Bringing TB genomics to the clinic: A comprehensive pipeline to predict antimicrobial susceptibility from genomic data, validated and accredited to ISO standards.

Kristy Horan[*1], *Linda Viberg*[2], *Susan Ballard*[1], *Maria Globan*[2], *Kerrie Stevens*[1], *Katherine Bond*[2], *Jessica R Webb*[3], *Thinley Dorji*[3], *Deborah Williamson*[2,5], *Michelle Sait*[1], *Ee Laine Tay*[7], *Justin T Denholm*[8,4], *Torsten Seemann*[1,3], *Benjamin P Howden*[§,1,3,6,9], *Norelle L Sherry*[§,1,3,6]

§ Contributed equally as senior authors

Affiliations

1. Microbiological Diagnostic Unit Public Health Laboratory (MDU-PHL), Department of Microbiology & Immunology, University of Melbourne at the Peter Doherty Institute for Infection & Immunity, Melbourne, Victoria, Australia
2. Victorian Infectious Diseases Reference Laboratory (VIDRL), at the Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia
3. Department of Microbiology and Immunology, The University of Melbourne, at the Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia
4. Victorian Tuberculosis Program, Melbourne Health, at the Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia
5. Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Melbourne, Victoria, Australia
6. Department of Infectious Diseases, Austin Health, Heidelberg, Victoria, Australia
7. Communicable Disease Epidemiology and Surveillance, Health Protection Branch, Public Health Division, Department of Health, Victoria, Australia
8. Department of Infectious Diseases, The University of Melbourne, at the Peter Doherty Institute for Infection and Immunity, Melbourne, Victoria, Australia
9. Centre for Pathogen Genomics, The University of Melbourne, Melbourne, Victoria, Australia

Abstract

Background: Whole genome sequencing shows promise to improve the clinical management of tuberculosis, but bioinformatic tools tailored for clinical reporting and suitable for accreditation to ISO standards are currently lacking.

Methods: We developed tbtAMR, a comprehensive pipeline for analysis of *Mycobacterium tuberculosis* genomic data, including inference of phenotypic susceptibility and lineage calling from both solid and broth (MGIT) cultures. We used local and publicly-availably real-world data (phenotype and genotype) and synthetic genomic data to determine the appropriate quality control metrics and extensively validate the pipeline for clinical use. We combined and curated the large global databases of resistance mutations, fine-tuned for clinical purposes, by minimising false-positives whilst maintaining accuracy.

Findings: tbtAMR accurately predicted lineages and phenotypic susceptibility for first- and second-line drugs, including from broth (MGIT) cultures. We designed and implemented a reporting template suitable for clinical and public health users and accredited the pipeline to ISO standards.

Interpretation: The tbtAMR pipeline is accurate and fit-for-purpose for clinical and public health uses. Report templates, validation methods and datasets are provided here to offer a pathway for laboratories to adopt and seek their own accreditation for this critical test, to improve the management of tuberculosis globally.

Funding: No specific funding was received for this study.

Introduction

Mycobacterium tuberculosis (Mtb) is the causative agent of tuberculosis (TB), a disease that is globally widespread predominantly affecting people in low- and middle-income communities and is the leading cause of death from a single infectious agent.[1]

Treatment of TB infections is lengthy and expensive, requiring combination therapy with more than one antimycobacterial drug prescribed for a minimum of six months for uncomplicated disease. The slow growth of Mtb *in vitro* means that clinicians treating patients with TB must choose treatment regimens before phenotypic drug susceptibility test (DST) results are available, which can take at least two months[2] in some settings. Treatment is usually commenced with four antimycobacterial drugs ('first-line agents'): rifampicin, isoniazid, pyrazinamide and ethambutol, followed by rationalisation to two agents (rifampicin and isoniazid, if susceptible), or an alternative regimen including second-line agents if resistant. Whole genome sequencing (WGS) has the advantage of providing comprehensive results from a Mtb sample in a single test, potentially significantly reducing turnaround time to DST for both first-line and other antimycobacterial drugs, especially when undertaken from Mycobacteria Growth Indicator Tube (MGIT) culture. In addition, WGS can provide increased reliability of resistance detection for some drugs, due to known challenges with phenotypic reproducibility, such as ethambutol and pyrazinamide[3]. WGS data is also being used to provide support for outbreak investigation and contact tracing. Therefore, the use of routine WGS for Mtb has the potential to improve patient management in a clinical setting and public health outcomes, and new Australian national guidelines recommend routine use and reporting of findings for these reasons[4].

There are many high-quality bioinformatics tools available for identification of genomic determinants of AMR in Mtb, including Mykrobe,[5] TB-Profiler[6] and ariba[7]. Additionally, databases of genomic determinants implicated in AMR in Mtb have become more comprehensive in recent years, including the WHO Mtb Mutation Catalogue[5,8]. However, they are not designed for reporting within a clinical and public health laboratory (CPHL), often being over-inclusive. This over-inclusivity is useful in research settings but is not suited to a CPHL environment, as high false-positive rates jeopardise clinical care by ruling out effective first-line therapies. Despite the urgent need for routine identification of Mtb AMR from genomic data to inform therapy and the wealth of available bioinformatic resources, few public health laboratories globally have such processes in place to report AMR from WGS data for Mtb. Many factors need to be considered when implementing such an AMR detection program, particularly where patient management may be impacted, including performance of tools in predicting drug-resistance, the impact of laboratory methodologies and existing processes, reporting structure,

data management in a CPHL, and test accreditation by national accreditation body to International Organization for Standardization (ISO) standards (ISO15189:2022).[9]

Here we demonstrate the validation of a real-time routine analysis pipeline for inference of phenotypic resistance in Mtb, based on detection of genomic variation. We have subsequently gained accreditation to ISO standard 15189:2022 for this pipeline, tbtAMR, for clinical and public-health reporting of genomic AMR detection in Mtb. The pipeline and validation approach are made available here for other laboratories globally to aid in the transition to WGS for genomic AMR reporting.

