beads.
12. Pipete lentamente cada tubo 10 vezes para misturar enquanto adiciona o sobrenadante ao IPB não diluído.
13. Incube os tubos em temperatura ambiente por 5 minutos, no mínimo.

Observação: Inicie o tempo de incubação imediatamente após o último tubo ser misturado.

CÓPIA CONTROLADA

14. Coloque os tubos na placa magnética e aguarde até que o líquido fique límpido (~5 minutos). Descarte o primeiro conjunto de tubos. Sem perturbar as beads, remova e descarte o sobrenadante com uma pipeta P200.
15. Realize **duas lavagens** da seguinte maneira:
    a. Com os tubos na placa magnética, adicione 200µL de etanol 80% recém-preparado sem misturar
       OBSERVAÇÃO: NÃO ADICIONE O ETANOL DIRETAMENTE SOBRE AS BEADS!
    b. Incube por 30 segundos
    c. Sem perturbar as beads, remova e descarte o sobrenadante
    d. Repita a lavagem
16. Use uma pipeta de 10µL para remover e descartar resíduos de etanol.
17. Seque as beads ao ar no suporte magnético por 5 minutos, no mínimo.
    OBSERVAÇÃO 1: Certifique-se de que o etanol tenha evaporado completamente. As beads devem ter uma aparência fosca. Se ainda estiverem brilhantes, deixe secar até ficarem foscas.
    OBSERVAÇÃO 2: NÃO SEQUE POR MAIS DE 5 MINUTOS! Secagem excessiva pode causar grandes rachaduras nas beads, o que compromete a qualidade da biblioteca.
18. Retire do suporte magnético.
19. Adicione 32µL de RSB diretamente nas beads.
20. Pipete lentamente (com uma pipeta P200 ajustada para 30µL) para ressuspender as beads.
21. Incube em temperatura ambiente por 2 minutos, no mínimo.
22. Coloque os tubos no suporte magnético e aguarde até que o líquido fique límpido (~2 minutos).
23. Transfira 30µL de sobrenadante para o terceiro conjunto de novos tubos Eppendorf Lo-Bind.

PONTO DE PARADA SEGURO
Se você for parar, sele os tubos e armazene-os entre -25°C e -15°C por até três dias.

5.5 POOL DE BIBLIOTECAS

Controle de qualidade 3: Quantificação das Bibliotecas

1. Quantifique as bibliotecas ou o pool (conjunto) de bibliotecas utilizando o ensaio *dsDNA HS Qubit* (consulte o POP_Qubit).

Controle de Qualidade 4: Análise dos Fragmentos

CÓPIA CONTROLADA

2. Analise as bibliotecas ou o pool de bibliotecas utilizando o *Laboratório Lasca GX/GXII Touch* (veja POP_Lab Chip) ou o *TapeStation 2200/4200* (veja POP_TapeStation).

Diluição das bibliotecas para a concentração inicial de 4nM

Observação

- Comprimento de leitura no MiSeq = 2 x 300 bp
- Índices Illumina DNA/RNA UD = códigos de índice de 10 pb
- Concentração inicial no MiSeq = 4 nM
- Concentração de carga final no MiSeq = 6-20 pM (Este POP usa 12 pM)

3. Calcule a molaridade das bibliotecas ou do pool de bibliotecas utilizando a seguinte fórmula:

$$nM = \left(\frac{\text{Concentração de DNA (ng/µl)}}{660 \times \text{Tamanho do fragmento (bp)}}\right) \times 10^6$$

4. Usando o valor da molaridade, calcule os volumes de RSB e da biblioteca necessários para diluir até a concentração inicial de 4nM para o MiSeq, em uma planilha Excel.
5. Dilua as bibliotecas até a concentração inicial, conforme indicado abaixo para cada plataforma Illumina:

Plataforma Illumina	Concentração inicial (nM)	Concentração de carga final (pM)
MiSeq	4	8 – 10 (V2) 6 -20 (V3)

***Desnaturar e diluir**

6. Adicione 5µL de cada biblioteca a um tubo Eppendorf Lo-Bind de 1,5mL. Misture bem e centrifugue brevemente.
7. Utilize o Qubit para garantir que o pool tenha a concentração inicial correta (exemplo, 4nM para o MiSeq).

OBSERVAÇÃO: Armazene o restante das bibliotecas a 4nM entre -25°C e -15°C.

6.6 DESNATURAÇÃO DAS BIBLIOTECAS E DO PHIX

6.6.1 Preparativos (Desnaturação)

1. Descongele o cartucho antes de iniciar a desnaturação e diluição das bibliotecas e do controle PhiX.

CÓPIA CONTROLADA

Após os QCs da biblioteca	Tipo de descongelamento
No mesmo dia	• Em banho-maria em temperatura ambiente. • Em banho-maria (acelera o processo de descongelamento). • Não use água fervente diretamente no cartucho!
No dia seguinte	• Na geladeira a 4°C durante a noite.

Kit de cartucho (saída média ou alta) 300 ciclos	-25°C a -15°C	Deixe em temperatura ambiente por 30 minutos. Inverta 10 vezes lentamente para misturar.
Flowcell	2°C a 8°C	Deixe atingir a temperatura ambiente por pelo menos 30 minutos.
HT1 (tampão de hibridização)	-25°C a -15°C	Descongele no gelo (uso no mesmo dia) ou na geladeira a 4°C (uso no dia seguinte).
RSB (tampão de ressuspensão)	-25°C a -15°C	Descongele e deixe em temperatura ambiente. Agite no vórtice.

2. Sempre prepare NaOH recém-diluído para desnaturação das bibliotecas.
3. Para evitar erros de pipetagem, siga as etapas abaixo para diluir 1mL de NaOH 0,2N.
    a. Misture 200µL de NaOH 10N com 800µL de água livre de nuclease para obter 1mL de NaOH 2N
    b. Misture 200µL de NaOH 1N com 800µL de água livre de nuclease para obter 1mL de NaOH 0,2N
4. Use a diluição dentro de 12 horas.

CÓPIA CONTROLADA

6.6.2 Método (Desnaturação)

Observação: Sempre faça o Qubit do estoque de PhiX (10 nM) dentro da data de validade para garantir diluições corretas.

Diluição de PhiX para 4 nM

1. Descongele um tubo de controle PhiX 10nM.
2. Combine os seguintes volumes em um tubo Eppendorf Lo-Bind:
   2µL de controle PhiX 10nM + 3µL de RSB = Total 5µL de PhiX a 4 nM
3. Agite no vórtex e centrifugue.

Observação: Armazene o controle PhiX 4nM entre -25°C e -15°C por até 3 meses.

Desnaturação de PhiX

1. Adicione 5µL de NaOH 0,2N no tubo Eppendorf Lo-Bind contendo 5µL de PhiX 4 nM.

2. Agite brevemente no vórtex e centrifugue por pulso.
3. Incube à temperatura ambiente por 5 minutos.
4. Adicione 990µL de HT1 pré-resfriado ao PhiX desnaturado **diretamente após a incubação**.
5. Inverta o tubo para misturar e centrifugue por pulso.

Observação: Armazene o PhiX desnaturado a 20pM entre -25°C e -15°C por até duas semanas. Após duas semanas, os números de *clusters* tendem a diminuir.

6.6.3 Desnaturação das Bibliotecas e do PhiX

1. Calcule o volume das bibliotecas agrupadas a 2nM necessário para preparar 1mL de bibliotecas agrupadas a 20 pM.

C1xV1 = C2xV2
V1= (1000 uL x 20 pM)/ 4000 pM
V1= 5 uL de blibliotecas a 2 nM agrupadas

2. Adicione 5 µL de NaOH 0,2 N às bibliotecas agrupadas.
3. Agite brevemente no vórtex e, em seguida, centrifugue por pulso.
4. Incube à temperatura ambiente por 5 minutos.
5. Adicione 970µL de HT1 ao tubo contendo as bibliotecas desnaturadas e reunidas.
   Volume total: 1mL a 20pM.
6. Agite brevemente no vórtex e centrifugue por pulso.
7. Calcule o volume de bibliotecas a 20pM e controle PhiX a 20pM (3%) para um volume final de 650µL a 12 pM.

C1xV1 = C2xV2
V1= (650 uL x 12 pM)/ 20 pM
V1= 390 de blibliotecas a 20 pM desnaturadas e agrupadas

a. Bibliotecas: 12 pM
b. 3% PhiX: 35,75µL x 0,03 = 1µL de PhiX diluído e desnaturado a 20pM.

Observação: O volume de 3% PhiX foi determinado com base no volume das bibliotecas desnaturadas e agrupadas.

a. Volume da biblioteca:
   35,75µL – 1µL = 34,8µL de biblioteca diluída e desnaturada a 20pM.

CÓPIA CONTROLADA

b. Volume do tampão de hibridização:
   550µL – 34,8µL = 515,2µL de tampão de hibridização.
c. Em um novo tubo Lo-Bind de 1,5mL, combine os seguintes componentes:

Componente	Volume (µL)
Biblioteca 20pM (Desnaturada e Diluída)	34,8
PhiX 20pM (Desnaturado e Diluído)	1
Tampão de hibridização	515,2*

*Pipete 500µL de tampão de hibridização primeiro, depois os 15,2µL restantes com uma pipeta P20.

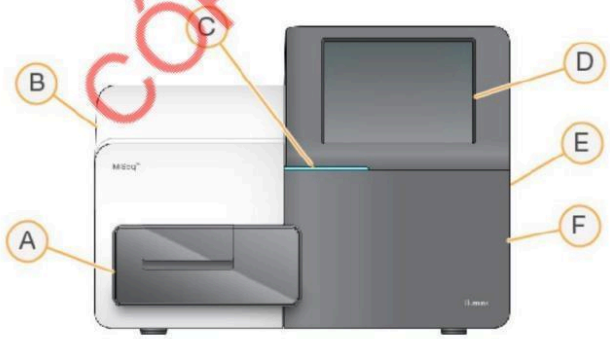
10. Agite brevemente no vórtex.

11. Coloque o tubo no gelo até que esteja pronto para ser adicionado ao cartucho e iniciar o sequenciamento.

6.7 CONFIGURAÇÃO DE CORRIDA

Observação: Consulte o guia do sistema MiSeq para obter instruções completas sobre como carregar o cartucho e a flowcell no instrumento.

6.7.1 O Sequenciador MiSeq



CÓPIA CONTROLADA

6.7.2 Configuração do Local Run Manager

1. Na tela inicial, selecione Sequência.
    a. O comando Sequence libera consumíveis de uma corrida anterior e abre a série de telas de configuração de corrida.
2. Na tela 'Modo de corrida', selecione 'Gerenciador de corrida local '.

6.7.3 Efetue login no Local Run Manager

1. Digite seu nome de usuário e senha.
    a. Nome de usuário: **verificar o nome do usuário local**
    b. Senha: **verificar a senha local**
2. Selecione Avançar.

6.7.4 Criar 'Nova corrida'

1. Clique em 'Criar corrida '.
2. Selecione 'Gerar FASTQ '.
3. A janela "Criar corrida" aparecerá na tela.
4. Preencha as caixas vazias necessárias, como 'Nome da corrida' e, se aplicável, 'Descrição da corrida '.
5. Selecione o 'Kit de preparação da biblioteca' selecionando 'Índices de DNA de CD ou kit apropriado usado durante a preparação da biblioteca'.
6. Certifique-se de que a opção 'Extremidade emparelhada' esteja marcada em 'Tipo de leitura'.
7. Em 'Read Lengths' certifique-se de que nas caixas 'READ1' e 'READ2' foi inserido 151 bp.
8. Além disso, 'INDEX1 e INDEX2, cada um tem um comprimento de leitura de 8 pb.
9. Role para baixo e importe uma planilha de amostra ou insira manualmente cada amostra a ser sequenciada com um ID de amostra exclusivo.
10. Continue adicionando uma 'Descrição da amostra' e selecione os índices i7 e i5 correspondentes usados durante a preparação da biblioteca por meio de um menu suspenso.
11. Após preencher a planilha de amostra, clique em 'Salvar corrida'.

