

Development and Validation of an ISO-Accreditable Genomic Pipeline for Predicting Antimicrobial Resistance in *Mycobacterium tuberculosis*

David Osei^{1*}, Akua Afriyie¹, Kofi Adu²

¹ Department of Genomics and Infectious Disease, Faculty of Medicine, University of Ghana, Accra, Ghana.

² Department of Antimicrobial Resistance and Bioinformatics, Faculty of Medicine, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.

*E-mail ✉ david.osei@ug.edu.gh

Abstract

Whole-genome sequencing is assuming an increasingly important role in guiding the clinical care of patients with tuberculosis. While bioinformatics resources for processing and clinically interpreting *Mycobacterium tuberculosis* genomic data are advancing, a persistent gap remains for user-friendly, adaptable tools that can be readily customized for local clinical reporting requirements and meet the rigorous demands of international laboratory accreditation. This study aimed to develop a dependable software application to determine *M. tuberculosis* lineages and antimicrobial resistance patterns directly from genomic sequences, engineered for clinical use and accessible to routine microbiology laboratories. We created *tbAMR*, a versatile and comprehensive data-driven platform for analyzing *M. tuberculosis* genomic data, encompassing both lineage identification and phenotypic drug-susceptibility prediction. The tool processes short-read sequencing data in FASTQ format or annotated VCF files derived from short- or long-read technologies. It systematically detects genomic variants—including single-nucleotide polymorphisms, insertions, deletions, large structural rearrangements, and gene loss or inactivation—and cross-references resistance-conferring mutations against the WHO catalog (or a custom database). It generates a structured report that classifies resistance profiles for clinical use. We validated it through extensive benchmarking against phenotypic and genomic results from our in-house collection ($n = 2005$) and large external repositories and published datasets ($n = 13\,777$). We also tested simulated sequence data with deliberately introduced known variants to establish optimal quality thresholds and confirm the pipeline's robustness for diagnostic deployment. We compared performance against established public tools, notably TBProfiler and Mykrobe. *tbAMR* delivered precise lineage assignments and reliable predictions of phenotypic susceptibility for first-line agents (sensitivity 94.6% [95% CI: 94.2–95.0], specificity 97.5% [97.3–97.7]) and second-line agents (sensitivity 83.7% [82.7–84.7], specificity 98.0% [97.9–98.1]). Its analytical speed and overall accuracy matched or approached those of other widely used bioinformatics platforms, such as TBProfiler (first-line sensitivity 94.2% [93.0–95.3], specificity 97.9% [97.6–98.2]) and Mykrobe (first-line sensitivity 91.5% [90.0–92.8], specificity 98.4% [98.2–98.6]). The software further provides configurable parameters that enable laboratories to adjust reporting thresholds according to their individual clinical contexts. *tbAMR* represents a practical solution for clinical and public health microbiology laboratories, offering straightforward customization by laboratory staff without specialized programming skills. The tool has successfully achieved ISO accreditation within our setting. It is currently integrated into routine workflows for genomic antimicrobial resistance reporting, delivering results in a timeframe comparable to conventional phenotypic testing (typically 3–4 weeks from positive culture). Comprehensive supporting materials—including reporting templates, detailed validation frameworks, and reference datasets—are provided to help other laboratories implement and accredit this important diagnostic approach, ultimately supporting enhanced tuberculosis control worldwide.

Keywords: *Mycobacterium tuberculosis*, Antimicrobial resistance, Whole-genome sequencing, Genomic drug susceptibility testing, Bioinformatics

Access this article online

<https://smerpub.com/>

Received: 17 May 2024; Accepted: 15 June 2024

Copyright CC BY-NC-SA 4.0

How to cite this article: Osei D, Afriyie A, Adu K. Development and Validation of an ISO-Accreditable Genomic Pipeline for Predicting Antimicrobial Resistance in *Mycobacterium tuberculosis*. J Med Sci Interdiscip Res. 2024;4(1):155-70. <https://doi.org/10.51847/azg1m67Gae>

Introduction

Mycobacterium tuberculosis represents the primary aetiological agent of tuberculosis, an illness with worldwide distribution that disproportionately impacts individuals residing in low- and middle-income settings.

It constitutes the foremost infectious cause of death globally [1].

Research in context

Evidence before this study

We performed a literature search on PubMed utilizing the keywords “*Mycobacterium tuberculosis*”, “clinical”, “bioinformatics”, “genomics”, and “drug resistance (OR antimicrobial)”. The search encompassed all publications released before July 31, 2024, without imposing language limitations (n=258). Studies selected for review applied genomic methods to forecast probable drug resistance profiles in *M. tuberculosis*. While many investigations emphasize the complexities of translating genomic information into clinically relevant AMR predictions, few address the operational challenges of deploying a fully validated and accredited WGS-based bioinformatics pipeline for routine diagnostic use. Sustainable frameworks addressing these implementation barriers are notably absent from the existing literature.

Added value of this study

This investigation is, to the best of our knowledge, the first effort to present detailed, actionable approaches for validating and accrediting routine AMR prediction from WGS data to guide clinical treatment decisions for *M. tuberculosis*. We engineered, thoroughly evaluated, and secured accreditation for tbAMR, a resilient and adaptable bioinformatics platform grounded in data-driven resistance inference. The tool supports effortless parameter modification and was refined through analysis of a substantial local dataset, complemented by external

public resources. Additionally, we released the entire analytical pipeline, along with supporting materials, for open use by the international public health community to facilitate analogous initiatives elsewhere. Input from infectious disease specialists, clinical microbiologists, and epidemiologists informed the design of a comprehensive end-to-end process that integrates sequencing, computational analysis, clinical interpretation, report generation, discrepancy management, and sustained quality monitoring.

Implications of all the available evidence

The current study fulfills key requirements for the clinical and public health laboratory reporting of genomic AMR information concerning *M. tuberculosis*. It additionally supplies a ready-to-implement pipeline and curated datasets that enable other facilities to establish genomic AMR testing services. By building on and enhancing existing open-source bioinformatics resources, we demonstrate the feasibility of delivering timely genomic AMR reports for *M. tuberculosis* in a low-prevalence, resource-rich context. This contribution is expected to be a foundational step toward wider adoption of genomic AMR reporting practices in public health and reference laboratories worldwide. Management of active tuberculosis demands extended and resource-intensive treatment courses involving multiple antimycobacterial medications for no less than six months [2]. Standard initial therapy consists of four first-line drugs—rifampicin, isoniazid, pyrazinamide, and ethambutol—followed by de-escalation to two agents (rifampicin plus isoniazid in susceptible cases).

Table 1. Antimycobacterial medications utilized for tuberculosis therapy

Drug category	Antimycobacterial agents
First-line therapy	Rifampicin, isoniazid, ethambutol, pyrazinamide
Second-line therapy for multidrug-resistant tuberculosis	
Group A	Fluoroquinolones (moxifloxacin or levofloxacin), bedaquiline, linezolid
Group B	Clofazimine, cycloserine or terizidone
Group C	Delamanid, imipenem-cilastatin or meropenem, ethambutol*, pyrazinamide*, amikacin or streptomycin, ethionamide, para-aminosalicylic acid
Other agents (currently not recommended by WHO for tuberculosis treatment)	Kanamycin, capreomycin, gatifloxacin, clavulanic acid

Data sourced from WHO [2]. *Where susceptibility is confirmed by drug susceptibility testing.

The comparatively slow replication rate of *M. tuberculosis* in culture often compels clinicians to initiate therapy for tuberculosis patients before the availability of

phenotypic drug susceptibility testing (DST) outcomes, which may require as long as two months in numerous laboratory environments [3]. Whole-genome sequencing

(WGS) offers the benefit of delivering a broad spectrum of information from a single *M. tuberculosis* isolate, encompassing species confirmation, lineage determination, and resistance profiling, thereby potentially shortening the interval to actionable DST results for both first-line and additional antimycobacterial agents, particularly when performed directly from mycobacterial growth indicator tube cultures. Furthermore, WGS can enhance the reliability of resistance identification for certain medications, for which phenotypic testing exhibits well-documented reproducibility limitations, notably with ethambutol and pyrazinamide [4]. WGS datasets also support epidemiological investigations, including outbreak detection and contact tracing [5]. Consequently, the systematic application of WGS to *M. tuberculosis* holds considerable promise for enhancing individual patient care within clinical settings and broader public health responses. In recognition of these advantages, recent national guidelines in Australia advocate for the routine implementation and reporting of such findings [6].