Methods

Implementation

tbtAMR software development

Numerous software options exist for interrogation of Mtb for the detection of AMR determinants, including Mykrobe[5] and TBProfiler[6], each with pros and cons. The software tool developed in our setting, tbtAMR (https://github.com/MDU-PHL/tbtamr), is a python package. It takes as input paired-end Illumina fastq files and implements TBProfiler using a custom database of mutations and generates an inferred antibiogram for clinical reporting (Figure 1).

Mutational catalogue

The mutational catalogue implemented within tbtAMR includes mutations from TB-Profiler[10] (v 4.0), Mykrobe[11] and the WHO catalogue[8]. The tbtAMR database consists of mutations and the antibiotics to which they confer resistance, as well as degree of inferred resistance (resistant or low-level resistant) and the confidence that the mutation is indicative of the degree of resistance for this prediction. Confidence scores are provided as High (odds ratio [OR] > 10), Moderate (1 < OR < 10) and Unconfirmed (0 < OR < 1), where OR is the OR data supplied by the WHO catalogue (Supplementary Table 1). Further details of database development and curation can be found in Supplementary methods.

Validation dataset

We included three data types in our validation dataset (Supplementary Figure 1). Firstly, we included sequences generated at MDU PHL and with phenotypic data generated at the Mycobacteria Reference Laboratory (MRL) (Supplementary Methods). Secondly, publicly available sequences were included, selected to be representative of all major global lineages and with phenotypic data available for first-line agents (with a subset also having data available for second-line agents). Lastly, we also used simulated genomic data to supplement the existing datasets. Paired-end sequence data was generated using TreeToReads software[12–15], directly from the H37rV strain (RefSeq accession NC_000962.3) or following introduction of 2–4 variants per 10,000 bases at known positions with an error-profile representative of the NextSeq500 instruments at MDU PHL. These simulated reads were mixed (Supplementary Figure 2) to simulate different allelic frequencies (0.01 to 0.99) across a range of average genome depth (10x to 200x).

Measuring the accuracy of Mtb genomic sequence recovery

For simulated genomes, we tested the recovery (identification) of introduced variants at a minimum depth of 10X and 20X, meaning that to make a base call there was least 10 or 20 reads across each position. For each position in the reference genome, a true positive (TP) result was observed if a variant was recovered where one had been introduced, a true negative (TN) result was observed where the same base as the reference was recovered where no variant had been introduced. A false positive (FP) result was observed where a variant was recovered when no variant had been introduced, and a false negative (FN) result was observed where a reference base was recovered but a variant was had been introduced (Supplementary Figure 2).

We also used publicly available sequences to determine the concordance of variant detection in sequences by tbtAMR with the approach described by Ezewudo *et al*. These sequences were used to support the limit of detection (average genome depth) of the tbtAMR pipeline. To simulate different quantities of DNA recovery, reads were randomly down-sampled to 0.8, 0.6, 0.4 and 0.2 of the original population, in triplicate (using seqkit sample). The variants were considered concordant if those identified from the down-sampled sequence were also identified in the original sample.

Assessing suitability of MGIT sequences for inference of AMR

To assess the utility of DNA generated from MGIT sequences for the interpretation of relationships and the inference of AMR, we evaluated matched sequences generated from solid cultures (gold standard) and MGITs for 66 primary samples (total of 161 sequences evaluated, as some samples had >1 sequence per culture type), where all sequences had passed the sequencing quality checks (Supplementary Figure 1). To establish whether sequences generated from MGIT culture provide a comparable degree of utility for identifying and interpreting relationships, we assessed the consistency of variant detection and the pairwise distance between the matched sequences from MGIT culture and solid culture.

Analysis of inferred phenotypic DST and phylogenetic lineage from genomic data

tbtAMR leverages the most recent definitions of phylogenetic lineage, including lineages 1–9 as well as animal adapted species/sub-species[6]. Phylogenetic lineage was assessed by the concordance of phylogenetic lineage determined by tbtAMR and compared with available public data[16]. The inference of phenotypic DST from genomic sequence data was assessed by determining whether the presence of a genomic resistance determinant, SNP (single nucleotide polymorphism) and/or insertion/deletions correlated with the phenotypic DST results, with sensitivity, specificity, negative predictive value (NPV), positive predictive value (PPV) and accuracy recorded (Supplementary Table 2). A true positive (TP) was recorded when a mutation was present and the phenotype was resistant; a

true negative (TN) was recorded when no mutation was detected and the phenotype was susceptible; a false positive (FP) was recorded when a mutation was present, but the phenotype was susceptible, and a false negative (FN) was recorded when no mutation was detected and the phenotype was resistant.