6.8 CARREGAMENTO DE CONSUMÍVEIS

6.8.1 Cartucho

CÓPIA CONTROLADA

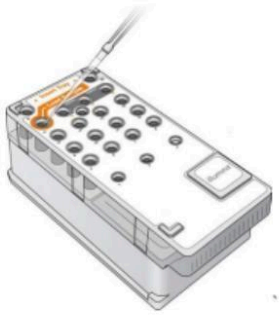


Figura 1: Cartucho MiSeq

1. Remova o cartucho do armazenamento a -25°C a -15°C.
2. Coloque o cartucho em banho-maria por 2 horas em temperatura ambiente.
   Observação: A estabilidade do cartucho, após descongelamento à temperatura ambiente, é de até 72 horas (cerca de 3 dias).
3. Inverta o cartucho 10 vezes para misturar os reagentes.
4. Inspecione os reservatórios grandes na parte inferior do cartucho para garantir que os reagentes estejam completamente descongelados e que os reservatórios estejam livres de cristais de gelo.
5. Bata levemente o cartucho na bancada para reduzir bolhas de ar e remover gotas de água. Certifique-se de que o cartucho esteja completamente seco antes de colocá-lo na máquina.

6.8.2 Flowcell

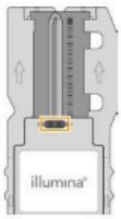


Figura 2: Flowcell Miseq

1. Remova um novo pacote de *flowcell* armazenado entre 2°C e 8°C.
2. Deixe a flowcell, ainda fechada em sua embalagem, em temperatura ambiente por 30 minutos.
   Observação: Evite repetidos ciclos de resfriamento e aquecimento da flowcell.
3. Retire a flowcell da embalagem de alumínio.
4. Coloque um novo par de luvas sem pó.

CÓPIA CONTROLADA

5. Segure a flowcell pelo cartucho de plástico e remova-a do recipiente
6. Limpe a superfície de vidro da flowcell com um lenço sem fiapos umedecido com álcool ou com um lenço Kimix levemente borrifado com etanol 70%.

7. Seque com um lenço de limpeza de lentes sem fiapos, tomando cuidado ao redor da junta preta da flowcell.



Figura 3: Procedimento de limpeza da flowcell

8. Inspecione as portas da flowcell para garantir que não haja obstruções e que a junta esteja devidamente assentada.
9. Abra a porta do compartimento da flowcell.
10. Pressione o botão de liberação à direita da trava da flowcell (Figura 4).
11. Se houver uma flowcell usada da corrida anterior, remova-a.
12. Certifique-se de que o estágio da flowcell esteja limpo. Se houver detritos, limpe o estágio com um lenço umedecido com álcool ou etanol 70% e um lenço macio.
13. Coloque a flowcell no estágio, alinhando-a sobre os pinos de alinhamento (Figura 3).

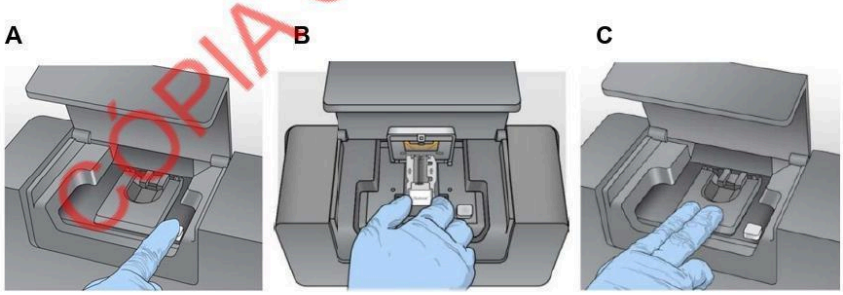


Figura 4: A) Destrave o compartimento da flowcell, B) Coloque a flowcell no estágio limpo e C) feche a trava da flowcell.

14. Feche a trava da flowcell para prendê-la (Figura 4).
15. Feche a porta do compartimento da flowcell.
16. Limpe o selo de alumínio que cobre o reservatório nº16, rotulado como **"Carregar biblioteca aqui"**, usando um lenço de papel que solte poucos fiapos.
17. Perfure o selo com uma ponta de pipeta limpa de 1mL.

CÓPIA CONTROLADA

18. Adicione 500µL das bibliotecas preparadas de 1,3pM com PhiX enriquecido (3%) no reservatório nº16 (Figura 5).
    a. Evite tocar no selo de alumínio ao dispensar as bibliotecas..

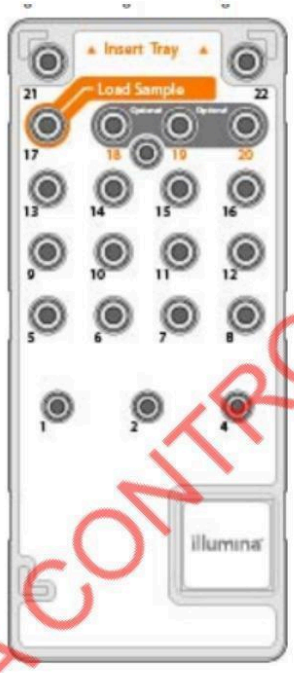


Figura 5: Carregue bibliotecas desnaturadas e diluídas com PhiX no cartucho.

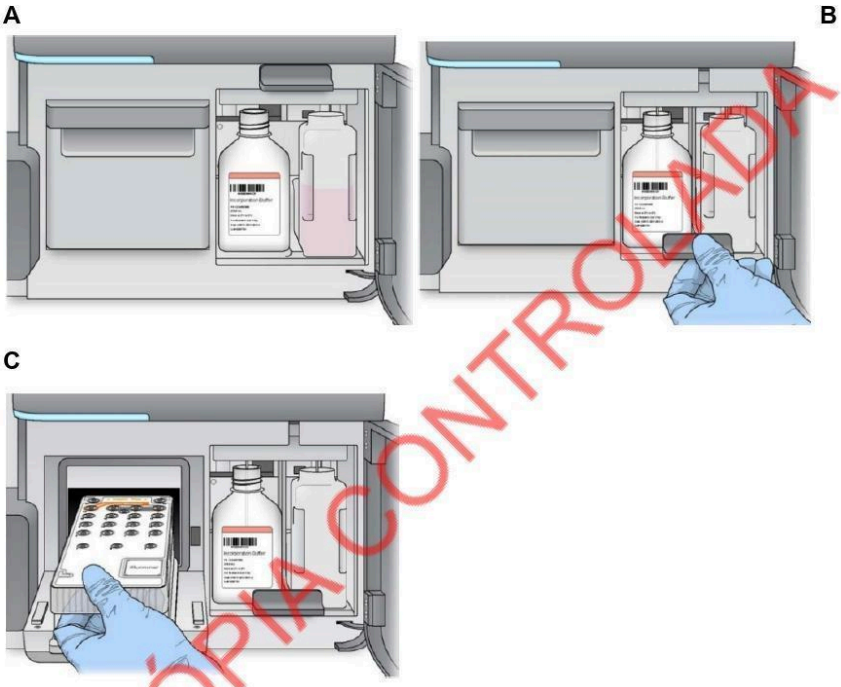
CÓPIA CONTROLADA

6.8.3 Carregar o cartucho de reagente

19. Abra a porta do compartimento de reagentes.

20. Se presente, remova o cartucho de reagente usado.
    Observação: Para facilitar o descarte seguro de reagente não utilizado que contenha formamida, o reservatório na posição #9 é removível. Descarte-o no recipiente designado para resíduos de formamida.

21. Deslize o novo cartucho de reagente para dentro do compartimento até que ele pare (Figura 6).



CÓPIA CONTROLADA

Figura 6: A) Carregue o cartucho de reagente e B) remova o conteúdo do frasco de resíduos e C) Insira o cartucho de reagentes.

22. Remova o frasco de reagente usado do compartimento (Figura 6).
23. Descarte o conteúdo de acordo com as normas aplicáveis.

Observação: Este conjunto de reagentes contém produtos químicos potencialmente perigosos. Ferimentos pessoais podem ocorrer por inalação, ingestão, contato com a pele e contato com os olhos. Utilize equipamentos de proteção, incluindo proteção ocular, luvas e um jaleco adequado para minimizar o risco de exposição. Manuseie os reagentes usados como resíduos químicos e descarte-os conforme as leis e regulamentações regionais, nacionais e locais aplicáveis.

24. Com a tampa rosqueada removida, deslize o frasco vazio de reagentes usados para dentro do compartimento até que ele pare.
25. Feche a porta do compartimento e selecione *Avançar*.

6.8.4 Iniciar corrida de sequenciamento

1. Selecione o nome de corrida apropriado na lista de execuções disponíveis.
    a. Use as setas para cima e para baixo para rolar pela lista ou digite o nome da corrida no campo *pesquisar*.
2. Selecione *Avançar*.
3. Confirme os parâmetros de corrida.
4. O *software* de controle verifica o número de ciclos especificados utilizando os seguintes critérios:
    a. O total de ciclos não excede o máximo permitido com base no cartucho de reagente carregado para a corrida.
    b. Os ciclos para Leitura 1 são maiores que os seis ciclos necessários para geração do modelo.
    c. Os ciclos de leitura do índice não excedem os ciclos de Leitura 1 e Leitura 2.

Observação: Certifique-se de especificar o número correto de ciclos de leitura de índice para as bibliotecas que estão sendo sequenciadas.

5. [Opcional] Selecione *Editar* para alterar os parâmetros de corrida. Quando terminar, selecione *Salvar*.
    a. Parâmetros de corrida: Altere o tipo de leitura ou o número de ciclos por leitura.
    b. Primers personalizados: Altere as configurações para primers personalizados.

Observação: Esses parâmetros devem ser ajustados apenas se forem usados índices mais longos. Por exemplo, se os índices forem de 10bp em vez dos 8bp padrão, os ciclos precisam ser ajustados de 151bp para 149bp para compensar os ciclos disponíveis no cartucho.

6. Selecione Avançar.

Observação: Não abra a porta do compartimento de reagentes ou da flowcell durante a verificação automatizada ou durante a corrida de sequenciamento.

6.8.5 Revisar Verificação Pré-corrida

1. Revise os resultados da verificação automatizada.
    a. Para interromper uma verificação em andamento, selecione *Cancelar*.

CÓPIA CONTROLADA

b. Se houver itens que não forem aprovados, será necessária uma ação corretiva antes de prosseguir.
c. Para reiniciar a verificação, selecione *Retry*. A verificação continuará a partir da primeira verificação incompleta ou com falha.

2. Para iniciar a corrida, selecione uma das seguintes opções:
   a. Se o sistema não estiver configurado para iniciar automaticamente após uma verificação bem-sucedida, selecione *Iniciar*.
   b. Se o sistema estiver configurado para iniciar automaticamente após uma verificação bem-sucedida, a corrida do sequenciamento começará automaticamente. Você não precisa estar presente. No entanto, se ocorrerem erros durante a verificação, a corrida não será iniciada automaticamente.

6.8.6 Monitorar o progresso da corrida

1. Monitore o progresso da corrida, as intensidades e as pontuações de qualidade conforme as métricas aparecem na tela (Figura 9).