Although there is a pressing requirement for timely genomic detection of *M. tuberculosis* antimicrobial resistance (AMR) to guide therapeutic choices, and despite the abundance of available bioinformatics resources, only a minority of clinical and public health microbiology (CPHM) laboratories worldwide have established protocols for the routine reporting of AMR profiles derived from WGS data. Implementing such a program requires careful consideration of multiple factors, especially when results directly influence patient treatment. These include the accuracy, user-friendliness, and adaptability of resistance-prediction tools, as well as compatibility with existing laboratory procedures, bioinformatics workflows, and data management systems, and accreditation by national bodies in accordance with International Organization for Standardization (ISO) requirements (ISO 15189:2022) [7]. Accreditation to ISO standards enables laboratories to verify their capacity to deliver consistent, reproducible, and high-caliber testing services. Other publications examine the accreditation pathway and the particular complexities associated with genomic assays in detail [8].

A range of well-established, high-performing bioinformatics platforms exists for detecting genomic markers of AMR in *M. tuberculosis*, such as Mykrobe [9], TBProfiler [10], and ARIBA [11], with additional tools recently assessed [12]. Moreover, an updated

iteration of the WHO *M. tuberculosis* mutation catalog [13] is now available. This version includes an enlarged set of resistance-associated mutations for second- and third-line agents (**Table 1**); (groups a–c) together with recommendations for interpreting intricate genomic variant combinations. Although this expanded knowledge base is valuable for in-depth analyses, clinical reporting often requires a more straightforward interpretation, as clinicians prefer concise, binary classifications (resistant or susceptible) over extensive explanatory notes. For this reason, end-users may need to adjust interpretation rules or protocols for individual variants. They may also wish to incorporate locally relevant mutations into the resistance database or alter classification terminology (for instance, replacing “Susceptible” with “Wildtype”) to align with institutional preferences. However, most existing tools for *M. tuberculosis* genomic DST (gDST) do not readily permit such modifications without specialist bioinformaticians altering the source code. This constraint limits CPHM laboratories and investigators from tailoring solutions to their specific operational contexts. In addition, high-volume sequencing facilities frequently process samples in large batches using established workflows; therefore, new bioinformatics tools must integrate seamlessly into these pipelines, remain easy to sustain, and allow straightforward reversion following any updates.

In this article, we validate a real-time, routine-use bioinformatics tool designed to infer phenotypic resistance in *M. tuberculosis* by identifying genomic variants in a CPHM laboratory environment. We have since obtained ISO 15189:2022 (Medical Laboratories) accreditation for *tbtAMR*, enabling its use in clinical and public health reporting of genomic AMR for *M. tuberculosis*. The pipeline and associated validation methodology are openly provided to help international laboratories adopt WGS-based genomic AMR reporting.

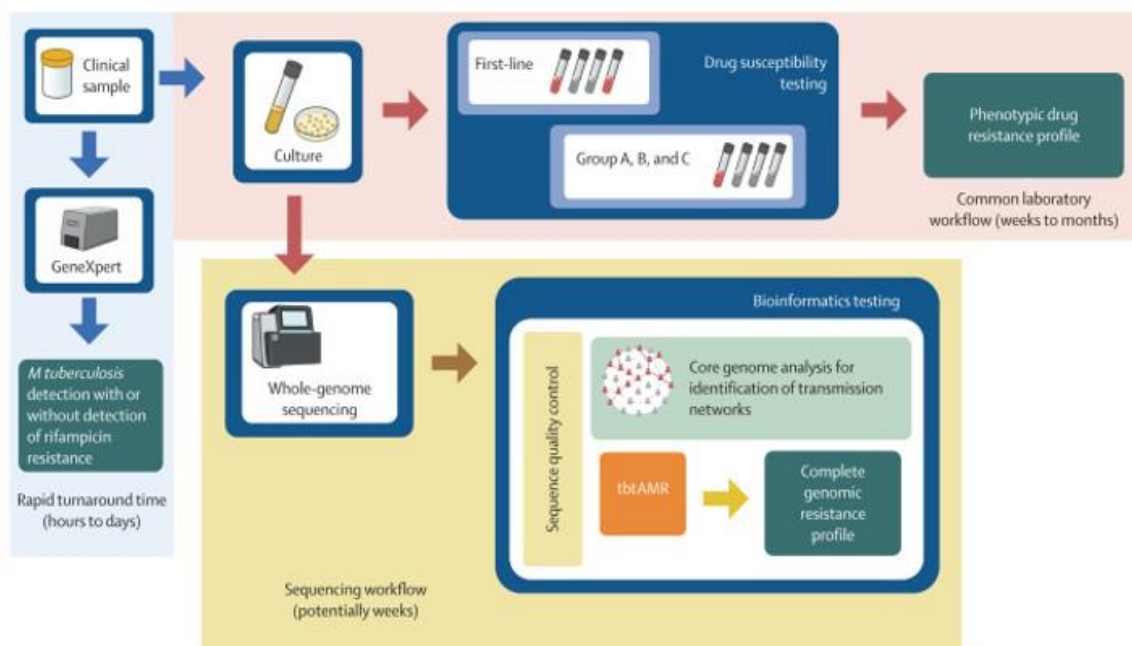
Materials and Methods

tbtAMR design

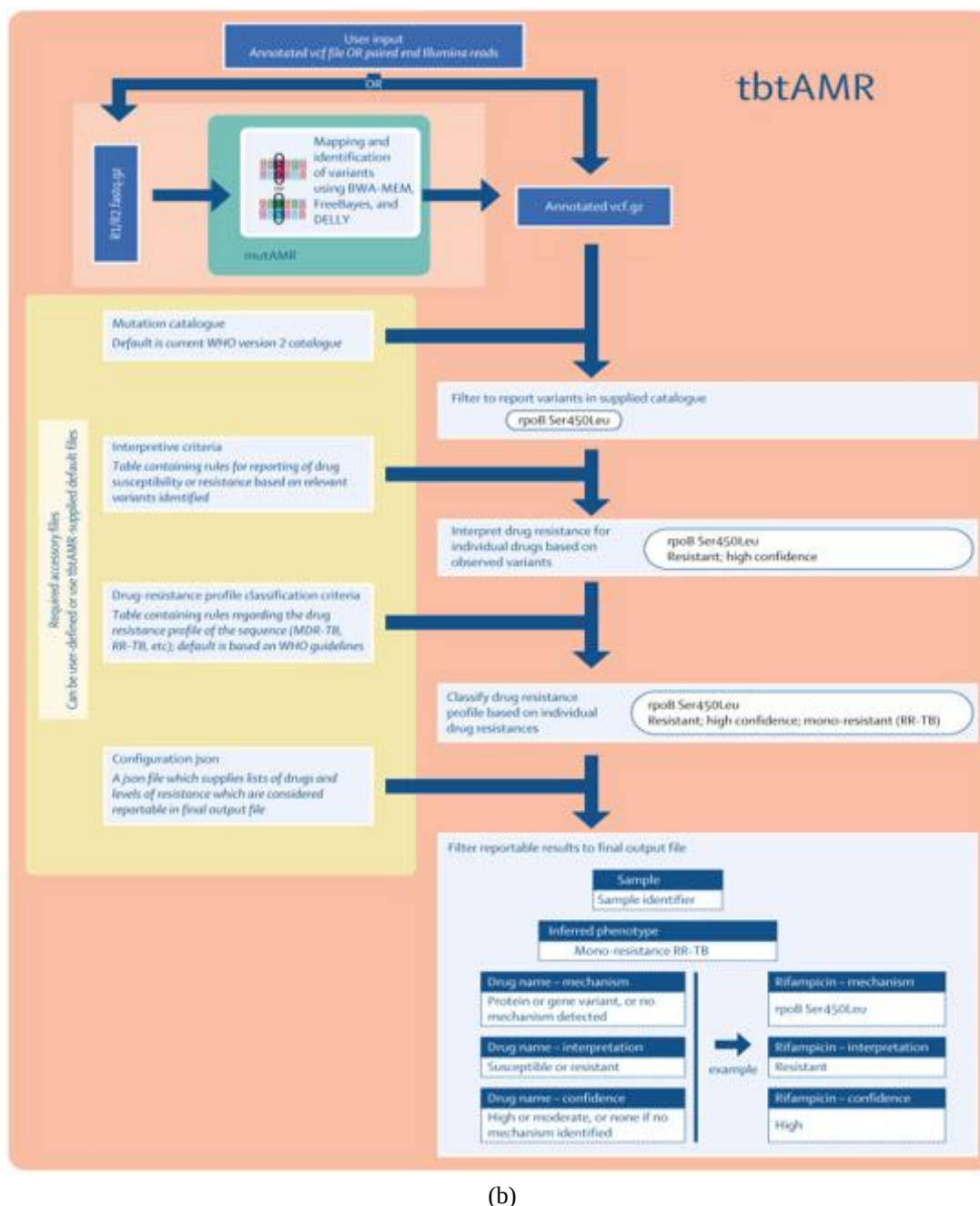
Several software tools are currently available to identify genetic elements associated with antimicrobial resistance (AMR) in *M. tuberculosis* genomic sequences. Prominent examples include Mykrobe [9] and TBProfiler [10], each offering particular strengths and drawbacks. During development of a dedicated standalone