Results

Validation of tbtAMR tool

Accuracy of sequence recovery by tbtAMR

Firstly, to assess the impact of minimum depth for base calling, we compared FP SNPs and the percentage of SNPs recovered across a range of average genome depth, maintaining the allelic frequency at 100%. We observed a higher number of FP SNPs, mean 6.6 (95% confidence interval (CI) 3.9-9.3) when using the standard criteria of 10X, which may result in over identification of resistance mechanisms. Increasing the minimum depth for base calling to 20X improved the number of FP SNPs to mean of 1.1 (CI 0.7-1.5), with little impact on the percentage of introduced SNPs recovered (Supplementary Figure 2). We further examined the impact of varying average genome depth and allelic frequency on the ability of tbtAMR to accurately identify mutations in the simulated sequences. Where average genome depth < 40 and allelic frequency < 10% sensitivity was only 2.1% (CI 1.48%-2.66%) and False Discovery Rate (FDR) was 41.1% (CI 31.1%-51.2%). However, maintaining 20 reads minimum depth for base-calling, an average genome depth of ≥40X and allele frequencies ≥10% gave the best balance (Figure 2A and Supplementary Figure 4), with an overall variant identification sensitivity of 96.3% (CI 95.5%-97.1%) and low FDR of 0.21% (CI 0.11%-0.31%). Hence, the minimum acceptable read depth for running tbtAMR was set to ≥40X and minimum allele frequency of 10%.

tbtAMR performance was compared to results published for the UVP pipeline, identifying 285/296 of the variants in the published dataset[17] (96.3% concordance). We determined that 3/11 discordances were likely due to low frequency mutations identified by the tbtAMR pipeline, and the remainder likely due to differences in SNP detection between tbtAMR and the UVP pipeline[17]. Down-sampling this public sequence data demonstrated that at a median depth of ≥40X, ≥99% concordance between SNPs detected in the original sequence by tbtAMR and the down sampled reads was observed, indicating tbtAMR was able to reproducibly detect genomic AMR determinants in Mtb sequences (Figure 2B).

Sequencing from liquid culture (MGIT) compared to solid culture

Firstly, we compared AMR determinants identified by tbtAMR of matched sequences from solid and MGIT cultures. Using average genome depth of ≥40X and a minimum allele frequency of 10% as established above, 93.9% (62/66) matched groups of sequences from MGIT and solid cultures exhibited detection of the same genomic determinants of AMR. In two of the four discordant samples, mutations were observed in *rrl*, a 23S rRNA gene (implicated in linezolid resistance). The remaining two discordances were observed in groups where no single lineage could be detected in one sequence from each group. It is possible that cases such as this reflect genuine in-host diversity (e.g. mixed infections) and should not be excluded from analysis, however, it is also possible that observations such as these are the result of laboratory error or contamination. In these cases, inferred AMR results will be reported on an Interim basis, while awaiting investigation and resequencing (for detailed discussion see below).

We routinely undertake identification of genomic relationships, for use in epidemiological investigations, using a validated and accredited pipeline (https://github.com/MDU-PHL/bohra). Therefore, it is important to establish the impact of using sequences from MGITs on this process. Thus, we also examined the SNP distances between matched groups (pairs/trios of sequences from MGITs and solid cultures) to determine the consistency of sequences between solid culture and MGITs from the same primary samples. The median distance between sequences from the same primary sample was 1 SNP (IQR +/- 3). Pairwise distances between matched sets fell within the 5 SNP threshold for 59/66 matched groups (89.4%); a further five matched groups fell within 12 SNPs (7.6%) and would therefore be identified as genomically-related in our routine analyses (Figure 3). Two remaining groups contained a single sequence each for which multiple phylogenetic lineages were observed, consistent with mixed infection or laboratory contamination. For the purposes of interpretation of genomic relationships, sequences where no single lineage can be identified will be failed and resequencing undertaken.

Accuracy of phylogenetic lineage calling by tbtAMR

tbtAMR accurately identified lineages compared to lineages reported from public datasets[16], (99.2% concordance). Of these discordant results (Supplementary Table 3), seven were cases where tbtAMR identified two different phylogenetic lineages in the sequence. In each case, the reported lineage was one of the lineages detected by tbtAMR, therefore these likely reflect mixed sequences, where tbtAMR was able to identify the minor allele.

Inference of phenotypic AMR from genomic data and discrepancy characterisation

Many studies have explored the inference of phenotypic AMR from genomic data using various tools[5,6,8,17,18]. We do not aim to recapitulate these high-quality studies here, but to assess the suitability of tbtAMR in our setting for inference and reporting of AMR in Mtb for clinical and public health.

We began with a dataset of 3374 sequences from both publicly available datasets[16–18] and sequences generated in-house (Supplementary Figure 1). Implementing the quality control thresholds and limits of detection outlined above, we excluded 371 sequences, leaving 3006 sequences of sufficient quality to include in the validation of tbtAMR for inference of phenotypic AMR (Figure 4A). This dataset contained > 500 sequences for each of the common international lineages (Figure 4B).

The performance of tbtAMR in inferring phenotypic resistance to first-line drugs was excellent, with accuracy > 95% across all first-line drugs (Table 2 and Figure 5A). For the purposes of validation for use in a CPHL, investigation and resolution of such discrepancies needs to be undertaken. Discrepancies observed in prediction of pyrazinamide, isoniazid and ethambutol have been observed in other studies [REF] and are consistent with our observations.

Phenotypic tests based on both liquid and solid culture methods for moxifloxacin resistant have been shown to have low levels of consistency and also inconsistent genomic mechanisms[19].Therefore for validation of moxifloxacin, we only included samples with phenotypic data available generated in our laboratories, to ensure that the comparisons made were consistent with methodology that was used for reporting resistance in our setting. The tbtAMR pipeline was able to predict moxifloxacin resistance phenotype from the detected genotype with very high accuracy (>98%), sensitivity (100%) and specificity (>98%) (Table 2 and Figure 5B). The positive predictive value (85.7%) was limited by the low proportion of resistance (genotypic and phenotypic) in the dataset (only 12 true positives and two potential false positives).

tbtAMR predicted resistance to second-line and other drugs with performance comparable to other studies[5,20]. The high number of false-negative results for prediction of phenotypic resistance (Table 2) are likely to be due to incomplete understanding of the mechanisms of resistance to these drugs, meaning that even when mutations are correctly identified, the functional impact is unknown and therefore prediction of resistance is not possible.