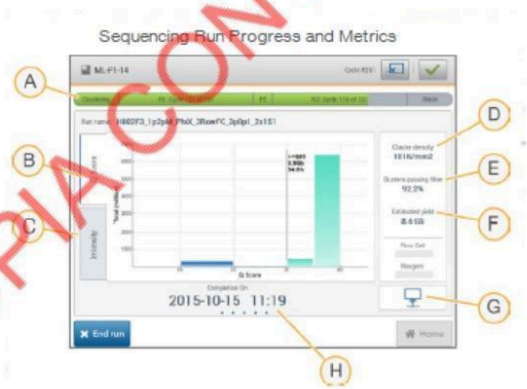


Figura 6: Progresso e métricas da corrida de sequenciamento

A: Progresso da corrida – Mostra a etapa atual e o número de ciclos concluídos para cada leitura. A barra de progresso não é proporcional à taxa de corrida de cada etapa.

B: Q-Score – Mostra a distribuição das pontuações de qualidade (Q-scores). Veja Pontuação de qualidade.

CÓPIA CONTROLADA

C: Intensidade – Mostra o valor das intensidades de cluster do 90º percentil para cada tile. As cores no gráfico representam cada base: vermelho é A, verde é C, azul é G e preto é T.

D: Densidade de cluster (K/mm²) – Mostra o número de clusters detectados na corrida.

E: Clusters que passam no filtro (%) – Mostra a porcentagem de clusters que passam no filtro. Veja Clusters que passam no filtro (Página 44).

F: Rendimento estimado (Gb) – Mostra o número de bases projetadas para a corrida.

G: Status da transferência de dados – Mostra o status da transferência de dados com base na configuração da análise.

H: Tempo para conclusão – Mostra a data e a hora de conclusão da corrida (aaaa-mm-dd hh).

Observação: Após selecionar Home, não é possível retornar para visualizar as métricas de corrida. No entanto, as métricas de corrida são acessíveis no BaseSpace, de um computador em rede usando o Sequencing Analysis Viewer, ou de um computador em rede usando o Local Run Manager.

6.8.7 Lavagem pós-corrida

1. Prepare uma solução de limpeza com tween 20 e água ultrapura seguindo os seguintes procedimentos:
    a. Adicione 5ml de Tween 20 em 45 mL de água ultrapura, esse volume resultará em uma solução 10% de tween 20.
    b. Adicione 25mL da solução 10% de tween 20 a 475mL de água ultrapura resultando em uma solução 0.5% tween 20 essa s
    c. Inverta 5 vezes para misturar
2. Após o preparo da solução de Limpeza, adicione 6mL de solução de lavagem a cada poço de lavagem e 500mL a garrafa de lavagem.
3. Clique em “Iniciar Lavagem”
4. Quando a execução for concluída, selecione Start Wash (Iniciar limpeza).
5. O software levanta automaticamente os aspiradores de líquidos do refrigerador de reagentes. Não selecione *Perform optional template line wash* (Executar limpeza opcional da linha do modelo) na tela Post-Run wash (Limpeza após a execução). A limpeza da linha de modelo requer um procedimento diferente.
6. Abra a porta do compartimento do reagente, a porta do refrigerador de reagentes e remova o cartucho de reagente usado do refrigerador.

CÓPIA CONTROLADA

7. Deslize a bandeja de limpeza no refrigerador de reagentes até o fundo e feche a porta do refrigerador de reagentes.
8. Levante a alça do aspirador de líquidos na frente do frasco do PR2 e do frasco de resíduos até que ela trave.
9. Remova o frasco do PR2 e substitua-o pelo frasco de limpeza.
10. Descarte o frasco do PR2 após cada execução. Não reutilize nenhum PR2 que sobrar.
11. Remova o frasco de resíduos e descarte o conteúdo de forma apropriada. Coloque o frasco de resíduos de volta no compartimento do reagente.
12. Abaixe lentamente a alça do aspirador de líquidos, deixando que o aspirador desça até os frascos de limpeza e de resíduos.
13. Feche a porta do compartimento do reagente.
14. Selecione Next (Avançar).
15. Quando a limpeza estiver concluída, deixe no instrumento a lâmina de fluxo, a bandeja e o frasco de limpeza usados contendo a solução de limpeza restante.
16. Os aspiradores de líquidos permanecem na posição para baixo, o que é normal. Deixe a solução de limpeza não utilizada na bandeja e no frasco de limpeza, para evitar que os aspiradores de líquidos sequem e que entre ar no sistema.
17. O reservatório será liberado da posição sob o cartucho de reagente.
18. Descarte o conteúdo do reservatório (com os produtos químicos restantes) em um recipiente de resíduos devidamente identificado com o rótulo "Resíduos contendo formamida" e armazene-o em um depósito de resíduos químicos até que sejam feitas as etapas para descarte.
19. Considerar as empresas contratadas externas, para a remoção de resíduos químicos. Se os pods forem esvaziados por meio de purga, podem ser descartados como resíduos de risco biológico no laboratório de controle de qualidade.

CÓPIA CONTROLADA

7. ABREVIATURAS

BSL3	Laboratório Nível de biossegurança 3
TB1	Buffer de marcação 1
TSB	Parada de marcação Amortecedor
TWB	Tampão de lavagem de marcação
BLT	Ligado a contas Transposomas

EDTA	Ácido etilenodiaminotetraacético
EPM	PCR aprimorado Mistura
HT1	Hibridização amortecedor
NaOH	Hidróxido de Sódio
NaCl	Cloreto de Sódio
ALN	Água Livre de Nuclease
PCR	Reação em cadeia da polimerase
RSB	Ressuspensão Amortecedor
POP	Procedimentos operacionail padrão
SPB	Purificação de amostra Contas

8. REFERÊNCIAS

Illumina, 2020. Illumina DNA PrepReference Guide. Disponivel em: https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/illumina_prep/illumina-dna-prep-reference-guide-1000000025416-09.pdf. Acesso 06 Nov 2024.

Illumina, 2020. MiSeq System Guide. Disponivel em: https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-system-guide-for-local-run-manager-15027617-05.pdf. Acesso 06 Nov 2024.

9. DOCUMENTOS DE APOIO

- REVIGET-POP001: Extração de DNA Micobacteriano Utilizando o Kit InstaGene.
- REVIGET-POP002: Controle de Qualidade do DNA Genômico.
- REVIGET-POP004: Análise de Fragmento Utilizando o Sistema TapeStation.
- REVIGET-POP005: Análise de Fragmento Utilizando o Sistema LabChip.

CÓPIA CONTROLADA

10. DISPONIBILIDADE

O original assinado versão de esse documento é mantido por Qualidade Garantia Oficial. Os documentos de treinamento estão disponíveis em pastas locais no laboratório **(adicionar nome do lab e local especifico da pasta)**.

11. HISTÓRICO DO DOCUMENTO

Versão nº.	Data	Localização do histórico de alterações	Autor/ Revisor	Aprovando Oficial	Data de aprovação	Próxima data de revisão
1.0	19.10.2022	Novo SOP em uso	Felicia Wells/	Tim Heupink	19.10.2022	2025

CÓPIA CONTROLADA

12. APENDICES

12.1 APENDICE 1: FLUXO DE TRABALHO RESUMIDO

12.2 APENDICE 2: MAPA DE TRABALHO NORMALIZAÇÃO E TAGMENTAÇÃO

12.3 APENDICE 3: MAPA DE TRABALHO LIMPEZA PÓS-TAGMENTAÇÃO

12.4 APENDICE 4: MAPA DE TRABALHO AMPLIFICAÇÃO DO DNA TAGMENTADO

12.5 APENDICE 5: MAPA DE TRABALHO LIMPEZA DAS BIBLIOTECAS

12.6 APENDICE 6: MAPA DE TRABALHO CONTROLE DE QUALIDADE DAS BIBLIOTECAS

12.7 APENDICE 3: MAPA DE TRABALHO DILUÍÇÃO E POOL DAS BIBLIOTECAS

12.8 APENDICE 3: MAPA DE TRABALHO DESNATURAÇÃO DO POOL DE BIBLIOTECAS E PHIX

12.9 APENDICE 3: MAPA DE TRABALHO PREPARO DO POOL DE SEQUENCIAMENTO E INICIO DA CORRIDA NO MISEQ

CÓPIA CONTROLADA

5.3. APRESENTAÇÃO DE TRABALHOS EM CONGRESSO

5.3.1 Validation Of The Use Of High-Performance Computing At The Evandro Chagas Institute For Genomic Surveillance Of Mycobacterioses

Autores: Davi Josué Marcon[1], Reginaldo Sales da Silva Junior 3, Jedson Ferreira Cardoso[3], Marília Lima da Conceição[1], Luana Nepomuceno Gondim Costa Lima 1 2, Edivaldo Costa Sousa Junior 4, Lourdes Maria Garcez 2 4 , Abhinav Sharma 1 5 , Emilyn Costa Conceição 1 5 , Karla Valéria Batista Lima 1 2.

Afiliação:

1. Evandro Chagas Institute, Bacteriology and Mycology Section, Ananindeua, PA, Brazil.
2. State University of Pará, Institute of Biological and Health Sciences, Graduate Program in Parasitic Biology in the Amazon, Belém, PA, Brazil.
3. Evandro Chagas Institute, Information Technology and Communication Section, Ananindeua, PA, Brazil.
4. Evandro Chagas Institute, Parasitology Section, Ananindeua, PA, Brazil.
5. Department of Science and Innovation - National Research Foundation Centre of Excellence for Biomedical Tuberculosis Research South African Medical Research Council Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cidade do Cabo, África do Sul.

Este resumo foi apresentado em formato de pôster no VI Simpósio Norte Nordeste em Bioinformática e premiado como 2º melhor pôster, o qual apresentou o resultado prévio do artigo 5.1 e resultados desenvolvidos durante os primeiros meses do projeto de mestrado.

Resumo Aprovado:

Title: Validation Of The Use Of High-Performance Computing At The Evandro Chagas Institute For Genomic Surveillance Of Mycobacterioses

Introduction:

In the era of genomics, the increasing accessibility and popularity of sequencing technologies and advancements in bioinformatics tools have paved the way for transformative developments. Despite this progress, challenges persist in the form of the need for standardized processes, specialized knowledge, and robust infrastructure for the analysis of extensive datasets. Much like the imperative for monitoring Covid-19 transmission, investing in genomic surveillance for combating mycobacterial diseases presents both a challenge and an avenue for significant development. The IEC-PA (Instituto Evandro Chagas, Pará) boasts a high-performance computing cluster featuring a three-blade functional architecture, with an initial focus on viral pathogen analysis but with the inherent potential for broader microorganism analysis.

Objective:

This study aims to validate the utility of the computational infrastructure at the Evandro Chagas Institute, Pará (IEC-PA), as a strategic approach for genomic surveillance of Mycobacterioses in Brazil. This involves leveraging high-performance computing and cutting-edge techniques.

Methodology:

Following technical meetings for knowledge exchange and system familiarization, we conducted an analysis of 1691 genomes of Mycobacterium tuberculosis using the Nextflow language (MTBSeq-nf) and the Tower platform. The conda package manager facilitated the seamless installation of essential packages. Both individual and group analyses were employed to assess the performance of the processes.

Results:

The individual analysis, encompassing quality control, mapping, and variant acquisition, consumed a total of 22 days, 21 hours, and 39 minutes. Group analysis (genomic distance calculation) required 4 days, 9 hours, and 26 minutes. Notably, only 1350 genomes passed the stringent quality control criteria (genomic coverage <75%). The computational infrastructure utilized 40 CPUs (X86 architecture) and 100GB of RAM. The adoption of Nextflow and Tower significantly enhanced the optimization of computational resources, streamlining process monitoring and statistical analysis.

Several challenges were identified, including data storage complexities, which demand increased attention to alternative storage solutions and standardized backup protocols. Additionally, the establishment of clear guidelines for cluster utilization, necessitating standardized operating procedures, emerged as a crucial consideration. Furthermore, there is a recognized need to expand existing resources.