application, priority was given to seamless integration with established laboratory procedures (**Figure 1a**), strong adaptability in output generation, and straightforward ongoing maintenance that staff without specialist coding skills could ideally perform. These considerations address common constraints observed in prior tools when applied in clinical and public health microbiology (CPHM) laboratory environments, where software maintenance and updates often pose significant challenges. Because AMR predictions derived from WGS data for *M. tuberculosis* directly influence therapeutic choices, the results must be appropriate for direct clinical application. This requires an interpretive framework suitable for medical reporting rather than the

dissemination of raw, unprocessed data except where specifically warranted. Accordingly, tbtAMR supplies four essential parameters to guide clinical decisions: (1) detection or absence of mutations, including small insertions or deletions (indels), extensive structural alterations, gene deletions or functional loss, and single nucleotide polymorphisms (SNPs); (2) the degree of confidence associating the mutation with resistance; (3) the intensity of resistance conferred (high-level or low-level); and (4) the comprehensive resistance classification for the isolate, determined by predictions for individual agents (for instance, following WHO categorisation).



(a)



(b)

Figure 1. Integration of tbtAMR into the *Mycobacterium tuberculosis* diagnostic and testing pathway within a reference laboratory.

(a) Standard diagnostic routes for *M. tuberculosis*. GeneXpert enables rapid identification of the organism and potential rifampicin resistance. This closed PCR system targets prevalent resistance mutations and typically returns results within hours to days at diagnostic facilities. Conventional culturing requires inoculation onto solid agar or, more frequently, liquid mycobacteria growth indicator tubes, with positive signals potentially

emerging only after 6–8 weeks. After culture confirmation, phenotypic DST begins and may take up to two weeks, or longer with stepwise testing protocols limited to group A, B, and C agents once initial resistance is established. Consequently, the full process from specimen collection to phenotypic results typically takes 6 weeks to 3 months. Optimized whole-genome sequencing can substantially compress this timeline,

yielding a detailed overview of resistance without additional assays while simultaneously bolstering public health measures such as contact tracing; (b) Overview of the tbAMR analytical workflow. Acceptable inputs comprise annotated variant call format (VCF) files aligned to the H37Rv v3 reference (generated from either short-read or long-read sequencing) or paired-end short-read FASTQ files, rendering the platform compatible with diverse sequencing technologies. The tool uses four supplementary files, with default versions included that users can replace with their own. By default, tbAMR incorporates version 2 of the WHO mutational catalog, supported by files that specify interpretation standards, resistance classification logic, and reporting filters. These supporting files govern variant evaluation, profile categorization, and output refinement for clinical reports. MDR=multidrug resistant. RR=rifampicin resistant. SNP=single nucleotide polymorphism. TB=tuberculosis. Variant nomenclature example: “rpoB_Ser450Leu” signifies substitution of serine by leucine at residue 450 within the RpoB protein sequence.

tbAMR is a Python package that operates independently with few external dependencies. It processes an annotated VCF file to derive an inferred antibiogram and an overall resistance classification based on customizable user criteria (**Figure 1b**). Version 2 of the WHO *M. tuberculosis* mutation catalog (published in 2023) serves as the default reference database within the software [13], with accompanying interpretive guidance shaping classification decisions [14]. Users can introduce a personalized mutation catalog and corresponding rules for variant assessment and profile assignment, improving adaptability and simplifying future updates.

Delivering clinically actionable interpretations of predicted drug resistance constitutes a fundamental objective of tbAMR. Unlike several alternative methods, the tool does not depend on fixed, embedded library components. Rather, it enables selection between standard libraries and user-adjusted versions. Operation of tbAMR is supported by four accessory files (elaborated below and depicted in **Figure 1b**, with availability online [15]): (1) a mutation catalogue or repository of resistance determinants; (2) interpretation guidelines for assigning susceptibility or resistance categories per drug; (3) classification protocols for determining the isolate’s overall resistance status (such as “fully susceptible” or “MDR-TB”); and (4) a configuration document that outlines the catalogue structure and report formatting.

The primary accessory file, the “mutation catalog,” is a CSV document that lists the pertinent drugs, implicated genes, specific variants, and associated confidence ratings for resistance inference. The default version supplied with tbAMR is based on WHO catalog version 2 [13].

Secondly, the “interpretation criteria” CSV file contains decision rules for analyzing data drawn from the mutation catalog once a variant has been identified. The default file supplied with the software is constructed from the supplementary notes provided with the WHO mutation catalog version 2. These rules fall into two categories: broadly applicable “default rules” that cover the majority of mutations, and “override rules” designed for specific exceptions and atypical situations. Default rules establish the standard approach for determining reportable drug resistance in typical scenarios. As an illustration, detection of a variant linked to resistance (or “Associated with resistance—Interim”) for a particular drug triggers a “Resistant” designation for that agent. The software reports results only for drugs with a default rule, allowing users to control the output scope. Override rules add supplementary logic to handle exceptions or special cases and take precedence over the default rules. For instance, identification of certain variants results in a “Low-level resistant” designation, consistent with explanatory notes in the WHO catalog. Users can incorporate their own rules—for example, to include a particular comment under specified conditions or to flag the presence of a variant without assigning any resistance interpretation.

Third, the “classification rules” CSV file defines criteria for assigning an overall resistance profile based on the pattern of detected resistances across the sequence. This includes categories such as mono-resistance to an individual first-line drug or multi-drug resistance. These default criteria follow the prevailing WHO recommendations for resistance profile classification (e.g., “No first-line drug resistance predicted”, “Mono-resistance predicted [HR-TB]”, “Multi-drug resistance predicted [MDR-TB]”).

Finally, the database configuration file (db_config.json) details the organization of the mutation catalog, the selection of drugs to be considered, the assigned confidence and resistance levels, and the structure for cascade reporting (**Figure 1b**).

tbAMR is distributed with a complete set of default accessory files based on the WHO catalog of mutations version 2 [13] and the current WHO guidelines for

resistance profile classification [2]. Users may, however, adapt or substitute these files with custom versions tailored to their specific operational requirements. Because the files are simple CSV documents, individuals without specialist bioinformatics expertise can readily make these adjustments, enhancing adaptability in both interpretation and reporting procedures.

The performance characteristics of *tbtAMR* were assessed using the default database and interpretation criteria; outcomes may differ when using a customized database and/or alternative interpretive rules. The software further offers optional capabilities for paired-end read alignment, variant calling, and annotation. These features depend on additional external tools, namely BWA-MEM [16], FreeBayes [17], SnpEff [18], DELLY [19], BCFtools, and SAMtools [20] (**Figure 1b**). Implementation of these extended functions is handled

by the in-house companion *mutAMR* tool. The full suite is distributed as a Conda package.