Report design

Reporting of AMR from WGS data for Mtb can potentially directly impact patient management, particularly when undertaken from MGIT culture with reduced turnaround times. Three key parameters are reported for each drug to assist in clinical decision making: (i) presence or absence of

mutations including small indels and SNPs, (ii) confidence that this mutation is associated with drug resistance, and (iii) the level of drug resistance conferred by this mutation.

We have leveraged the confidence levels from the WHO mutation catalogue and supplemented it with an inferred phenotype interpretation of 'susceptible', 'low-level resistant' and 'resistant' to allow for clinically informative reporting (Supplementary Table 1). These interpretation levels are based on available laboratory evidence, supported by the literature for a particular phenotype.

As AMR genotypes and inferred antibiograms for Mtb were new to most clinicians, additional interpretive comments were constructed for inclusion in reports – for example, explaining the difference between confidence level and level of resistance in the inferred phenotype (Supplementary Methods and Supplementary Files 1). Clinical and laboratory colleagues were consulted about the draft report design to ensure the formatting and explanations were clear and unambiguous.

Implementation process

Once validation and report design were complete, Mtb genomic AMR inference using the tbtAMR tool was implemented into existing genomic workflows, including standard operating procedure documentation (SOP), staff training, incorporation into the laboratory information management system, and testing of reporting and feedback loops (e.g. reporting of samples that failed QC).

External clients and stakeholders (diagnostic laboratories, clinicians, Department of Health/TB program) were notified of the new test through existing channels, and the contact details for the medical microbiologists were included on the reports to facilitate further discussions if required.

To address any discrepancies in phenotype and genotypic results, a multidisciplinary expert panel for Mtb AMR was formed, meeting monthly to discuss individual cases and provide expert recommendations to clinicians where discrepancies were identified. This provides the opportunity for ongoing dialogue between the wet-lab microbiologists, clinicians, epidemiologists, medical microbiologists and bioinformaticians to monitor for AMR determinants of interest, discuss individual cases and discrepancies, and undertake prospective validation of novel or uncharacterised AMR determinants.

In preparation for ISO-equivalent accreditation, revalidation and reverification processes were also defined for tbtAMR (Supplementary Methods). Subsequently, the Mtb genomic AMR inference workflow has now been accredited to ISO 15189:2022 standards.

Discussion

We have demonstrated here the development of a custom pipeline with a fit-for-purpose mutational catalogue, to maximise the balance between PPV and NPV for prediction of AMR in Mtb. Performance of inferred AMR using our process is consistent with or exceeds previously published results on similar datasets[5,16]. This pipeline has been robustly validated and implemented to ISO standards, with reverification strategies for ongoing improvement, and reporting processes put in place.

While there are many excellent tools available for detection of resistance mechanisms in Mtb, the outputs and/or databases and levels of interpretation were not strictly suited to reporting in a high-throughput CPHL. Reporting and data storage methods are amongst some of the considerations which prevent 'out-of-box' use of existing bioinformatics software. We assessed the two most popular existing software tools, TB-profiler and Mykrobe and elected to leverage aspects of TB-profiler to identify genomic variation, adding a custom collection of mutations. We then implemented reporting logic with interpretive comments specific for use in a clinical setting, based on our validation and supported by literature and laboratory evidence.

Clinical and public health laboratories operate rigorous quality systems to ensure that results are robust, accurate and fit-for-purpose. Comprehensive validation of new tests or pipelines is an essential component of this quality framework, optimising the performance metrics of the process prior to implementation. However, this can also be a dynamic process, as software and database updates occur regularly in the bioinformatics field. As such, it is important to establish a regular process for pipeline reverification before any updates or modification to the code base of tbtAMR or its dependencies are approved (Supplementary Figure 5). To maintain the integrity of tbtAMR, we utilise conda, a software management framework, implementing tbtAMR in a centralised conda environment, which is only updated upon successful testing of the proposed updates.

AMR determinants are often reported as a binary 'present/absent' variable, inferring a 'resistant' or 'susceptible' phenotype respectively, but this does not always accurately reflect the complexity of genotype-phenotype correlations. It is important to ensure that mutations used to infer resistance have solid evidence supporting a role as genomic determinants of resistance. The databases underlying both TB-profiler and Mykrobe have been extensively curated and incorporate many, although not all, of the variants described in the WHO catalogue. In addition to variants highly likely to confer resistance, our database also contains many mutations that are being investigated for their role in resistance. However, we have classified these mutations as 'Not reportable' since there is currently insufficient evidence for their causative role in resistance (Supplementary Methods). This allows us to provide meaningful reports to clinicians (based on high confidence mutations), whilst still being able

to prospectively monitor and validate additional mutations. This is particularly relevant for second- and third-line drugs, such as para-aminosalicylic acid and cycloserine, as well as the newer critical antimycobacterial drugs bedaquiline, delaminid and linezolid, where a lack of data, locally and globally limits our ability to adequately validate inference of resistance to these drugs. Even for drugs which are well studied, discrepancies can still occur, and further research and iterative validation will only improve the performance of tbtAMR and other tools used by Mtb genomics community, ultimately leading to improved patient outcomes.

tbtAMR represents a robust and accurate pipeline that uses high-quality open-source tools, a database tailored to achieve the best balance between PPV and NPV for patient management, and simplified outputs for clinical and public health microbiology. We have accredited this pipeline to ISO-standards in our laboratory, and have provided the validation methods and dataset for use by other laboratories endeavouring to make a similar transition. However, our results also highlight areas where further research can enhance our knowledge of AMR determinants in Mtb, and the nuanced interactions between determinants and phenotypes with new second-line Mtb drugs is a key priority. Further enhancements of diagnostic processes, such as sequencing direct from samples, will allow us to maximise the benefit of genomic technologies for patient management. Our ongoing engagement with stakeholders, including researchers, clinicians and public health officials ensures that the processes established can be improved and refined where needed to continue to improve patient management and inform timely public health responses.