Conclusion:

The IEC-PA cluster demonstrates promising potential for large-scale genomic analysis, with a particular focus on the surveillance of mycobacterial diseases such as tuberculosis, leprosy, and other related conditions. Its capabilities extend to contributing valuable insights into transmission patterns, genetic diversity, emergence of novel genotypes, drug resistance diagnosis and monitoring, as well as the identification of new microbial species. Continuous improvements, especially in addressing the identified challenges, will further solidify the cluster's relevance and efficacy for the broader scientific and health communities.

Keywords:

Genomic Surveillance, Mycobacteria, Bioinformatics, High-Performance Computing, Nextflow.

Poster apresentado:

VI SIMPÓSIO NORTE-NORDESTE DE BIOINFORMÁTICA

VALIDATION OF THE USE OF HIGH-PERFORMANCE COMPUTING AT THE EVANDRO CHAGAS INSTITUTE FOR GENOMIC SURVEILLANCE OF MYCOBACTERIAS

Davi Josué Marcon[1] , Reginaldo Sales da Silva Junior[3] , Jedson Ferreira Cardoso[3] , Marilia Lima da Conceição[1] , Luana Neponucemo Gondim Costa Lima[1,2] Edivaldo Costa Sousa Junior[4] , Lourdes Maria Garcez[2,4] , Abhinav Sharma[1,5] , Emilyn Costa Conceição[1,5] . Karla Valéria Batista Lima[1,2]

1 Evandro Chagas Institute, Bacteriology and Mycology Section, Ananindeua, PA, Brazil.
2 State University of Pará, Institute of Biological and Health Sciences, Graduate Program in Parasitic Biology in the Amazon, Belém, PA, Brazil. ;
3 Evandro Chagas Institute, Information Technology and Communication Section, Ananindeua, PA, Brazil.
4 Evandro Chagas Institute, Parasitology Section, Ananindeua-PA, Brazil.
5 Department of Science and Innovation - National Research Foundation Centre of Excellence for Biomedical Tuberculosis Research South African Medical Research Council Centre for Tuberculosis Research, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cidade do Cabo, África do Sul.

Introduction

In the era of genomics, the increasing accessibility and popularity of sequencing technologies and advancements in bioinformatics tools have paved the way for transformative developments. Despite this progress, challenges persist in the form of the need for standardized processes, specialized knowledge, and robust infrastructure for the analysis of extensive datasets. Much like the imperative for monitoring Covid-19 transmission, investing in genomic surveillance for combating mycobacterial diseases presents both a challenge and an avenue for significant development. The IEC-PA (Instituto Evandro Chagas, Pará) boasts a high-performance computing cluster featuring a three-blade functional architecture, with an initial focus on viral pathogen analysis but with the inherent potential for broader microorganism analysis.

Objective

This study aims to validate the utility of the computational infrastructure at the Evandro Chagas Institute, Pará (IEC-PA), as a strategic approach for genomic surveillance of Mycobacterioses in Brazil. This involves leveraging high-performance computing and cutting-edge techniques.

Methodology

Job submission and Package managment

nextflow

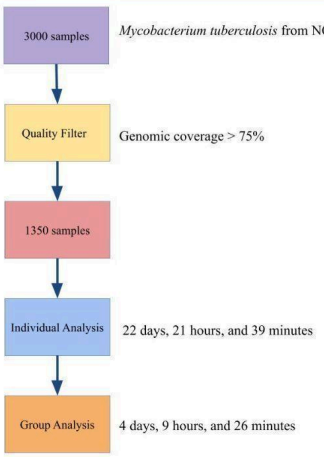
CONDA®

Analysis Pipeline

mycobactopia-org / MTBseq-nf

MTBSeq made simple and easy using Nextflow.

Results



Conclusions

- The IEC-PA cluster shows promising potential for large-scale genomic analysis.
- It focuses particularly on the surveillance of mycobacterial diseases such as tuberculosis and leprosy.
- Its capabilities include identifying transmission patterns, genetic diversity, emerging genotypes, drug resistance diagnosis and monitoring, as well as the discovery of new microbial species.
- Ongoing improvements are needed to address identified challenges.
- It holds significance for both the scientific and healthcare communities.

Acknowledgements

INSTITUTO EVANDRO CHAGAS
MS-SECRETARIA DE VIGILÂNCIA EM SAÚDE

UEPA
UNIVERSIDADE DO ESTADO DO PARÁ

Stellenbosch UNIVERSITY IYUNIVESITHI UNIVERSITEIT

CAPES

CNPq

5.3.2 Simulação Computacional De Variantes Relacionadas À Resistência À Bedaquilina Em Mycobacterium Tuberculosis

Autores: Danna Karen Corrêa dos Santos 1, Davi Josué Marcon 2,3 , José Anderson Monteiro Pantoja 4, Karla Valéria Batista Lima 2,3,4.

Afiliação:

1. Faculdade de Biomedicina, Universidade do Estado do Pará. Belém, Pará, Brasil
2. Seção de Bacteriologia do Instituto Evandro Chagas, Ananindeua, Pará,Brasil
3. Programa de Pós graduação em Biologia Parasitária na Amazônia, Belém, Brasil
4. Faculdade de Biologia, Universidade do Estado do Pará, Belém, Pará, Brasil

Este resumo foi publicado nos anais do IX Congresso de Norte/Nordeste de Infectologia, tendo sido selecionado para apresentação oral e classificado como um dos 6 finalistas de melhor apresentação.

O qual apresentou os primeiros resultados do Trabalho de Conclusão de Curso: **Análise estrutural in silico baseada em modelagem molecular: Explorando variantes de resistência à bedaquilina em Mycobacterium tuberculosis.**

Resumo Aprovado

Simulação computacional de variantes relacionadas à resistência à bedaquilina em *Mycobacterium tuberculosis*

Introdução/Objetivo: A emergência de estirpes resistentes de *Mycobacterium tuberculosis* tem agravado os desafios terapêuticos, especialmente com a resistência a novos fármacos, como a bedaquilina (BDQ). Utilizada no tratamento de pacientes com tuberculose resistente à rifampicina, multirresistente ou extensivamente resistente, a eficácia da BDQ vem sendo comprometida pela resistência emergente. Este estudo tem como objetivo analisar as modificações estruturais de proteínas do genoma de *M. tuberculosis* relacionadas à resistência à BDQ, utilizando análises *in silico*, com o intuito de explorar as variantes genéticas responsáveis pela resistência ao fármaco.

Métodos: Foram analisados 41 isolados de *M. tuberculosis* com Concentração Inibitória Mínima (CIM) ≥1 e ≥2, selecionados de 12.289 amostras do banco de dados do consórcio CRyPTIC. A identificação de variantes foi realizada utilizando o pipeline MAGMA 1.1.1 para alinhamento e busca por mutações. A modelagem estrutural foi feita com o ColabFold 1.5.5, e a docagem molecular com o AutoDock 4.2, a visualização das interações proteína-ligante foi realizada com LigPlot+ 2.2 e ChimeraX 1.8. Foram avaliados parâmetros de energia livre de ligação (ΔG) e constante de inibição (Ki) para correlacionar com os valores de CIM.

Resultados: Detectaram-se 19 variantes nos genes *mmpL5*, *Rv0678*, *atpE*, *mtrB*, *Rv1979c*, e *lpqB*, que estão potencialmente associadas à resistência à BDQ. Embora descritos no catálogo da OMS como relacionados à resistência, não foram observadas mutações nos genes *mtrA*, *mmpS5* e *pepQ* nas cepas avaliadas. Variantes como atpE_var2 apresentaram valores elevados de Ki (653,32 μM), indicando resistência. As variantes de *Rv0678* mostraram um padrão interessante, com Rv0678_var1 exibindo Ki = 145.95 μM, e Rv0678_var2 com Ki = 531.79 μM, ambas demonstrando resistência significativa, enquanto Rv0678_var6 apresentou maior afinidade pela BDQ (ΔG = -6,13 kcal/mol) e menor Ki (31,89 μM), sugerindo baixo nível de resistência.

Conclusão: O estudo forneceu contribuições inéditas quanto ao uso da simulação computacional no estudo da resistência e das interações entre as variantes de *M. tuberculosis* e BDQ, destacando a importância da modelagem e docagem para predição e correlação de mutações com a resistência. Os achados podem ser aplicados para o monitoramento contínuo da resistência, contribuindo para um melhor entendimento dos mecanismos moleculares que influenciam a eficácia da BDQ.

5.3.3 Vigilância Genômica Revela Perfil de Transmissão dos primeiros Casos de Tuberculose Extensivamente Resistentes no Estado do Pará

Autores: Davi Josué Marcon 1,2, Abhinav Sharma 3, Alex Brito Souza 2, Rafaella Bonfim Barros4, Valnete das Graças Dantas Andrade4, Lúcia Helena Martins Tavares Monteiro 5, Luana Nepomuceno Gondim Lima 1,2, Ricardo José de Paula Souza e Guimarães 6, Emilyn Costa Conceição 3, Carlos Augusto Abreu Alberio7, Karla Valéria Batista Lima 1,2.

Afiliação:

1. Seção de Bacteriologia do Instituto Evandro Chagas, Ananindeua, Pará,Brasil
2. Programa de Pós graduação em Biologia Parasitária na Amazônia, Belém, Brasil
3. Department of Science and Innovation - National Research Foundation Centre of Excellence for Biomedical Tuberculosis Research, SAMRC Centre for Tuberculosis Research, Division of MBHG, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa.
4. Laboratório Central do Pará, LACEN-PA, Belém, Pará, Brasil.
5. Coordenadora do Programa de Controle da Tuberculose, Secretaria de Estado de Saúde Publica do Pará, Belém, Pará, Brasil.
6. Laboratório de Geoprocessamento do Instituto Evandro Chagas, Ananindeua, Pará,Brasil
7. Hospital Universitário João de Barros Barreto, Universidade Federal do Pará, Belém, Pará, Brasil.

Este trabalho foi apresentado no 59º Congresso da Sociedade Brasileira de Medicina tropical, e compreende aos resultados preliminares da dissertação, tendo sido apresentado em formato de Pôster.

Resumo Aprovado

Eixo 13 | Tuberculoses e outras Microbactérias Humanas e Veterinárias

Vigilância Genômica Revela Perfil de Transmissão dos Primeiros Casos de Tuberculose Extensivamente Resistente no Estado do Pará

Resumo:

Com a introdução dos novos medicamentos no tratamento da tuberculose droga resistente (TBDR) a partir de 2021 no Brasil, tornou-se crucial a vigilância genômica (VG) para monitoramento e disseminação da resistência antimicrobiana. Este estudo objetivou iniciar a VG de Micobactérias provenientes de indivíduos tratados para TBDR entre out/2021 e dez/2022 no estado do Pará. O projeto foi aprovado pelo Comitê de ética/IEC (parecer 3.950.565). O DNA micobacteriano foi extraído no Laboratório Central do Pará, as bibliotecas foram preparadas com o kit Illumina Nextera-XT e o sequenciamento genômico (WGS) foi realizado no NextSeq 550. Os genomas foram analisados utilizando o MAGMA v.3.1 e TB-Profiler v.6.10. Para análise de transmissão foram utilizados os cutoffs de 5 e 12 Single Nucleotide Polymorphisms (SNPs). Também foi realizada distribuição espacial dos casos. Dos 103 casos notificados como TBDR, 87 obtiveram cultura viável, das quais realizamos WGS para 40 isolados (casos). Os dados genômicos de duas amostras sugerem eventos de infecção mista (duas linhagens de *Mycobacterium tuberculosis* L1 e L4) e de co-infecção *M. tuberculosis* e *M. chelonae* que foram analisados separadamente. O perfil genético de sensibilidade antimicrobiana obtido pela VG evidenciou que 7,5% (N=3) dos isolados não apresentavam os genes de resistência conhecidos ou definidos como alta significância segundo a Organização Mundial da Saúde, 20% (N=8) apresentaram monorresistência à rifampicina (RIF) ou isoniazida (INH). Multidroga-resistência (MDR) foi evidenciada em 47,5% (N=19), pré-XDR (pré-TB extensivamente resistente em 20% (N=8) e extensivamente resistente (XDR) em 5% (N=2) com resistência à bedaquilina (BDQ) e clofazimina em um isolado e resistência à Linezolida (LZD) em outro. Dos 38 isolados de *M. tuberculosis*, 36.84% (N=14) foram agrupados em cinco clusters (C1-C5) com cutoff de 12 SNPs sugerindo transmissão da TB entre estes. Deste, o perfil de resistência por cluster foi: RIF (C1), MDR (C2), XDR e pré-XDR (C3), MDR (C4 e C5). Os membros do C2 residem no município de Marabá e os demais na capital Belém. Este é o primeiro relato de resistência à BDQ e LZD no Brasil com confirmação genômica de transmissão da TB pré-XDR e XDR já no primeiro ano de inclusão dos novos medicamentos (BDQ e LZD) no esquema de tratamento da TBDR, reforçando a importância da VG para precisão em saúde pública. Sugerimos atenção de saúde pública nas áreas com casos pré-XDR e XDR para rastreamento de contatos.