Validation dataset

Implementation of genomic drug susceptibility testing for clinical and public health applications, including ISO accreditation, required demonstration that *tbtAMR* could produce dependable outputs for strain lineage, underlying resistance mechanisms, the predicted antibiogram (gDST), and the composite resistance classification. Achieving this objective first required confirming high-confidence variant detection from sequencing reads using *tbtAMR*'s dedicated variant-calling component (*mutAMR*). We then verified that the system could accurately derive lineage information, resistance determinants, and gDST profiles for individual isolates (**Figure 2**). Validation drew on three complementary data sources.

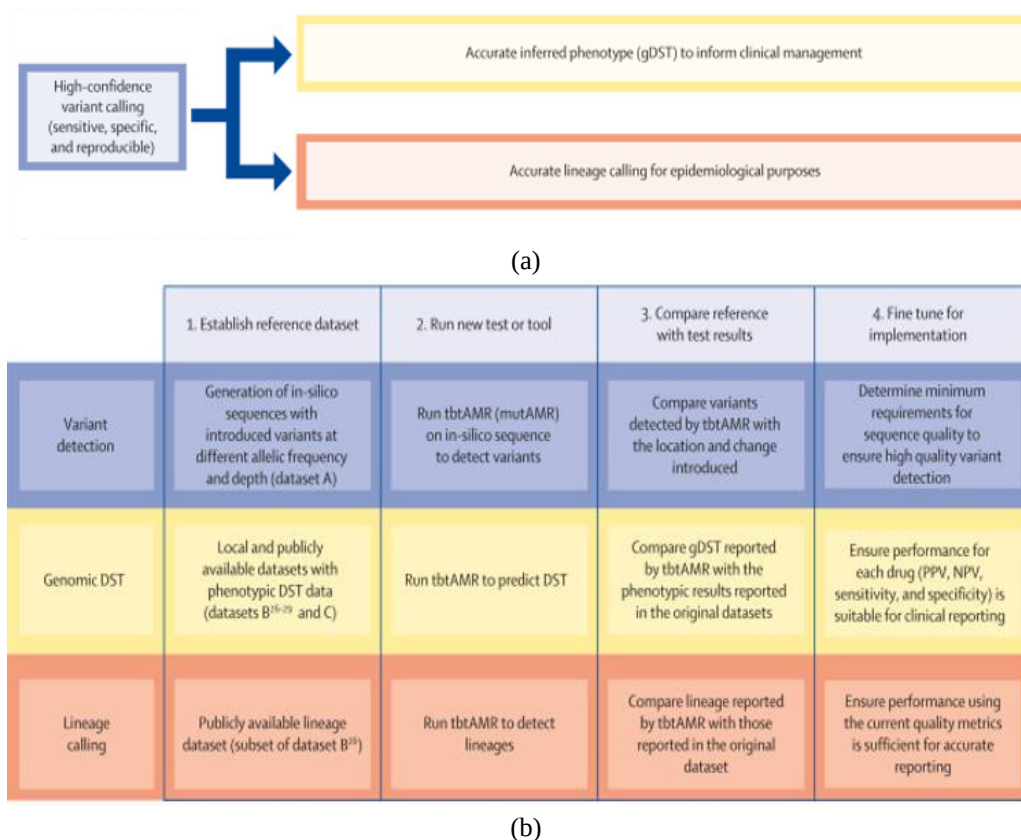


Figure 2. Protocol for validating *tbtAMR* for clinical and public health microbiology applications

(a) Key validation requirements for *tbtAMR* to deliver a trustworthy and clinically robust testing framework. This process centers on confirming the reliability of variant calling, the foundational step for all subsequent analyses,

before applying these calls in lineage assignment and gDST generation; (b) The validation workflow consists of four main phases: (1) Assembly of a high-quality reference dataset to benchmark performance of the novel

bioinformatics solution (tbtAMR). Three distinct datasets were used: artificially generated sequences harboring predefined variants (dataset A), locally produced clinical isolates (dataset c), and publicly sourced clinical isolates (dataset B). (2) Application of tbtAMR to all sequences within each dataset. (3) Systematic comparison of tbtAMR outputs against the reference standards. (4) Optimization for routine deployment, encompassing establishment of minimum quality control metrics for sequencing data and comprehensive performance assessment for every test element (specifically, each antituberculosis drug) to verify clinical utility. DST = drug susceptibility testing. gDST = genomic DST. NPV=negative predictive value. PPV = positive predictive value.

First, we examined the precision of variant detection by tbtAMR using computationally simulated genomic sequences (dataset A). Paired-end reads were produced via TreeToReads software [21-24] based on the H37Rv reference strain (RefSeq accession NC_000962.3), or after deliberate insertion of 2–4 variants per 10 000 bases at specified locations. Error characteristics mirrored those observed on NextSeq500 instruments at the Microbiological Diagnostic Unit Public Health Laboratory (MDU PHL; n=1830). Because resistance profiling must detect emerging resistant subpopulations during therapy, these reads were combined in proportion to create varying allelic fractions (0.01–0.99) across a spectrum of average sequencing depths (20–200×, in 20× increments).

Second, we incorporated 13 777 publicly available genome sequences. These were selected to encompass the breadth of major global lineages and included accompanying phenotypic susceptibility results for first-line drugs, with a portion also covering second-line agents (groups a, b, and c) [25-28] (dataset B provided paired phenotypic DST and genomic data, with a subset supporting lineage validation) [25]. In addition, we analyzed 2005 sequences generated in-house at MDU PHL from Jan 1, 2018, to Dec 31, 2022, with matched phenotypic DST data from the Mycobacterium Reference Laboratory (dataset C). This collection included all culture-positive specimens processed by the reference laboratory during the study interval.

Measuring the accuracy of M. tuberculosis genomic sequence recovery

In the simulated genome set, we evaluated the tool's effectiveness in recovering introduced variants at

minimum coverage thresholds of 10× and 20× (requiring at least 10 or 20 supporting reads per position for a confident base call). Following read alignment to the reference genome, the fidelity of base calling by mutAMR was rigorously examined.

Analysis of inferred phenotypic DST and phylogenetic lineage from genomic data

Evaluation focused on the correspondence between phenotypes predicted from genomic resistance determinants—including single nucleotide polymorphisms, insertions, and deletions—and the outcomes obtained through conventional phenotypic DST. Phenotypic findings that were indeterminate or absent were removed before each analysis. To measure agreement, we directly compared susceptibility profiles inferred from detected genomic variants with experimental phenotypic results and calculated sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV). Identified mismatches between genomic predictions and phenotypic resistance were ideally to be clarified through repeat phenotypic testing. As this step was not feasible for public datasets, these cases remained unresolved, contributing to the observed discrepancies in performance metrics. tbtAMR additionally assigns phylogenetic lineages, covering lineages 1–9, as well as animal-adapted species or subspecies [10]. Agreement between these tbtAMR-derived lineage assignments and the lineage metadata supplied with the public sequences was similarly examined [25].

Report design

Because AMR genotypes and inferred antibiograms for *M. tuberculosis* are relatively novel for most clinicians, we prepared tailored explanatory statements for inclusion in the final reports. These statements addressed key concepts, such as the distinction between confidence grading and resistance intensity in predicted phenotypes. We sought input from clinical and laboratory personnel during the review of prototype reports to verify that the presentation and accompanying explanations were readily comprehensible and precise in the institutional environment. After validation and report refinement, the tbtAMR-based genomic AMR inference process for *M. tuberculosis* was incorporated into standard laboratory genomic analysis pathways. This included revising standard operating procedures, providing targeted staff training, integrating with the laboratory information

management system, and validating end-to-end reporting and feedback processes (for example, protocols for samples that failed quality thresholds).

Implementation and accreditation

A dedicated multidisciplinary expert panel on *M. tuberculosis* AMR was constituted to review and resolve any discordance between genotypic and phenotypic data. Panel membership included representatives from MDU PHL (one medical microbiologist, one genomic scientist, and three bioinformaticians) and the Mycobacterium Reference Laboratory (one medical microbiologist and two scientists experienced in mycobacterial culture and molecular diagnostics); all members are listed as authors on this publication. The panel convenes monthly to evaluate individual patient cases and deliver specialist advice to clinicians regarding any observed inconsistencies. Such regular interaction promotes sustained exchange among laboratory scientists performing wet-bench work, treating clinicians, epidemiologists, medical microbiologists, and computational specialists. It supports ongoing surveillance of clinically relevant AMR markers, detailed case discussions, and forward-looking assessment of emerging or poorly defined resistance mechanisms. Preparatory steps for ISO-equivalent accreditation included establishing formal revalidation and reverification procedures for tbtAMR. The complete *M. tuberculosis* genomic AMR inference workflow was then formally submitted for assessment against the ISO 15189:2022 criteria.