Figures and tables

Table 1 – Performance of tbtAMR pipeline to predict phenotypic susceptibility for first- and second-line antimycobacterial drugs

Drug	Accuracy (%, 95%CI)	Sensitivity (%, 95%CI)	Specificity (%, 95%CI)	PPV (%, 95%CI)	NPV (%, 95%CI)
Rifampicin (n= 2957)	99.29 (98.92, 99.54)	97.34 (95.17, 98.55)	99.57 (99.24, 99.76)	97.08 (94.85, 98.36)	99.61 (99.29, 99.79)
Isoniazid (n= 2926)	98.09 (97.52, 98.52)	97.86 (96.21, 98.8)	98.13 (97.51, 98.6)	91.79 (89.19, 93.81)	99.54 (99.17, 99.74)
Pyrazinamide (n= 2962)	96.35 (95.62, 96.97)	80.97 (76.23, 84.95)	98.15 (97.57, 98.6)	83.67 (79.06, 87.42)	97.78 (97.15, 98.28)
Ethambutol (n= 2965)	96.12 (95.36, 96.76)	95.08 (91.76, 97.1)	96.22 (95.44, 96.88)	71.1 (66.17, 75.59)	99.5 (99.15, 99.71)
Moxifloxacin (n= 170)	98.82 (95.81, 99.68)	100.0 (75.75, 100.0)	98.73 (95.5, 99.65)	85.71 (60.06, 95.99)	100.0 (97.6, 100.0)
Amikacin (n= 279)	96.77 (93.98, 98.29)	95.56 (89.12, 98.26)	97.35 (93.96, 98.86)	94.51 (87.78, 97.63)	97.87 (94.66, 99.17)
Kanamycin (n= 279)	96.77 (93.98, 98.29)	95.7 (89.46, 98.31)	97.31 (93.86, 98.85)	94.68 (88.15, 97.71)	97.84 (94.57, 99.16)
Streptomycin (n= 98)	89.8 (82.23, 94.36)	95.52 (87.64, 98.47)	77.42 (60.19, 88.6)	90.14 (81.02, 95.14)	88.89 (71.94, 96.15)
Cycloserine (n= 98)	81.63 (72.83, 88.05)	45.45 (29.84, 62.01)	100.0 (94.42, 100.0)	100.0 (79.61, 100.0)	78.31 (68.3, 85.82)
Capreomycin (n= 278)	97.12 (94.43, 98.53)	94.44 (87.65, 97.6)	98.4 (95.41, 99.46)	96.59 (90.45, 98.83)	97.37 (93.99, 98.87)
Ethionamide (n= 180)	89.44 (84.1, 93.14)	84.78 (71.78, 92.43)	91.04 (85.0, 94.8)	76.47 (63.24, 86.0)	94.57 (89.22, 97.35)
Para-aminosalicylic acid (n= 98)	95.92 (89.97, 98.4)	20.0 (3.62, 62.45)	100.0 (96.03, 100.0)	100.0 (20.65, 100.0)	95.88 (89.87, 98.38)