Palavras-chave: *Mycobacterium tuberculosis,* Micobactérias Não Tuberculosas, Vigilância Genômica, Bedaquilina, Linezolida, Tuberculose Extensivamente Resistente.

Agências Financiadoras: Capes, CNPq, Laboratório Central do Pará, Instituto Evandro Chagas, Secretaria de vigilância em saúde e ambiente/Ministério da Saúde.

Poster apresentado:

MEDTROP
59º Congresso da SBMT • São Paulo | SP
Medicina Tropical sob olhar de Saúde Única
22 a 25 setembro 2024

Vigilância Genômica Revela Perfil de Transmissão dos Primeiros Casos de Tuberculose Extensivamente Resistente no Estado do Pará

Davi Josué Marcon[1,2*], Abhinav Sharma, Alex Brito Souza1,2, Rafaella Bonfim Barros Valnete das Graças Dantas Andrade4, Lúcia Helena Martins Tavares Monteiro5, Luana Nepomuceno Gondim Lima1,2, Ricardo José de Paula Souza e Guimarães6, Emilyn Costa Conceição3, Carlos Augusto Abreu Alberio7 e Karla Valéria Batista Lima1,2

1. SEBAC-IEC, Ananindeua, Pará,Brasil; **2.** PPGBPA, Belém, Brasil; **3.** SAMRC Centre for Tuberculosis Research, Division of MBHG, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa; **4.** LACEN-PA, Belém, Pará, Brasil; **5.** PCT, SESPA, Belém, Pará, Brasil; **6.** SAEPI-IEC, Ananindeua, Pará,Brasil; **7.** HUJBB-UFPA, Belém, Pará, Brasil. *davijosuemarcon@gmail.com

INTRODUÇÃO

Com a introdução dos novos medicamentos no tratamento da tuberculose droga resistente (TBDR) a partir de 2021 no Brasil, tornou-se crucial a vigilância genômica (VG) para monitoramento e disseminação da resistência antimicrobiana incluindo a busca por mutações relacionadas a resistência aos novos fármacos

OBJETIVO

Este estudo objetivou iniciar a VG de Micobactérias provenientes de indivíduos tratados para TBDR entre out/2021 e dez/2022 no estado do Pará.

MATERIAL E MÉTODOS



RESULTADOS

Figura 1 - Árvore filogenética

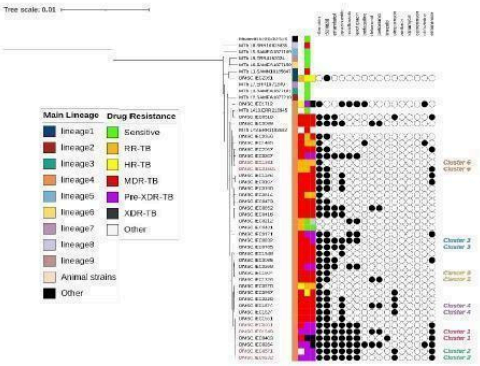
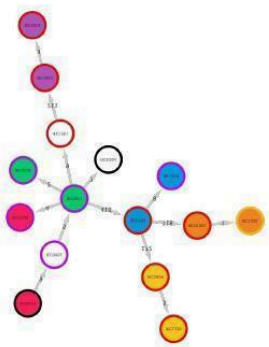
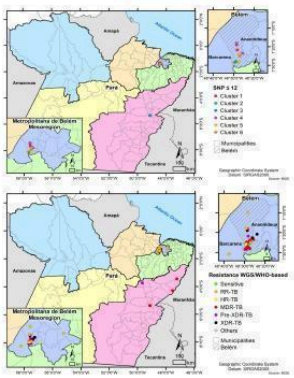


Figura 2 - Reconstrução de rede de transmissão



Figuras 3 e 4 - Distribuição de clusters de transmissão e perfis de susceptibilidade



CONCLUSÕES

Este é o primeiro relato de resistência à BDQ e LZD no Brasil com confirmação genômica de transmissão da TB pré-XDR e XDR já no primeiro ano de inclusão dos novos medicamentos (BDQ e LZD) no esquema de tratamento da TBDR, reforçando a importância da VG para precisão em saúde pública. Sugerimos atenção de saúde pública nas áreas com casos pré-XDR e XDR para rastreamento de contatos.

FINANCIAMENTO

CAPES
LACEN LABORATÓRIO CENTRAL DO ESTADO DO PARÁ
CNPq
INSTITUTO EVANDRO CHAGAS
MS-SECRETARIA DE VIGILÂNCIA EM SAÚDE

PRINCIPAL REFERÊNCIA

BRASIL. **NOTA INFORMATIVA No 9/2021-CGDR/.DCCI/SVS/MS**. Brasília: Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Doenças de Condições Crônicas e Infecções Sexualmente Transmissíveis, Coordenação-Geral de Vigilância das Doenças de Transmissão Respiratória de Condições Crônicas, 2021.

5.3.4 Caso complexo de Tuberculose e Micobacteriose (TB/MNT) em referência no Estado do Pará.

Autores: Davi Josué Marcon 1,2, Danna Karen 2, Alex Brito Souza 2, Jeanne Cabral 2, Rafaella Bonfim Barros 4 ,Abhinav Sharma 3 Valnete das Graças Dantas Andrade 4, Lúcia Helena Martins Tavares Monteiro 5, Carlos Augusto Abreu Alberio 7, Karla Valéria Batista Lima 1,2.Emilyn Costa Conceição 3

Afiliação:

1. Seção de Bacteriologia do Instituto Evandro Chagas, Ananindeua, Pará,Brasil
2. Programa de Pós graduação em Biologia Parasitária na Amazônia, Belém, Brasil
3. Department of Science and Innovation - National Research Foundation Centre of Excellence for Biomedical Tuberculosis Research, SAMRC Centre for Tuberculosis Research, Division of MBHG, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa.
4. Laboratório Central do Pará, LACEN-PA, Belém, Pará, Brasil.
5. Coordenadora do Programa de Controle da Tuberculose, Secretaria de Estado de Saúde Publica do Pará, Belém, Pará, Brasil.
6. Laboratório de Geoprocessamento do Instituto Evandro Chagas, Ananindeua, Pará,Brasil
7. Hospital Universitário João de Barros Barreto, Universidade Federal do Pará, Belém, Pará, Brasil.

Este trabalho foi apresentado no 59º Congresso da Sociedade Brasileira de Medicina tropical, e compreende aos resultados preliminares da dissertação, tendo sido apresentado em formato de Pôster.

Resumo Aprovado

Eixo 13 | Tuberculoses e outras Microbactérias Humanas e Veterinárias

Caso complexo de coinfecção TB-MNT em referência no estado do Pará

Objetivos:

Descrever um caso de infecção pulmonar causada por M tuberculosis e micobactéria não tuberculose.

Relato do caso:

Paciente diabético de 46 anos, sexo masculino, apresentou sintomas sugestivos de TB em mar/2022, sendo diagnosticado com tuberculose pulmonar - motivo pelo qual se iniciou tratamento com esquema padrão RHZE. A radiografia de tórax realizada no momento do diagnóstico apresentou cavitação unilateral.

Dois testes Xpert MTB/RIF realizados em abril/2022 resultaram em presença de MTB e resistência à rifampicina, resultando em alteração no esquema de tratamento para seis meses utilizando bedaquilina (Bdq), levofloxacino (Lfx), linezolida (Lzd) e terizidona (Trd) complementado com Vitamina B6 (B6), com fase manutentiva de 12 meses com Lfx, Lzd, Trd e B6. No terceiro mês de tratamento, foi solicitado um teste de sensibilidade antimicrobiana (TSA), que resultou em resistência fenotípica à isoniazida, rifampicina e etambutol, com sensibilidade à estreptomicina. O TSA de 2a linha foi rejeitado em jan/2023 por apresentar resultados inconsistentes e novas Identificações moleculares foram realizadas em dez/22 e mai/23 confirmando coinfecção causada por M. *abscessus*.As baciloscopias foram negativas durante todo o curso do tratamento. As culturas de acompanhamento tiveram resultado variado, alternando entre positivas e negativas. Sendo negativas do 5º ao 8º mês, positiva no 9º, novamente negativas no 10º e 11º mês e positivas nos meses subsequentes até o 15º mês de tratamento. A vigilância genômica reavaliou a amostra inicial de 2022 em mai/2024, identificando a cepa como M. chelonae com mutações que predizem resistência a beta-lactâmicos, aminoglicosídeos e rifamicinas. Ao observar os resultados, sugere-se uma infecção inicial mista causada por *M. tuberculosis* e *M. chelonae*, seguida de uma sucessão de infecção por *M. abcessus.*

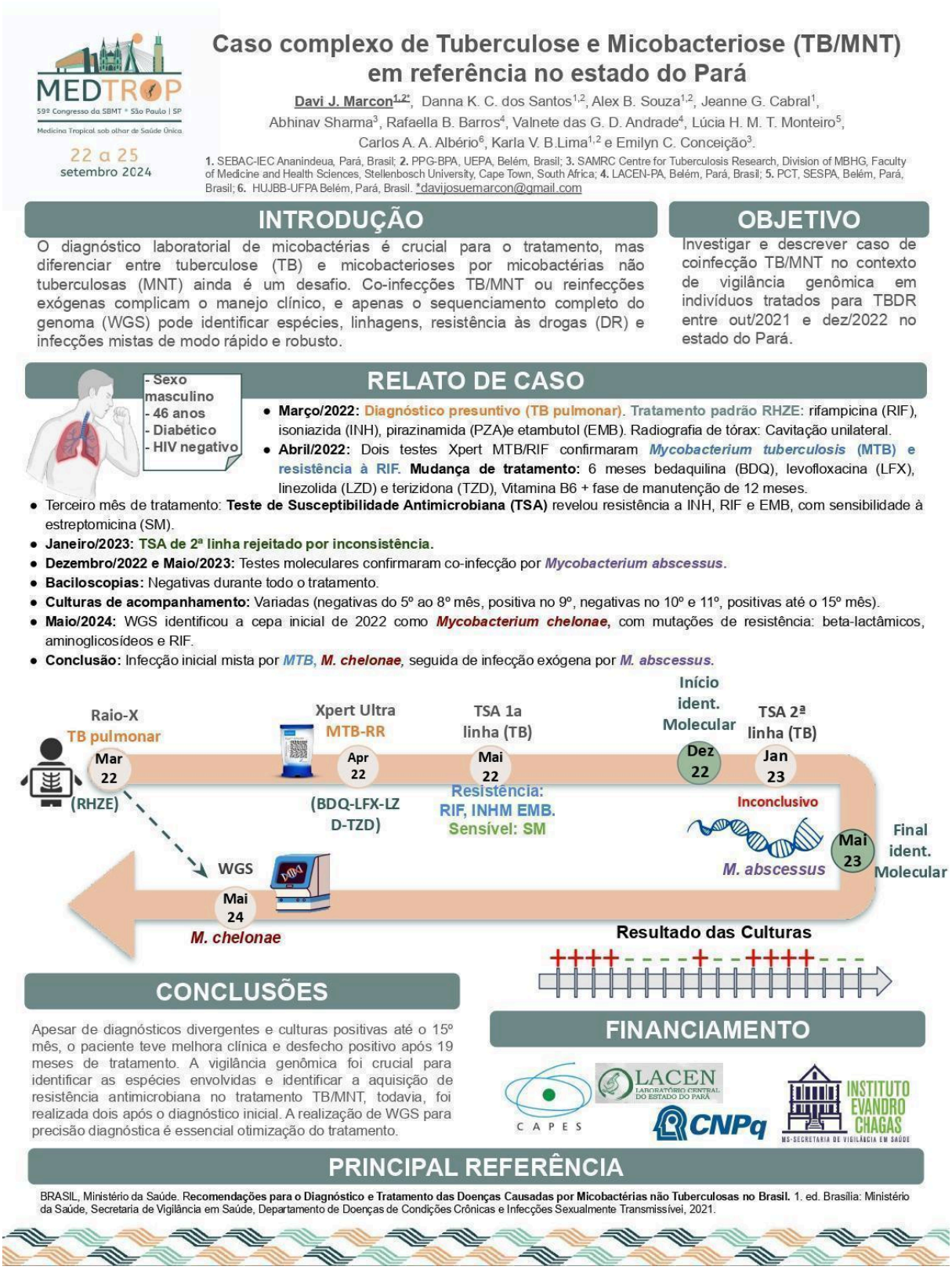
Conclusão:

Mesmo com resultados divergentes no diagnóstico e com culturas positivas até o 15º mês de tratamento, o paciente apresentou melhora clínica e desfecho favorável após 19 meses de tratamento. A vigilância genômica colaborou com o esclarecimento das espécies envolvidas e o na investigação de casos complexos bem como a efetividade do tratamento baseado em BDQ para infecções causadas por TB e MNT.

Palavras-chave: *Mycobacterium tuberculosis,* Micobactérias Não Tuberculosas, Vigilância Genômica, Bedaquilina, Linezolida, Identificação molecular.

Agências Financiadoras: Capes, CNPq, Laboratório Central do Pará, Instituto Evandro Chagas, Secretaria de vigilância em saúde e ambiente/Ministério da Saúde.

Poster apresentado:



5.3.5 From Research to Surveillance: Leveraging the State of Whole Genome Sequencing for Precision Public Health in a High-Burden Tuberculosis Setting

Autores: Karla Valéria Batista Lima 1,2, Abhinav Sharma 3, Davi Josué Marcon 1,2, Ricardo José de Paula Souza e Guimarães 1, Annelies van Rie 4, Robin Warren 3, Philip Noel Suffys 5, Emilyn Costa Conceição 3.

Afiliação:
1. Instituto Evandro Chagas, Ananindeua-PA, Brazil.
2. Universidade do Estado do Pará, Instituto de Ciências Biológicas e da Saúde, Pós-Graduação em Biologia Parasitária na Amazônia, Belém-PA, Brazil.
3. SAMRC, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa.
4. Faculty of Medicine and Health Sciences, University of Antwerp, Antwerp, Belgium
5. Laboratório de Bacteriologia, Instituto de Infectologia Evandro Chagas, Fundação Oswaldo Cruz, Rio de Janeiro-RJ, Brazil.

O trabalho derivado do projeto "**GEMIBRA - Genomas de Micobactérias do Brasil**", do qual o discente faz parte, foi apresentado em dois eventos internacionais: Congresso Europeu de Micobacteriologia (2024) e Nextflow Summit (2024).

Resumo Aprovado

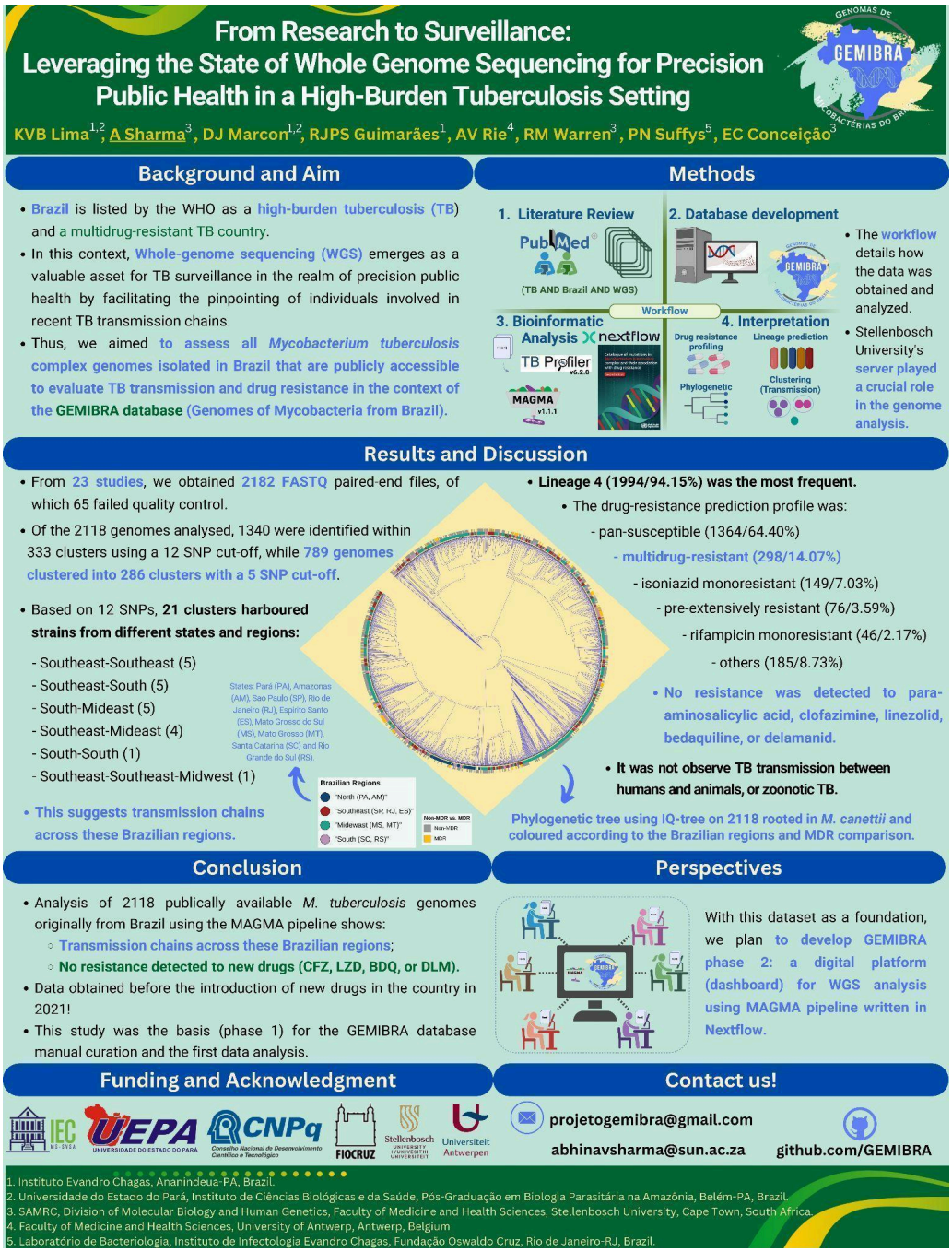
From Research to Surveillance: Leveraging the State of Whole Genome Sequencing for Precision Public Health in a High-Burden Tuberculosis Setting

Abstract 250/250 words

Tuberculosis (TB) continues to pose a significant public health challenge in 30 high-burden TB countries, including Brazil. Whole-genome sequencing (WGS) offers promising prospects in the realm of precision public health. By facilitating the pinpointing of individuals involved in recent TB transmission chains, WGS emerges as a valuable asset for TB surveillance. Through a literature review, we aimed to assess all *Mycobacterium tuberculosis* complex genomes isolated in Brazil that are publicly accessible to evaluate TB transmission and drug resistance. We employed the MAGMA pipeline utilizing two distinct cut-off analyses involving 5 and 12 Single Nucleotide Polymorphisms (SNPs). Drug resistance profiling (DRP) was obtained using TB-Profiler v.6.2.0. From 23 studies, we obtained 2182 FASTQ paired-end files, of which 65 failed quality control. Of the 2118 genomes analysed, 1340 were identified within 333 clusters using a 12 SNP cut-off, while 789 genomes clustered into 286 clusters with a 5 SNP cut-off. Based on 12 SNPs, 21 clusters harboured strains from different states and regions: Southeast-Southeast (5), Southeast-South (5), South-Mideast (5), Southeast-Midwest (4), South-South (1), and Southeast-Southeast-Midwest (1). This suggests transmission chains across these Brazilian regions. Lineage 4 (1994/94.15%) was the most frequent. The DRP was: pan-susceptible (1364/64.40%), multidrug-resistant (298/14.07%), mono-isoniazid (149/7.03%), pre-extensively resistant (76/3.59%), mono-rifampicin (46/2.17%) and others (185/8.73%). No resistance was detected to para-aminosalicylic acid, clofazimine, linezolid, bedaquiline, or delamanid. This is preliminary data that will be accessible for research and TB surveillance purposes in Brazil within the GEMIBRA (Genome of Mycobacteria in Brazil) web platform v.2 which is currently under development.

Keywords: Tuberculosis, Whole genome sequencing, Precision Public Health, Genomic Surveillance, Brazil.

Poster Apresentado:



5.4 CURSOS MINISTRADOS

5.4.1 1o Minicurso de Introdução à Bioinformática:

Realizado na seção de Bacteriologia e Micologia do Instituto Evandro Chagas, teve como objetivo introduzir conceitos básicos de bionformática aos colaboradores envolvidos em projetos de genômica bacteriana. O discente ministrou aulas introdutórias sobre *pipelines*, e servidores, além de realizar atividade prática de introdução a linha de comando para análise de genomas bacterianos.

5.4.2 1º Workshop da Rede de Vigilância Norte e Nordeste:

O discente participou ministrando a palestra "**Acelerando o fluxo de diagnóstico a partir do Escarro**" no LACEN-PA e na Fiocruz Bahia, a última disponível online em: https://www.youtube.com/watch?v=eSn9s6Kv_cg entre os minutos 2:07:00 e 2:57:00.

Além disso, participou ministrando o treinamento prático e teórico em extração de DNA, preparo de bibliotecas, sequenciamento genômico e bioinformática nos sítios de treinamento LACEN-PA e LACEN-BA.

Os certificados de palestras e treinamentos deste evento ainda encontram-se em fase de produção.

5.4.3 Rotinas Moleculares: Da Extração ao Sequenciamento Capilar

O discente ministrou aulas teórico-práticas durante o treinamento realizado em 2024 na seção de Bacteriologia do Instituto Evandro Chagas, no qual foram introduzidas técnicas de avaliação e interpretação de resultados de sequenciamento capilar e identificação de organismos utilizando análise de homologia de sequências.

Os certificados deste treinamento ainda encontram-se em fase de produção.

6 CONCLUSÕES

A implementação do *wrapper* MTBSeq-nf tornou mais acessível o uso do software MTBseq, permitindo sua execução em diversas nuvens como Azure, AWS e Google Cloud. Além disso, o uso de parâmetros centrados no usuário e linhas de comando simplificadas, extenderam as funcionalidades originais do programa. A análise posterior dos resultados, demonstrou que o uso paralelizaclismo na execução, não alterou de maneira significativa os resultados, sendo estes reprodutiveis nas plataformas testadas.