Ethics approval

Approval was obtained from the University of Melbourne Human Research Ethics Committee (study number 1954615). Because the investigation relied exclusively on isolated data, the committee waived the requirement for patient consent.

Role of the funding source

The study sponsors played no part in the study design, data collection or analysis, interpretation of results, or drafting of the manuscript.

Results and Discussion

In variant detection, tbtAMR can align and identify variants from paired-end Illumina whole-genome sequencing data via its mutAMR module. We assessed performance by measuring how effectively mutAMR recovered deliberately introduced single-nucleotide polymorphisms (SNPs) in the simulated sequence set (dataset A) and by identifying the lowest average sequencing depth required for reliable outcomes. Many variant detection pipelines apply a minimum depth threshold of 10× [29, 30], which requires at least 10 aligned reads at a given genomic position before a variant can be confidently called. Nevertheless, analysis across a spectrum of average sequencing depths and allelic fractions revealed a notable number of false-positive SNPs (mean 6.6, 95% CI: 3.9–9.3) when the conventional 10× threshold was employed; such errors could lead to incorrect attribution of resistance mechanisms. Raising the minimum depth requirement for base calling to 20× substantially enhanced accuracy, lowering the mean number of false-positive SNPs to 1.1 (0.7–1.5) while exerting minimal influence on the recovery rate of the introduced SNPs (**Figures 3a and 3b**). Additional assessments examined the combined effects of average genome coverage and allelic frequency on mutAMR's ability to correctly detect mutations in simulated data. At average depths below 40× paired with allelic frequencies under 10%, variant detection sensitivity fell markedly to 2.1% (95% CI: 1.48–2.66), accompanied by a high false discovery rate (FDR) of 41.1% (31.1–51.2).

In contrast, application of a 20-read minimum depth threshold, an average genome depth of at least 40×, and allelic frequencies of 10% or greater yielded optimal results (**Figures 3c and 3d**), achieving an overall sensitivity of 96.3% (95.5–97.1) and a very low FDR of 0.21% (0.11–0.31). Identification of additional variant classes, including deletions, by mutAMR aligned closely with the performance of TBProfiler and Mykrobe when tested on publicly available sequences. Based on these findings, we established the minimum acceptable criteria for detecting resistance-associated variants as an average read depth of 40× and a minimum allele frequency of 10%.

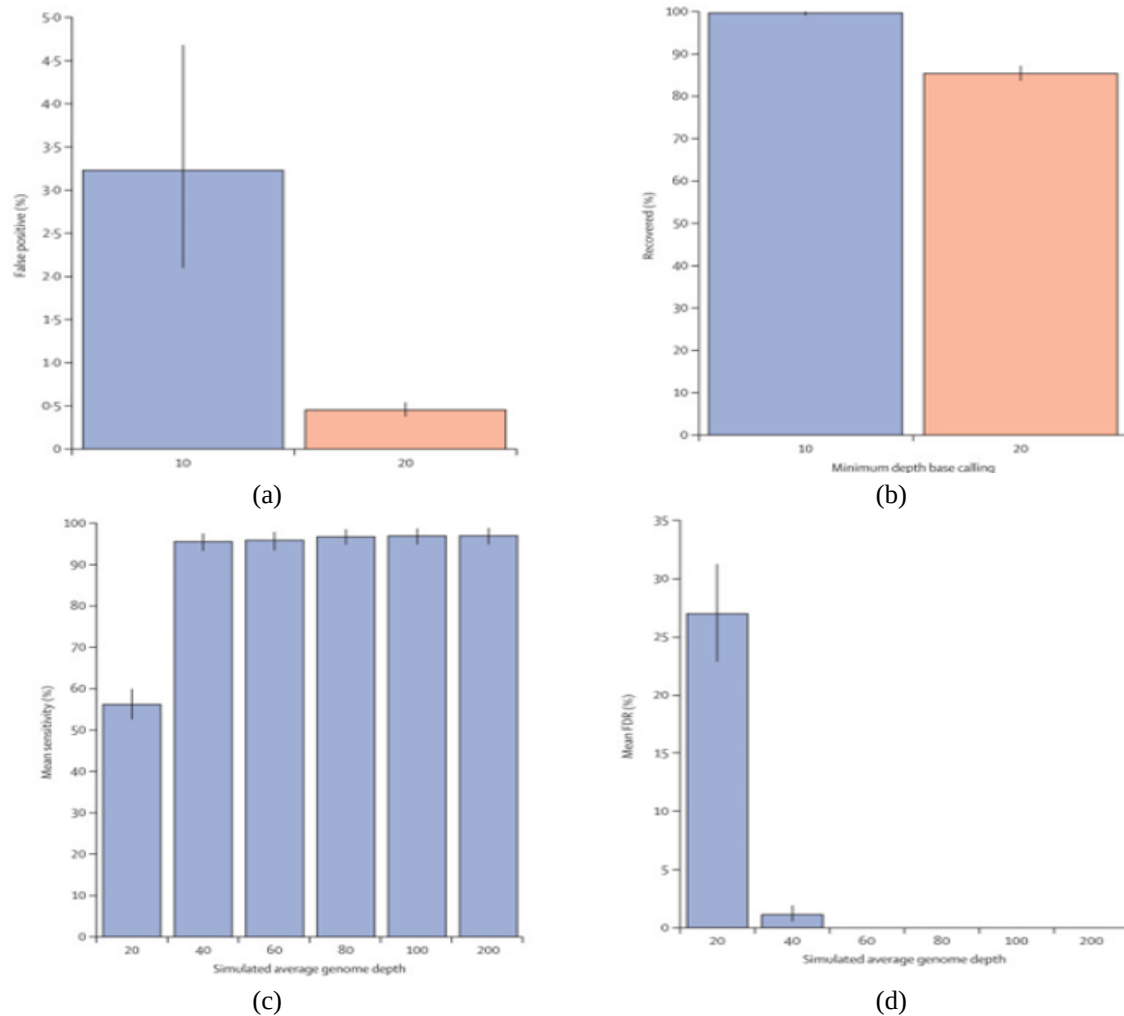


Figure 3. Performance of mutAMR in recovering SNPs from *Mycobacterium tuberculosis* genome sequences

Performance assessment of mutAMR within tbAMR for accurate mutation recovery in *M. tuberculosis* sequences involved analysis of simulated genomes spanning average depths of 10× to 200× and allelic frequencies from 10% to 100%. We investigated the effects of different minimum read depth thresholds for base calling at individual genomic sites (10× versus 20×) in terms of the rate of false-positive SNP detections (a) and the success rate in recovering the introduced SNPs (b). With a 20× minimum depth requirement for base calling, both SNP detection sensitivity (c) and false discovery rate (FDR) (d) were evaluated at average genome coverage levels of $\geq 40\times$ combined with allelic frequencies of $\geq 10\%$ (error bars denote 95% CIs). FDR=false discovery rate. SNP=single-nucleotide polymorphism.

For phylogenetic lineage determination in *M. tuberculosis*, tbAMR achieved excellent agreement with

the lineage designations reported in the public datasets (subset of dataset B; 98.5% concordance [871/884]) [25]. The 13 cases showing discordance primarily consisted of 12 instances in which tbAMR identified two separate phylogenetic lineages within a single sequence. In all such cases, the publicly reported lineage matched one of those detected by the tool. These observations may indicate mixed populations or insufficient coverage at critical sites, distinguishing lineages from the reference (lineage 4) and allowing detection of the minor variant. It is noteworthy that the identification of mixed lineages does not compromise the sequence's utility for AMR variant detection.

In evaluating tbAMR's suitability for *M. tuberculosis* AMR prediction and reporting, the study draws upon extensive prior literature examining genomic inference of phenotypic resistance through diverse analytical

platforms [9, 27, 31, 32]. Rather than replicating these comprehensive earlier efforts, the current analysis specifically examines tbAMR's capacity to deliver reliable AMR inference and reporting suitable for clinical and public health applications.