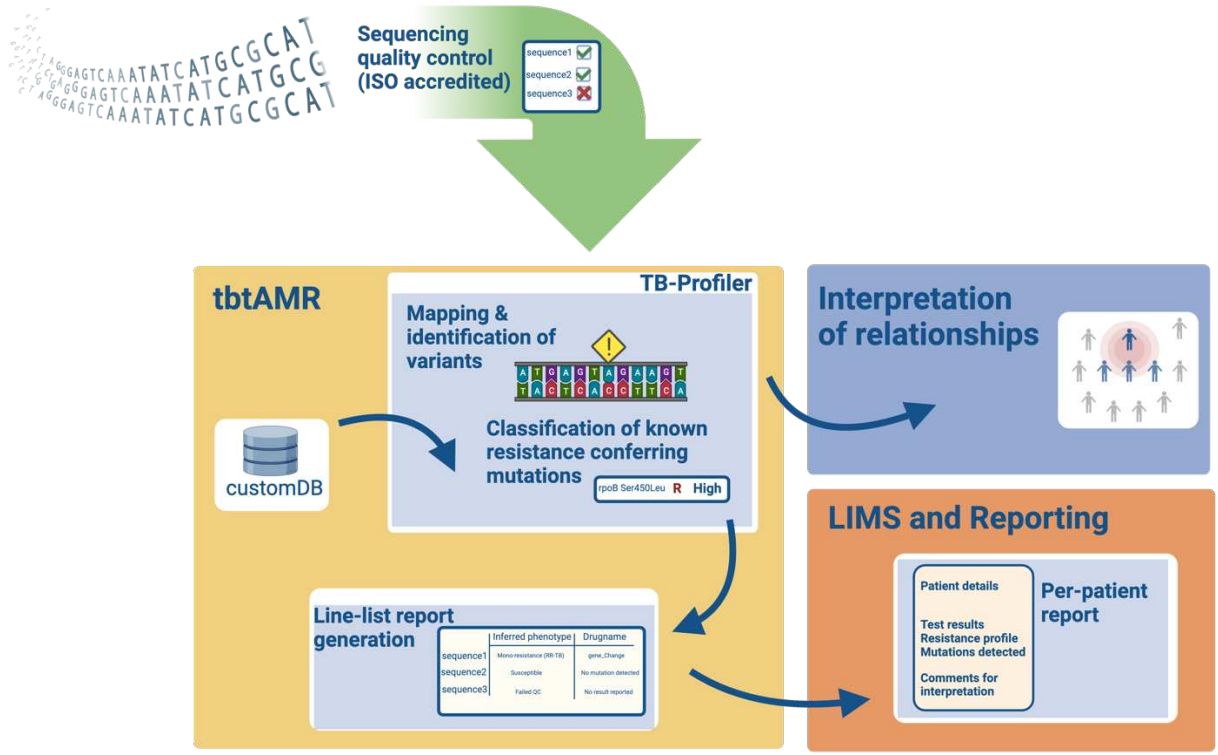
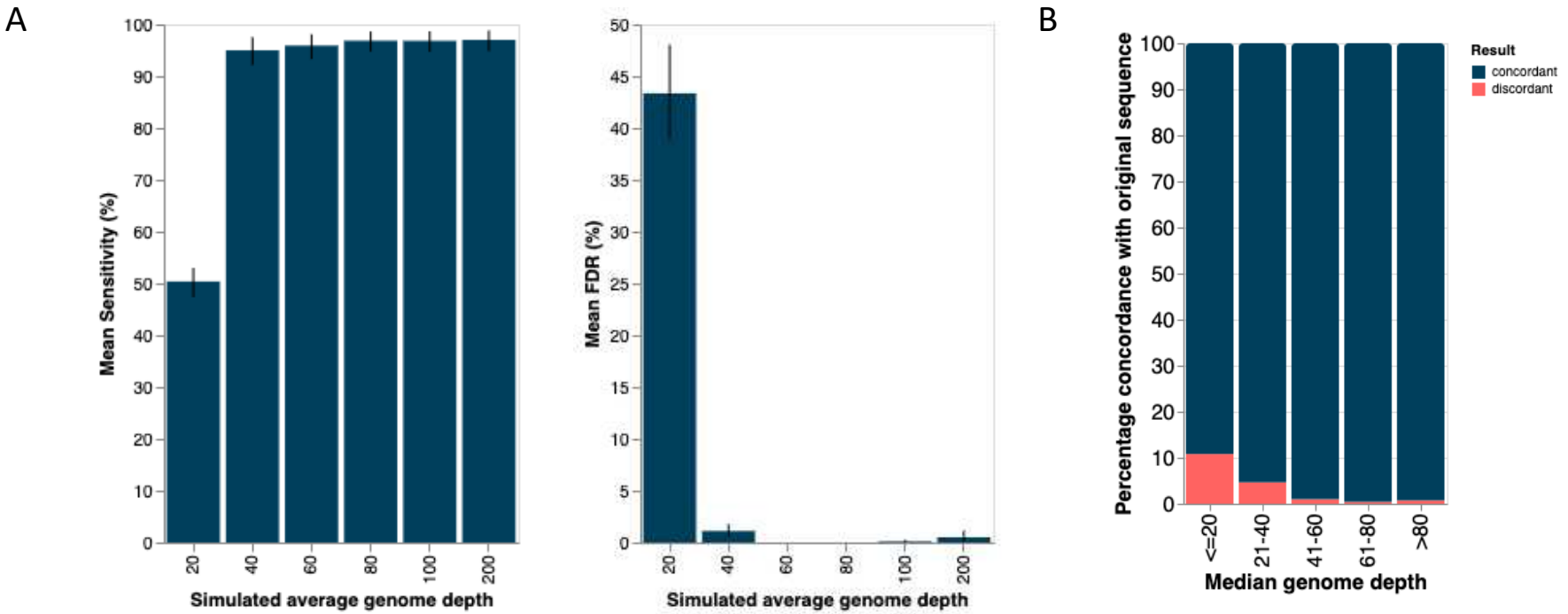


Figure 1 – Outline of data flow for the tbtAMR tool

tbtAMR takes paired-end fastq files that have passed sequencing quality control as input. TB-profiler is used to identify variants and known resistance-conferring mutations reported, based on the tbtAMR custom database. Using logic generated during data exploration, inferred phenotypes are recorded, and line lists are generated which can be bulk uploaded into the laboratory information management system (LIMS) for automated per-patient reports. In addition, the paired-end fastq files are also incorporated into the ISO accredited surveillance software utilised in our setting. ISO, International Standards Organization; DB, database.



341 **Figure 2 – Performance of tbtAMR for identification SNPs in *M. tuberculosis* genomes**

342 A) Simulated paired-end fastq (read) files were used to assess the performance of tbtAMR in recovering introduced variants. Using a minimum read criteria of
343 20X for individual base calling, tbtAMR had SNP calling sensitivity (left panel) of ≥95% and false discovery rate (FDR) (right panel) of ≤1.0% when simulated
344 average genome depth was ≥40X and allelic frequency ≥10%. B). Recovery of resistance-conferring mutations at varying simulated average genome coverage
345 levels was compared to the resistance-conferring mutations identified in the original sequence (red indicates discordance with original sequence, blue
346 indicates concordance with original sequence).