Os primeiros relatos de TB-XDR e a Associação do SIT2517 com estudos anteriores, revelam a evolução da resistência da tuberculose no Pará, o SIT raro e apenas descrito no Pará, anteriormente só era relacionado a TB-MDR, e neste estudo, foi relacionado a pré-XDR. Com perfil de resistência primária, ou seja, transmitida de um indivíduo para outro, este perfil pré-XDR, esteve relacionado a posterior aquisição de TB-XDR, com o surgimento de mutações de baixa frequência em indivíduos da coorte estudada.

O uso translacional do WGS para a investigação de casos complexos em estudo de caso, demonstrou de maneira prática a utilidade da ferramenta para a saúde individualizada, permitindo que, sejam identificados fatores genéticos associados a complexidade do caso. A identificação de MNTs ou múltiplos clones de TB em amostra pode ser feita de maneira efetiva com a especificidade que técnicas baseadas em cultivo não permitiriam.

Complementarmente aos trabalhos publicados, o desenvolvimento de POPs e treinamentos baseados nestes, são marcos de contribuição do estudo piloto para o projeto para o fortalecimento da Rede de Vigilância Norte Nordeste. A qual, expandirá a vigilância genômica no estado do para até 2025, e em outros estados. Os produtos gerados a partir da vigilância genômica na introdução da bedaquilina, são um exemplo do uso de WGS para a saúde pública para o enfrentamento da tuberculose droga-resistente, fornecendo base para a formulação de políticas públicas e justificando a expansão em outros projetos. Através de artigos, resumos, protocolos operacionais padrão e treinamentos realizados.

7. FONTES DE FINANCIAMENTO

Este trabalho contou com financiamento da CAPES, bolsa de Mestrado (88887.825814/2023-00), CNPq bolsa de produtividade em pesquisa Chamada 09/2022, proc. 311165/2022-2; Instituto Evandro Chagas e LACEN-PA com infraestrutura, concessão de reagentes e material de consumo; suporte financeiro do CNPq Edital 29/2023, proc. 445784/2023-7 sob o projeto REVIGET-N/NE e auxílio Oracle sobre o ID 3083687.

REFERÊNCIAS BIBLIOGRÁFICAS

AFOLAYAN, A. O. *et al.* Overcoming Data Bottlenecks in Genomic Pathogen Surveillance. **Clinical Infectious Diseases**, [*s. l.*], v. 73, n. Supplement_4, p. S267–S274, 2021.

ALLUÉ-GUARDIA, A.; GARCÍA, J. I.; TORRELLES, J. B. Evolution of Drug-Resistant Mycobacterium tuberculosis Strains and Their Adaptation to the Human Lung Environment. **Frontiers in Microbiology**, [*s. l.*], v. 12, 2021. Disponível em: https://www.frontiersin.org/articles/10.3389/fmicb.2021.612675. Acesso em: 24 dez. 2022.

ALMEIDA DA SILVA, P. E.; PALOMINO, J. C. Molecular basis and mechanisms of drug resistance in Mycobacterium tuberculosis: classical and new drugs. **Journal of Antimicrobial Chemotherapy**, [*s. l.*], v. 66, n. 7, p. 1417–1430, 2011.

ANDRIES, K. *et al.* A Diarylquinoline Drug Active on the ATP Synthase of Mycobacterium tuberculosis. **Science**, [*s. l.*], v. 307, n. 5707, p. 223–227, 2005.

BAGCCHI, S. WHO's Global Tuberculosis Report 2022. **The Lancet Microbe**, [*s. l.*], v. 4, n. 1, p. e20, 2023.

BRASIL. **BEDAQUILINA PARA PACIENTES COM TUBERCULOSE RESISTENTE À RIFAMPICINA, MULTIRRESISTENTE OU EXTENSIVAMENTE RESISTENTE**. Brasília: [*s. n.*], 2020. (Relatórios para sociedade). v. 208

BRASIL, M. da S. **Boletim Epidemiológico**. Brasília: Secretaria de Vigilância em Saúde, 2022.

BRASIL, M. da S. **Boletim Epidemiológico de Tuberculose - Número Especial**. Brasília: Departamento de HIV/Aids, Tuberculose, Hepatites Virais e Infecções Sexualmente Transmissíveis Secretaria de Vigilância em Saúde e Ambiente, 2024. Disponível em: https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/boletins/epidemiologicos/especiais/2024/boletim-epidemiologico-de-tuberculose-numero-especial-mar-2024.pdf. .

BRASIL, M. da S. **Manual de recomendações para o controle da tuberculose no Brasil**. 2. ed. Brasília: Editora MS, 2019.

BRASIL. **Manual de Recomendações para o Diagnóstico Laboratorial de Tuberculose e**

Micobactérias não Tuberculosas de Interesse em Saúde Pública no Brasil. 1. ed. Brasília: Ministério da Saúde, 2022.

BRASIL. **NOTA INFORMATIVA Nº 1/2025-CGTM/.DATHI/SVSA/MS**. Brasília: Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Doenças de Condições Crônicas e Infecções Sexualmente Transmissíveis, Coordenação-Geral de Vigilância das Doenças de Transmissão Respiratória de Condições Crônicas, 2025.

BRASIL. **NOTA INFORMATIVA Nº 9/2021-CGDR/.DCCI/SVS/MS**. Brasília: Ministério da Saúde, Secretaria de Vigilância em Saúde, Departamento de Doenças de Condições Crônicas e Infecções Sexualmente Transmissíveis, Coordenação-Geral de Vigilância das Doenças de Transmissão Respiratória de Condições Crônicas, 2021.

BRASIL. **Pretomanida para o tratamento da tuberculose resistente a medicamentos**. [*S. l.*]: Ministério da saúde, Secretaria de Ciência, Tecnologia, Inovação e Complexo da Saúde - SECTICS Departamento de Gestão e Incorporação de Tecnologias em Saúde - DGITS Coordenação-Geral de Avaliação de Tecnologias em Saúde - CGATS, 2023.

BRASIL, M. da S., Secretaria de Vigilância em Saúde e Ambiente, Departamento de HIV/Aids, Tuberculose, Hepatites Virais e Infecções Sexualmente Transmissíveis. **Recomendações Técnicas para Laudo e Interpretação do Teste de Hibridização com Sonda em Linha (Line Probe Assay – LPA) para Tuberculose**. Brasília: Ministério da Saúde, 2023.

BROSCH, R. *et al.* Use of a Mycobacterium tuberculosisH37Rv Bacterial Artificial Chromosome Library for Genome Mapping, Sequencing, and Comparative Genomics. **Infection and Immunity**, [*s. l.*], v. 66, n. 5, p. 2221–2229, 1998.

COHEN, K. A. *et al.* Deciphering drug resistance in Mycobacterium tuberculosis using whole-genome sequencing: progress, promise, and challenges. **Genome Medicine**, [*s. l.*], v. 11, n. 1, p. 45, 2019.

CONCEIÇÃO, E. C. *et al.* Molecular epidemiology of Mycobacterium tuberculosis in Brazil before the whole genome sequencing era: a literature review. **Memórias do Instituto Oswaldo Cruz**, [*s. l.*], v. 116, 2021. Disponível em: http://www.scielo.br/j/mioc/a/5by7ghP7tGCxby96vBTJhzr/?lang=en. Acesso em: 1 jan. 2023.

CONCEIÇÃO, E. C. *et al.* **Whole genome sequencing for comprehensive management of Bedaquiline-resistant tuberculosis: a case report**. Itália: [*s. n.*], 2022.

CONSORTIUM, Cr. A data compendium associating the genomes of 12,289 Mycobacterium tuberculosis isolates with quantitative resistance phenotypes to 13 antibiotics. **PLOS Biology**, [*s. l.*], v. 20, n. 8, p. e3001721, 2022.

CONSORTIUM, Cr. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. **New England Journal of Medicine**, [*s. l.*], v. 379, n. 15, p. 1403–1415, 2018.

GARCEZ, J. C. D. *et al.* Surveillance Quality Indicators Highlight the Need for Improving Tuberculosis Diagnostics and Monitoring in a Hyperendemic Area of the Brazilian Amazon Region. **Tropical Medicine and Infectious Disease**, [*s. l.*], v. 7, n. 8, p. 165, 2022.

GUIMARAES, A. M. S.; ZIMPEL, C. K. Mycobacterium bovis: From Genotyping to Genome Sequencing. **Microorganisms**, [*s. l.*], v. 8, n. 5, p. 667, 2020.

GUYEUX, C. *et al.* Newly Identified Mycobacterium africanum Lineage 10, Central Africa. **Emerging Infectious Diseases**, [*s. l.*], v. 30, n. 3, p. 560–563, 2024.

HEUPINK, T. H. *et al.* The MAGMA pipeline for comprehensive genomic analyses of clinical Mycobacterium tuberculosis samples. **PLOS Computational Biology**, [*s. l.*], v. 19, n. 11, p. e1011648, 2023.

KIM, S. *et al.* Clinical Characteristics and Treatment Outcomes of Pulmonary Diseases Caused by Coinfections With Multiple Nontuberculous Mycobacterial Species. **Journal of Korean Medical Science**, [*s. l.*], v. 39, n. 20, p. e167, 2024.

MALLICK, J. S. *et al.* Acquired bedaquiline resistance during the treatment of drug-resistant tuberculosis: a systematic review. **JAC-Antimicrobial Resistance**, [*s. l.*], v. 4, n. 2, p. dlac029, 2022.

MERTANIASIH, N. M. *et al.* Nontuberculous mycobacterial species and Mycobacterium Tuberculosis Complex Coinfection in Patients with Pulmonary Tuberculosis in Dr. Soetomo Hospital, Surabaya, Indonesia. **The International Journal of Mycobacteriology**, [*s. l.*], v. 6, n. 1, p. 9, 2017.

MINISTÉRIO DA SAÚDE. **PORTARIA SCTIE-MS Nº 43, DE 6 DE JULHO DE 2021**.

[*S. l.: s. n.*], 2021.

NGABONZIZA, J. C. S. *et al.* A sister lineage of the Mycobacterium tuberculosis complex discovered in the African Great Lakes region. **Nature Communications**, [*s. l.*], v. 11, n. 1, p. 2917, 2020.

NGUYEN, T. V. A. *et al.* Bedaquiline Resistance: Its Emergence, Mechanism, and Prevention. **Clinical Infectious Diseases**, [*s. l.*], v. 66, n. 10, p. 1625–1630, 2018.

NGUYEN, T. V. A. *et al.* Delamanid Resistance: Update and Clinical Management. **Clinical Infectious Diseases**, [*s. l.*], v. 71, n. 12, p. 3252–3259, 2020.

NIMMO, C. *et al.* Opportunities and limitations of genomics for diagnosing bedaquiline-resistant tuberculosis: a systematic review and individual isolate meta-analysis. **The Lancet Microbe**, [*s. l.*], v. 5, n. 2, p. e164–e172, 2024.

OMS. **Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance**. 1. ed. Geneva: Organização Mundial da SAúde, 2022a. Disponível em: https://www.who.int/publications-detail-redirect/9789240028173. Acesso em: 23 dez. 2022.

OMS. **Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance**. 1. ed. Geneva: Organização Mundial da Saúde, 2022b. Disponível em: https://www.who.int/publications-detail-redirect/9789240028173. Acesso em: 23 dez. 2022.

OMS. **Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance, second edition**. Geneva: Organização Mundial da Saúde - OMS, 2023.

OMS. **Technical manual for drug susceptibility testing of medicines used in the treatment of tuberculosis**. Geneva: Organização Mundial da Saúde, 2018a. Disponível em: https://apps.who.int/iris/handle/10665/275469. Acesso em: 24 dez. 2022.