AMR validation utilized a combined set of 15 323 sequences sourced from public repositories [26-28, 33] and locally generated data produced under ISO-accredited sequencing protocols. After applying the predefined quality control standards and detection limits, we excluded 608 sequences due to insufficient quality. This process retained 14 715 high-quality sequences for assessing tbAMR in predicting phenotypic AMR (datasets b and c).

The tool demonstrated robust performance in inferring phenotypic resistance across all first-line antimycobacterial drugs (Table 2, Figure 4) (overall

sensitivity 94.6% [95% CI 94.2–95.0], specificity 97.5% [97.3–97.7]) and maintained high accuracy for second-line agents (sensitivity 83.7% [82.7–84.7], specificity 98.0% [97.9–98.1]). Benchmarking against established tools such as TBProfiler and Mykrobe confirmed comparable levels of accuracy and computational efficiency (TBProfiler first-line sensitivity 94.2% [93.0–95.3], specificity 97.9% [97.6–98.2]; Mykrobe first-line sensitivity 91.5% [90.0–92.8], specificity 98.4% [98.2–98.6]). The comparatively lower positive predictive value for ethambutol (81.0%, 95% CI 79.3–82.6) stemmed from variants at codon 360 in embB associated with susceptible phenotypes. Prediction of pyrazinamide resistance showed modestly reduced sensitivity (80.0%, 75.4–84.0) compared with other first-line drugs, consistent with findings from alternative pipelines.

Table 2. Performance of the tbAMR pipeline for predicting phenotypic susceptibility to first-line and second-line antimycobacterial drugs.

Antimycobacterial drug	Sensitivity, % (95% CI)	Specificity, % (95% CI)	Positive predictive value (PPV), % (95% CI)	Negative predictive value (NPV), % (95% CI)
First-line medications				
Rifampicin (n = 12,505)	98.1 (97.6–98.4)	98.1 (97.7–98.3)	96.3 (95.7–96.8)	99.0 (98.8–99.2)
Isoniazid (n = 12,545)	94.2 (93.5–94.9)	99.1 (98.9–99.3)	98.5 (98.1–98.8)	96.5 (96.0–96.9)
Ethambutol (n = 10,348)	90.6 (89.2–91.8)	95.0 (94.5–95.5)	81.0 (79.3–82.6)	97.7 (97.4–98.0)
Pyrazinamide (n = 3,174)	80.8 (76.4–84.5)	98.8 (98.3–99.1)	89.2 (85.4–92.2)	97.6 (96.9–98.1)
Second-line medications				
Bedaquiline (n = 9,339)	43.2 (33.3–53.6)	98.2 (97.9–98.5)	18.5 (13.8–24.4)	99.5 (99.3–99.6)
Linezolid (n = 8,017)	40.9 (31.2–51.4)	99.9 (99.8–100.0)	83.7 (70.0–91.9)	99.3 (99.1–99.5)
Moxifloxacin (n = 7,349)	93.3 (91.5–94.7)	97.3 (96.9–97.7)	84.1 (81.8–86.1)	98.9 (98.7–99.2)
Clofazimine (n = 8,591)	15.5 (10.3–22.7)	98.4 (98.1–98.7)	13.1 (8.6–19.3)	98.7 (98.4–98.9)
Amikacin (n = 10,054)	88.0 (85.6–90.0)	98.4 (98.2–98.7)	83.5 (80.9–85.8)	98.9 (98.7–99.1)
Delamanid (n = 8,882)	17.0 (10.8–25.9)	99.9 (99.8–100.0)	69.6 (49.1–84.4)	99.1 (98.9–99.3)
Ethionamide (n = 9,112)	86.9 (85.1–88.5)	94.5 (93.9–95.0)	74.9 (72.8–76.9)	97.4 (97.1–97.8)
Streptomycin (n = 123)	95.1 (88.0–98.1)	78.6 (64.1–88.3)	89.5 (81.3–94.4)	89.2 (75.3–95.7)
Capreomycin (n = 304)	92.7 (86.3–96.3)	98.5 (95.6–99.5)	97.1 (91.9–99.0)	96.0 (92.3–98.0)
Kanamycin (n = 10,397)	86.6 (84.4–88.5)	97.2 (96.8–97.5)	77.4 (74.9–79.7)	98.5 (98.2–98.7)

PPV = positive predictive value; NPV = negative predictive value.

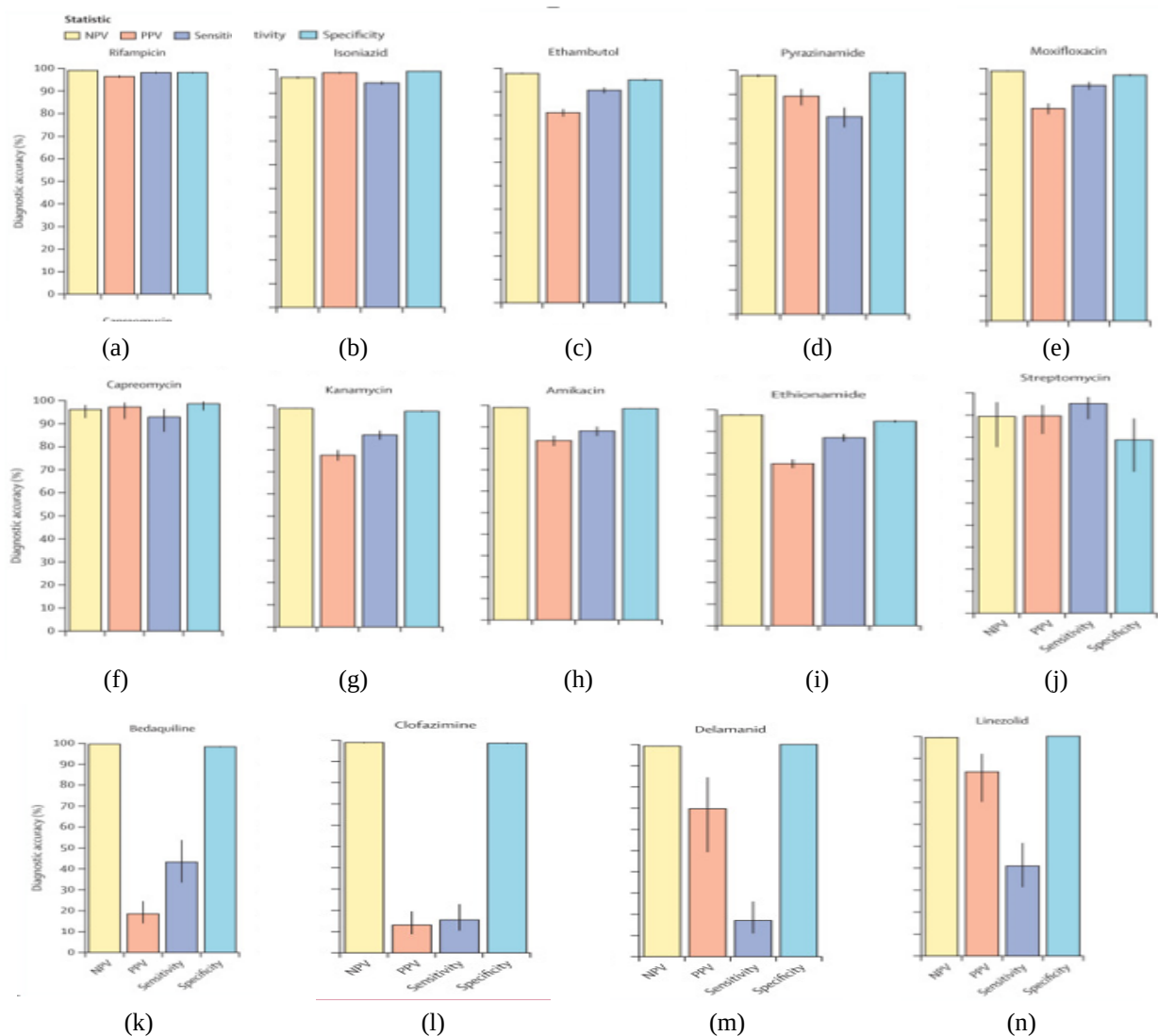


Figure 4. Performance of tbAMR in the prediction of phenotypic susceptibility.