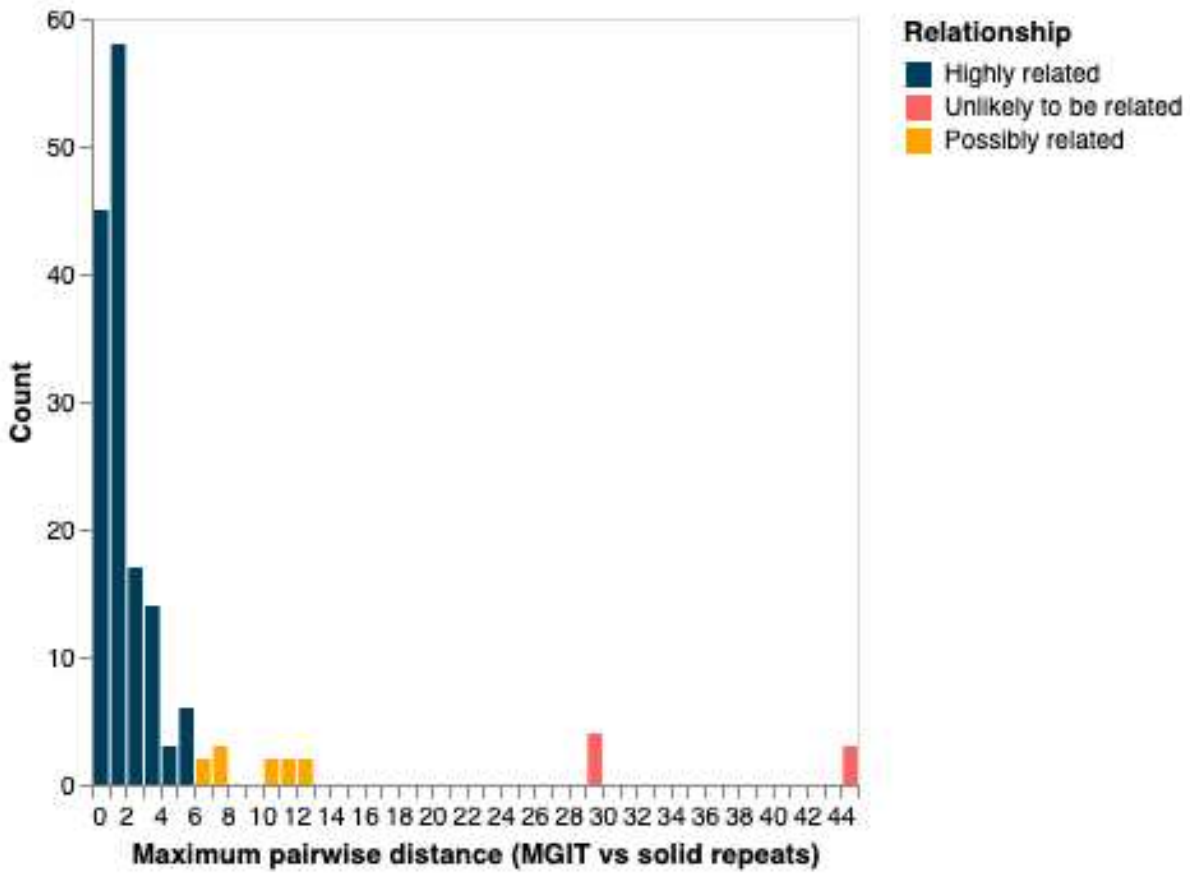


Figure 3 – Distance between sequences from matched MGIT and solid culture.

Pairwise SNP distance was calculated between sequences obtained Mycobacterial Growth Indicator Tubes (MGITs) and solid culture of the same clinical sample and the distribution plotted. Blue indicates sequences which would correctly be called highly genomically related, orange indicates sequences which would be considered possibly related and red indicates matched pairs where the pairwise distance would indicate no recent genomic relationship (evidence of mixed sequences; see Results).