OMS. **The use of next-generation sequencing technologies for the detection of mutations associated with drug resistance in Mycobacterium tuberculosis complex: technical guide**. Geneva: Organização Mundial da SAúde, 2018b.

OMS. **WHO announces updated definitions of extensively drug-resistant tuberculosis**. [*S. l.*], 2021. Disponível em: https://www.who.int/news/item/27-01-2021-who-announces-updated-definitions-of-extensively-drug-resistant-tuberculosis. Acesso em: 24 dez. 2022.

OMS. **WHO operational handbook on tuberculosis. Module 4: treatment - drug-resistant tuberculosis treatment**. Geneva: Organização Mundial da SAúde, 2022c.

RIOJAS, M. A. *et al.* Phylogenomic analysis of the species of the Mycobacterium tuberculosis complex demonstrates that Mycobacterium africanum, Mycobacterium bovis, Mycobacterium caprae, Mycobacterium microti and Mycobacterium pinnipedii are later heterotypic synonyms of Mycobacterium tuberculosis. **International Journal of Systematic and Evolutionary Microbiology**, [*s. l.*], v. 68, n. 1, p. 324–332, 2018.

ROBISHAW, J. D. *et al.* Genomic surveillance to combat COVID-19: challenges and opportunities. **The Lancet Microbe**, [*s. l.*], v. 2, n. 9, p. e481–e484, 2021.

SATAPATHY, P. *et al.* Emerging bedaquiline resistance: A threat to the global fight against drug-resistant tuberculosis. **Journal of Biosafety and Biosecurity**, [*s. l.*], v. 6, n. 1, p. 13–15, 2024.

SESPA. **BOLETIM EPIDEMIOLÓGICO DA TUBERCULOSE - Nº 01**. Belém: SECRETARIA DE ESTADO DE SAÚDE PÚBLICA, 2021.

SMITH, T.; WOLFF, K. A.; NGUYEN, L. Molecular Biology of Drug Resistance in Mycobacterium tuberculosis. *In*: PIETERS, J.; MCKINNEY, J. D. (org.). **Pathogenesis of Mycobacterium tuberculosis and its Interaction with the Host Organism**. Berlin, Heidelberg: Springer, 2013. (Current Topics in Microbiology and Immunology). p. 53–80. Disponível em: https://doi.org/10.1007/82_2012_279. Acesso em: 5 jan. 2023.

SNOBRE, J. *et al.* Frameshift mutations in the mmpR5 gene can have a bedaquiline-susceptible phenotype by retaining a protein structure and function similar to wild-type Mycobacterium tuberculosis. **Antimicrobial Agents and Chemotherapy**, [*s. l.*], v. 68, n. 12, p. e00854-24, 2024.

SONNENKALB, L. *et al.* Bedaquiline and clofazimine resistance in Mycobacterium tuberculosis: an in-vitro and in-silico data analysis. **The Lancet Microbe**, [*s. l.*], v. 4, n. 5, p.

e358–e368, 2023.

TORTORA, G. J.; CASE, C. L.; FUNKE, B. R. **Microbiologia - 12ª Edição**. [*S. l.*]: Artmed Editora, 2016. Disponível em: https://books.google.com.br/books?id=L98_DQAAQBAJ.

TU, P. H. T. *et al.* Bedaquiline resistance probability to guide treatment decision making for rifampicin-resistant tuberculosis: insights from a qualitative study. **BMC infectious diseases**, [*s. l.*], v. 22, n. 1, p. 876, 2022.

VAN RIE, A. *et al.* Sequencing Mycobacteria and Algorithm-determined Resistant Tuberculosis Treatment (SMARTT): a study protocol for a phase IV pragmatic randomized controlled patient management strategy trial. **Trials**, [*s. l.*], v. 23, n. 1, p. 864, 2022.

YILMAZ, N.; UÇAR, E. Y.; SAĞLAM, L. Mycobacterium Tuberculosis and Nontuberculous Mycobacteria Coinfection of the Lungs. **Turkish Thoracic Journal**, [*s. l.*], v. 18, n. 1, p. 23–26, 2017.

ANEXO A: CERTIFICADOS DAS ATIVIDADES MENCIONADAS NO TEXTO



VI SNNB 23

2023 · SALINÓPOLIS · PARÁ

VI SIMPÓSIO NORTE-NORDESTE DE
BIOINFORMÁTICA

Genômica Populacional e
Biodiversidade Amazônica

CERTIFICADO

Certificamos que

DAVI JOSUÉ MARCON

Recebeu **Menção de Honra** pela apresentação do trabalho

Validation of the Use of High-Performance Computing at the Evandro Chagas Institute for Genomic Surveillance of Mycobacterioses

com os coautores

Davi Josué Marcon, Reginaldo Sales da Silva Junior, Jedson Ferreira Cardoso, Marilia Lima da Conceição, Luana Neponucemo Gondim Costa Lima, Edivaldo Costa Sousa Junior, Lourdes Maria Garcez, Abhinav Sharma, Emilyn Costa Conceição, Karla Valéria Batista Lima

na modalidade de pôster **no VI Simpósio Norte-Nordeste de Bioinformática,**

realizado no dia 30, 31 de outubro e 1 de novembro de 2023, em Salinópolis-PA.

Gilderlanio Santana de Araújo
Coordenação Geral do VI SNNB

Ândrea Kely Ribeiro dos Santos
Coordenação Geral do VI SNNB

INFECTONORTE/NORDESTE
IX CONGRESSO NORTE/NORDESTE
DE INFECTOLOGIA

14 A 16 NOVEMBRO/2024
GRAND MERCURE HOTEL
BELÉM, PA

CERTIFICADO

A Comissão Avaliadora de Trabalhos tem a enorme satisfação de atribuir **MENÇÃO HONROSA** ao trabalho intitulado **SIMULAÇÃO COMPUTACIONAL DE VARIANTES RELACIONADAS À RESISTÊNCIA À BEDAQUILINA EM MYCOBACTERIUM TUBERCULOSIS,** dos autores *Danna Karen Corrêa dos Santos; Davi Josué Marcon; José Anderson Monteiro Pantoja e Karla Valéria Batista Lima,* do Instituto Evandro Chagas, pelo mérito de ter sido um dos 06 finalistas do **IX Congresso Norte/Nordeste de Infectologia**, realizado no período de 14 a 16 de novembro, no Hotel Grand Mercure, na cidade de Belém/PA.

Belém, PA, 16 de novembro de 2024.

Sociedade Brasileira de Infectologia

Alberto Chebabo
Presidente da SBI

Alexandre Naime Barbosa
Presidente da Comissão Científica

Alessandre Guimarães
Presidente do Congresso e da Sociedade Paraense de Infectologia

MEDTROP
59º Congresso da SBMT • São Paulo | SP
Medicina Tropical sob olhar de Saúde Única

22 a 25
setembro 2024

Certificamos que o trabalho intitulado

Vigilância Genômica Revela Perfil de Transmissão dos Primeiros Casos de Tuberculose Extensivamente Resistente no Estado do Pará

de autoria de **Davi Josué Marcon, Abhinav Sharma, Alex Brito Souza, Rafaella Bonfim Barros, Valnete das Graças Dantas Andrade, Lúcia Helena Martins Tavares Monteiro, Luana Nepomuceno Gondim Lima, Ricardo José de Paula Souza e Guimarães, Emilyn Costa Conceição , Carlos Augusto Abreu Alberio, Karla Valéria Batista Lima** foi apresentado por Davi Josué Marcon no **59º MEDTROP - Congresso da Sociedade Brasileira de Medicina Tropical**, realizado no Centro de Convenções Rebouças, de 22 a 25 de setembro de 2024, na forma de apresentação **E-pôster**.

São Paulo, 25 de setembro de 2024.

Para verificar a autenticidade deste certificado acesse: validacertificados.iweventos.com.br e use o código: 28c2ace93b

Prof. Dr. Marcus Vinicius Lacerda
Coordenador da Comissão Científica

Profa. Dra. Hiro Goto
Presidente do MEDTROP 2024

Rosália Morais Torres
Presidente SBMT

CERTIFICADO

Reuniões Satélites

CHAGASLEISH Reunião Anual de Pesquisa Aplicada em Doença de Chagas e Leishmanioses 2024 Alvos e Indicadores de Uma Só Saúde

Entomol10

Fórum Social Brasileiro de Enfrentamento das Doenças Infecciosas e Negligenciadas

XI Workshop Nacional da Rede-TB 22 a 23 de setembro de 2024

Realização

SBMT Sociedade Brasileira de Medicina Tropical

MEDTROP
59º Congresso da SBMT • São Paulo | SP
Medicina Tropical sob olhar de Saúde Única

22 a 25
setembro 2024

CERTIFICADO

Certificamos que o trabalho intitulado

Caso complexo de coinfecção TB-MNT em referência no estado do Pará

de autoria de **Davi Josué Marcon, Danna Karen Corrêa dos Santos , Alex Brito Souza, Jeanne Gonçalves Cabral, Abhinav Sharma, Rafella Bonfim Barros, Valnete das Graças Dantas Andrade, Lúcia Helena Martins Tavares Monteiro, Carlos Augusto Abreu Albério , Karla Valéria Batista Lima, Emilyn Costa Conceição** foi apresentado por Davi Josué Marcon no **59º MEDTROP - Congresso da Sociedade Brasileira de Medicina Tropical**, realizado no Centro de Convenções Rebouças, de 22 a 25 de setembro de 2024, na forma de apresentação **E-pôster**.

São Paulo, 25 de setembro de 2024.

Para verificar a autenticidade deste certificado acesse: validacertificados.iweventos.com.br e use o código: acc769b438

Prof. Dr. Marcus Vinicius Lacerda
Coordenador da Comissão Científica

Profa. Dra. Hiro Goto
Presidente do MEDTROP 2024

Rosália Morais Torres
Presidente SBMT

Reuniões Satélites

CHAGASLEISH 2024 – Reunião Anual de Pesquisa Aplicada em Doença de Chagas e Leishmanioses – Alvos e Indicadores de Uma Só Saúde

Entomol10 – X Workshop de Genética e Biologia Molecular de Insetos Vetores de Doenças Tropicais

Fórum Social Brasileiro de Enfrentamento das Doenças Infecciosas e Negligenciadas

XI Workshop Nacional da Rede-TB – 21 a 23 de setembro de 2024

Realização

SBMT – Sociedade Brasileira de Medicina Tropical

CERTIFICADO

Certificamos que **Davi Marcon** foi instrutor(a) do 1° Minicurso de Introdução à Bioinformática intitulado "1° Minicurso de Introdução à Bioinformática", realizado pela Seção de Bacteriologia e Micologia - SABAC do Instituto Evandro Chagas/SVSA/MS, de 12 a 28/04/2023, com carga horária de 12 horas.

Ananindeua-PA, 28 de abril de 2023.

FERNANDA DO ESPÍRITO SANTO SAGICA
Chefe da Seção de Ensino, Informação Científica e Memória – SIEIC/IEC/SVSA/MS

LÍVIA CARÍCIO MARTINS
Diretora do Instituto Evandro Chagas IEC/SVSA/MS

INSTITUTO EVANDRO CHAGAS
SECRETARIA DE VIGILÂNCIA EM SAÚDE E AMBIENTE

SUS

MINISTÉRIO DA SAÚDE

GOVERNO FEDERAL
BRASIL
UNIÃO E RECONSTRUÇÃO

Conteúdo:	Carga Horária:
Conteúdo programático: - Conceitos básicos de biologia molecular - Conceitos básicos de informática - Ômicas - Sequenciadores e seus dados - Introdução ao Linux e ao Conda - Sequenciamento por terminação de cadeia e sua análise - Montagem de genomas NGS - Tratamento e finalização da montagem de genomas - Processamento de dados brutos por web-service - Análise funcional de dados genômicos	12 horas