The phenotypes inferred by the tbAMR pipeline were evaluated against experimental drug-susceptibility phenotypic testing outcomes. True positives were cases in which a mutation was detected, and the corresponding phenotype indicated resistance. True negatives corresponded to the absence of any mutation when the isolate was phenotypically susceptible. False positives were noted when a mutation was identified despite phenotypic susceptibility, while false negatives occurred when no mutation was reported in a resistant isolate. We determined sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) for each drug and presented them as percentages (error bars show 95% CIs) (PPV = positive predictive value, NPV = negative predictive value).

For second-line agents belonging to groups A, B, and C (Table 1), tbAMR delivered predictive performance that aligned closely with findings from earlier research [9, 27, 31, 32], including expected levels of false-negative and false-positive classifications of phenotypic resistance (Table 2). Although the tool's results were generally consistent with those obtained using alternative bioinformatics platforms, and adoption of the revised catalog enhanced outcomes compared with the original WHO version 1 catalog, the reliability of resistance predictions for bedaquiline, delamanid, linezolid, and clofazimine proved inadequate for standard clinical reporting in a public health microbiology context. This was primarily attributable to unacceptably high rates of both false-positive and false-negative results for these

essential agents used to manage drug-resistant tuberculosis.

Upon validation, *tbtAMR* was successfully incorporated into the laboratory's routine genomic processing pipeline and attained accreditation in accordance with ISO 15189:2022.

We constructed an integrated pipeline for comprehensive analysis and structured reporting of *M. tuberculosis* antimicrobial resistance, intended for application in clinical public health microbiology (CPHM) environments. The workflow features highly adaptable interpretive rules and classification frameworks, enabling optimization of positive and negative predictive values when forecasting resistance profiles in this pathogen. The overall accuracy of resistance predictions generated by this approach aligns closely with that observed for other widely adopted analytical platforms [9, 32, 33]. Following rigorous validation, the pipeline has been fully operationalized and accredited to ISO standards in our facility, supported by defined reverification protocols and established reporting systems to facilitate sustained quality assurance.

The global expansion of whole-genome sequencing (WGS) capabilities within CPHM laboratories has accelerated notably since the COVID-19 pandemic, leading to substantial growth in both sequencing infrastructure and generated data volume [34]. Nevertheless, important deficiencies in specialized skills remain prevalent within the CPHM workforce [32, 35, 36], particularly in low- and middle-income countries (LMICs). Such limitations restrict the capacity to implement, customize, and sustainably manage open-source bioinformatics resources. Proprietary solutions, while available, frequently involve high financial costs. Although open-source alternatives are freely accessible, their practical deployment in regulated diagnostic laboratory settings often encounters significant obstacles. These tools are typically developed and maintained by independent academic groups, which can result in irregular maintenance, complex installations, or incomplete alignment with the specific demands of clinical public health microbiology services [37]. Achieving and maintaining compliance within an ISO-accredited laboratory requires extensive staff commitment and institutional investment [8], as external updates cannot be selectively adopted. These issues are especially acute in resource-constrained environments [38]. By supplying detailed validation methodologies, reference datasets, reverification frameworks, and

illustrative report examples, *tbtAMR* and its supporting materials aim to alleviate several of these practical challenges for CPHM laboratories. Moreover, its modest set of external dependencies facilitates straightforward deployment on individual laptops or seamless embedding within larger, more sophisticated bioinformatics systems. Flexibility is a core principle of *tbtAMR*'s architecture. Users may use the provided default parameters and files, or substitute them with institution-specific criteria to accommodate local requirements. The accessory files were intentionally developed to allow straightforward editing by non-programmers, including medical microbiologists and scientists with genomic expertise. This design supports context-appropriate adjustments to interpretation thresholds. For instance, laboratories serving regions with elevated tuberculosis prevalence may adopt more sensitive criteria for resistance reporting than those in low-prevalence areas, potentially classifying a broader range of low-level resistance variants as clinically significant. Cascade reporting sequences can likewise be reconfigured to correspond with locally available therapeutic options. The authors maintain that enabling end-users to calibrate these settings directly offers a more suitable solution than depending on external developers' determinations of universal best practice.

Antimicrobial resistance determinants are frequently categorized in a binary manner as either present or absent, leading to the inference of resistant or susceptible phenotypes, respectively. However, this simplification does not always capture the nuanced relationships between genotype and phenotype. Therefore, only mutations with robust supporting evidence should be used to predict resistance. The updated WHO catalog and associated guidance offer valuable insights into variants associated with borderline resistance, such as the *rpoB* His445Asn mutation [13], along with confidence tiers that help interpret results. The modular architecture of *tbtAMR* enables the generation of clinically informative reports focused on high-confidence mutations while simultaneously permitting ongoing surveillance and prospective validation of additional variants across all confidence categories. This dual capability is especially pertinent for newer second-line agents, including bedaquiline, delamanid, and linezolid, where limited data—both locally and internationally—currently constrain reliable resistance prediction for routine clinical public health microbiology (CPHM) reporting. Even for extensively studied drugs, inconsistencies may

arise. For instance, known issues with the reproducibility of phenotypic DST for ethambutol and pyrazinamide can influence gDST performance, as observed in the present study and prior reports [4, 9, 27, 31, 32]. Continued research and iterative validation are expected to further refine the performance of tbAMR and comparable tools within the *M. tuberculosis* genomics community, ultimately improving clinical outcomes.

The application of genomic DST for *M. tuberculosis* in CPHM laboratories involves careful trade-offs between immediate benefits for individual patient care and the practical constraints of limited laboratory resources. Cost and operational efficiency considerations typically govern sequencing protocols in such settings. Consequently, routine achievement of the high read depths necessary for confident detection of variants present at allelic frequencies below 10% is rarely feasible [39]. Moreover, reduced coverage increases the risk of false-positive variant calls (**Figure 2**) [40], potentially resulting in inaccurate gDST reporting. The default parameters incorporated in tbAMR were selected to balance these factors appropriately for CPHM laboratories, whether in high-income or low- and middle-income country (LMIC) contexts. Nevertheless, these parameters remain fully adjustable to accommodate varying resource availability and operational priorities, including those encountered in research environments.

The present evaluation and the performance of tbAMR are inherently constrained by current global understanding of resistance mechanisms and their phenotypic manifestations in *M. tuberculosis*. As a result, perfect genotype–phenotype concordance is not attainable with this or any other existing tool. When used with standard short-read sequencing depths typical of diagnostic and public health laboratories (as opposed to ultra-deep sequencing), tbAMR, like similar platforms, has limited ability to detect mixed *M. tuberculosis* populations. Despite these inherent constraints, the tool has been purposefully developed for practical diagnostic and patient management applications and can be readily adopted, validated, and accredited in diverse laboratory settings.

In summary, tbAMR constitutes a versatile, reliable, and accurate analytical pipeline that builds upon the validation performed in this study and other high-quality investigations [27, 31–33] to deliver optimal predictive performance for clinical decision-making and streamlined outputs tailored to CPHM needs. The tool has also been integrated into the Center for Pathogen

Genomics bioinformatics portal, providing a graphical user interface and extending access to bioinformatics resources for LMICs and other resource-limited environments. While this development does not eliminate the need for medical microbiologists and scientists in such settings to cultivate genomic expertise to use tbAMR outputs effectively, the intuitive graphical interface, accompanied by comprehensive documentation, is intended to facilitate broader adoption in high-tuberculosis-burden countries.

Conclusion

In conclusion, we developed a practical tool for deployment in real-world CPHM settings. It allows users to customize reporting parameters and outputs to local requirements readily and is accessible via both command-line and web-based interfaces. By openly sharing validation methodologies and datasets, the authors aim to support other laboratories in implementing and accrediting similar workflows in line with ISO standards, thereby enhancing patient care. The findings also identify priority areas for future research to deepen understanding of AMR determinants in *M. tuberculosis*, particularly the genotype–phenotype relationships for newer second-line agents. Additional advances, such as direct sequencing of clinical specimens, will further enhance the clinical utility of genomic technologies. Sustained collaboration among researchers, clinicians, and public health authorities will be critical for the continued refinement of these processes, ultimately improving patient management and strengthening timely public health interventions.

Acknowledgments: None

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Bennett JE, Dolin R, Blaser MJ. Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases. 9th ed. Elsevier; 2020.
2. WHO. WHO consolidated guidelines on tuberculosis. Module 4: treatment and care. Geneva:

- World Health Organization; 2025 [cited 2025 Sep 11]. Available from: <https://iris.who.int/bitstream/handle/10665/380799/9789240107243-eng.pdf?sequence=1>
3. WHO. TB guidelines [Internet]. Geneva: World Health Organization; 2021 [cited 2022 Mar 5]. Available from: <https://www.who.int/publications/digital/global-tuberculosis-report-2021/featured-topics/tb-guidelines>
 4. Carroll K, Pfaller M, Landry M. Manual of clinical microbiology. 12th ed. Wiley; 2019.
 5. Denholm JT, de Vries G, Anthony R, Smith A, Jones B, Brown C, et al. Considering best practice standards for routine whole-genome sequencing for TB care and control. *IJTL Open*. 2024;1:431-7.
 6. Donnan EJ, Marais BJ, Coulter C, White D, Black E, Green F, et al. The use of whole genome sequencing for tuberculosis public health activities in Australia: a joint statement of the National Tuberculosis Advisory Committee and Communicable Diseases Genomics Network. *Commun Dis Intell*. 2023;47:47.
 7. ISO. ISO 15189:2022 Medical laboratories—Requirements for quality and competence [Internet]. Geneva: International Organization for Standardization; 2022 [cited 2023 Jul 4]. Available from: <https://www.iso.org/standard/76677.html>
 8. Ballard SA, Sherry NL, Howden BP. Public health implementation of pathogen genomics: the role for accreditation and application of ISO standards. *Microb Genom*. 2023;9:001097.
 9. Hunt M, Bradley P, Lapierre SG, He H, Lim I, Foo J, et al. Antibiotic resistance prediction for *Mycobacterium tuberculosis* from genome sequence data with Mykrobe. *Wellcome Open Res*. 2019;4:191.
 10. Phelan JE, O'Sullivan DM, Machado D, Ramos J, Fernandes P, Rodrigues M, et al. Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. *Genome Med*. 2019;11:41.
 11. Hunt M, Mather AE, Sánchez-Busó L, Page A, Parkhill J, Keane JA, et al. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microb Genom*. 2017;3:e000131.
 12. Spies R, Crook DW, Peto TEA, Hoosdally S, Iqbal Z, Smith EG, et al. A systematic evaluation of automated *Mycobacterium tuberculosis* complex whole genome sequencing analysis pipelines. SSRN [Preprint]. 2024 Dec 23 [cited 2025 Mar 18]. Available from: <https://doi.org/10.2139/ssrn.5064085>
 13. WHO. Catalogue of mutations in *Mycobacterium tuberculosis* complex and their association with drug resistance (version 2) [Internet]. Geneva: World Health Organization; 2023 [cited 2025 Dec 10]. Available from: <https://www.who.int/publications/i/item/9789240082410>
 14. Walker TM, Miotto P, Köser CU, Feuerriegel S, Niemann S, Cirillo DM, et al. The 2021 WHO catalogue of *Mycobacterium tuberculosis* complex mutations associated with drug resistance: a genotypic analysis. *Lancet Microbe*. 2022;3:e265-73.
 15. Horan K. MDU-PHL/tbtamr [software]. 2024 Oct 23 [cited 2024 Nov 12]. Available from: <https://github.com/MDU-PHL/tbtamr>
 16. Vasimuddin M, Misra S, Li H, Aluru S. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS); 2019 May 20-24; Rio de Janeiro, Brazil. IEEE; 2019. p. 314-24 [cited 2024 Nov 12]. Available from: <https://ieeexplore.ieee.org/document/8820962>
 17. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv [Preprint]*. 2012 Jul 20 [cited 2024 Oct 1]. Available from: <http://arxiv.org/abs/1207.3907>
 18. Cingolani P, Platts A, Wang L, Coon M, Nguyen T, Wang L, et al. 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 (Austin)*. 2012;6:80-92.
 19. Rausch T, Zichner T, Schlattl A, Stütz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics*. 2012;28:i333-9.
 20. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *Gigascience*. 2021;10:giab008.
 21. Huang W, Li L, Myers JR, Marth GT. ART: a next-generation sequencing read simulator. *Bioinformatics*. 2012;28:593-4.

22. McTavish EJ. TreeToReads [software]. 2022 Sep 14 [cited 2022 Nov 16]. Available from: <https://github.com/snacktavish/TreeToReads>
23. Rambaut A, Grassly NC. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. *Comput Appl Biosci*. 1997;13:235-8.
24. Sukumaran J, Holder MT. DendroPy: a Python library for phylogenetic computing. *Bioinformatics*. 2010;26:1569-71.
25. The CRyPTIC Consortium. A data compendium associating the genomes of 12,289 *Mycobacterium tuberculosis* isolates with quantitative resistance phenotypes to 13 antibiotics. *PLoS Biol*. 2022;20:e3001721.
26. The CRyPTIC Consortium and the 100,000 Genomes Project. Prediction of susceptibility to first-line tuberculosis drugs by DNA sequencing. *N Engl J Med*. 2018;379:1403-15.
27. Coll F, Phelan J, Hill-Cawthorne GA, Nair M, Mallard K, Ali S, et al. Genome-wide analysis of multi- and extensively drug-resistant *Mycobacterium tuberculosis*. *Nat Genet*. 2018;50:307-16.
28. Ezewudo M, Borens A, Chiner-Oms Á, Miotto P, Cirillo DM, Phelan J, et al. Integrating standardized whole genome sequence analysis with a global *Mycobacterium tuberculosis* antibiotic resistance knowledgebase. *Sci Rep*. 2018;8:15382.
29. Seemann T. Snippy [software]. 2022 Mar 19 [cited 2022 Mar 26]. Available from: <https://github.com/tseemann/snippy>
30. Phelan J. jodyphelan/TBProfiler [software]. 2022 Feb 15 [cited 2022 Feb 28]. Available from: <https://github.com/jodyphelan/TBProfiler>
31. Shaw B, von Bredow B, Tsan A, Garner O, Yang S. Clinical whole-genome sequencing assay for rapid *Mycobacterium tuberculosis* complex first-line drug susceptibility testing and phylogenetic relatedness analysis. *Microorganisms*. 2023;11:2538.
32. Mahé P, El Azami M, Barlas P, Tournoud M. A large scale evaluation of TBProfiler and Mykrobe for antibiotic resistance prediction in *Mycobacterium tuberculosis*. *PeerJ*. 2019;7:e6857.
33. Barilar I, Battaglia S, Borroni E, Cabibbe AM, Miotto P, Cirillo DM, et al. Quantitative measurement of antibiotic resistance in *Mycobacterium tuberculosis* reveals genetic determinants of resistance and susceptibility in a target gene approach. *Nat Commun*. 2024;15:488.
34. Black A, MacCannell DR, Sibley TR, Bedford T. Ten recommendations for supporting open pathogen genomic analysis in public health. *Nat Med*. 2020;26:832-41.
35. Sundermann AJ, Rosa R, Harris PNA, Harrison LH, Mathers AJ, Peaper DR, et al. Pathogen genomics in healthcare: overcoming barriers to proactive surveillance. *Antimicrob Agents Chemother*. 2025;69:e0147924.
36. Pronyk PM, de Alwis R, Rockett R, Basu S, Bhattacharya S, Garg P, et al. Advancing pathogen genomics in resource-limited settings. *Cell Genom*. 2023;3:100443.
37. Siepel A. Challenges in funding and developing genomic software: roots and remedies. *Genome Biol*. 2019;20:147.
38. Helmy M, Awad M, Mosa KA. Limited resources of genome sequencing in developing countries: challenges and solutions. *Appl Transl Genomics*. 2016;9:15-9.
39. Goossens SN, Heupink TH, De Vos E, Van Rie A, Verbruggen E, De Smet K, et al. Detection of minor variants in *Mycobacterium tuberculosis* whole genome sequencing data. *Brief Bioinform*. 2022;23:bbab541.
40. Seah YM, Stewart MK, Hoogestraat D, Crandall H, Chen L, Cookson BT, et al. In silico evaluation of variant calling methods for bacterial whole-genome sequencing assays. *J Clin Microbiol*. 2023;61:e0184222.