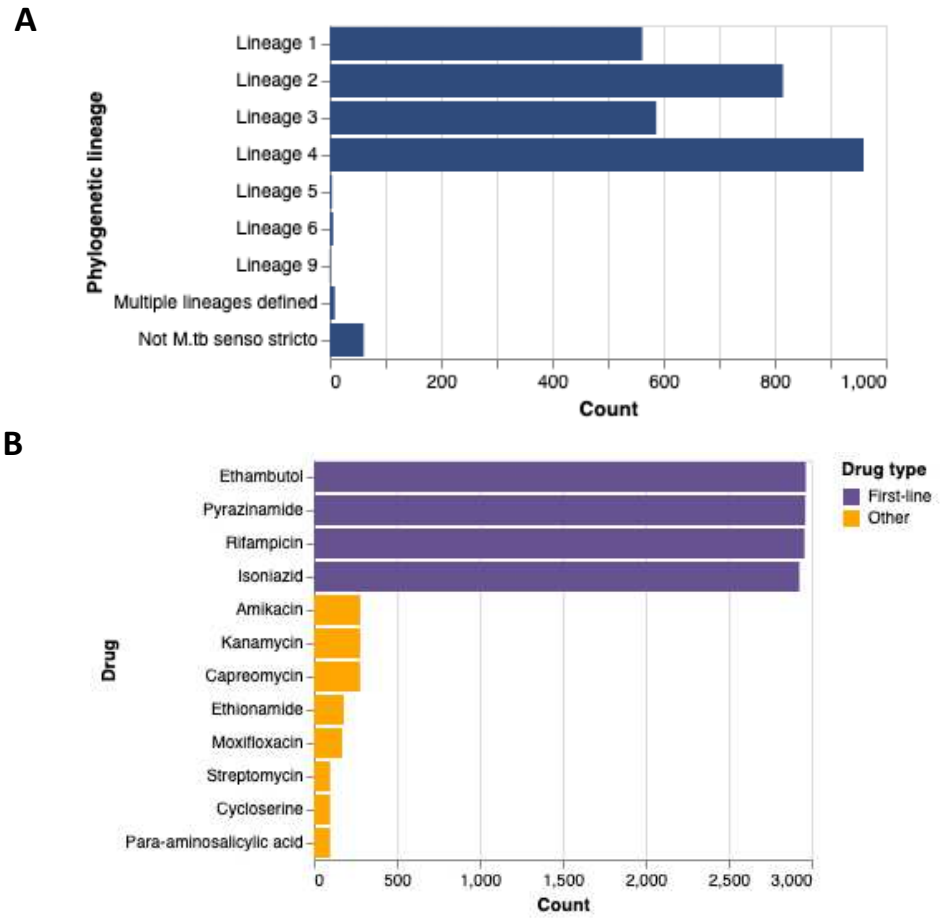
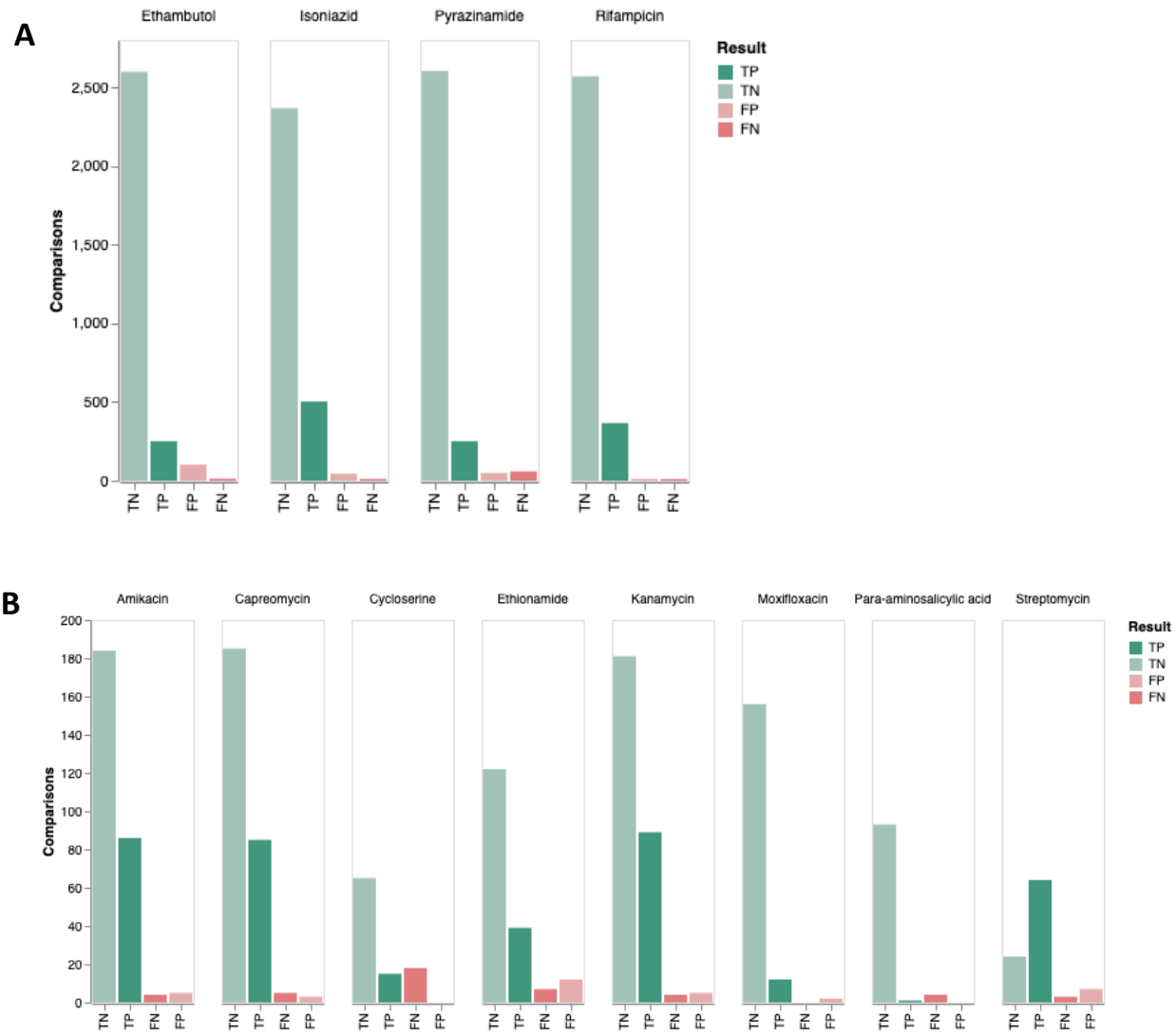


Figure 4 – Lineages and phenotypic antimicrobial resistance profiles represented in validation dataset

3016 Mtb complex sequences were used to validate the inference of AMR. A) Phylogenetic lineage was determined using tbtAMR pipeline, and grouped by lineage. Non-Mtb *sensu stricto* lineages were grouped together. B) The validation dataset represented by the phenotypic resistance for each drug.



360 **Figure 5 – Performance of tbtAMR inference of phenotypic DST**

361 Inferred phenotypes predicted by the tbtAMR pipeline was compared to phenotypic susceptibility data
362 for and results of the comparison classified into TP, TN, FP and FN for A) first-line drugs and B) other
363 drugs.

364

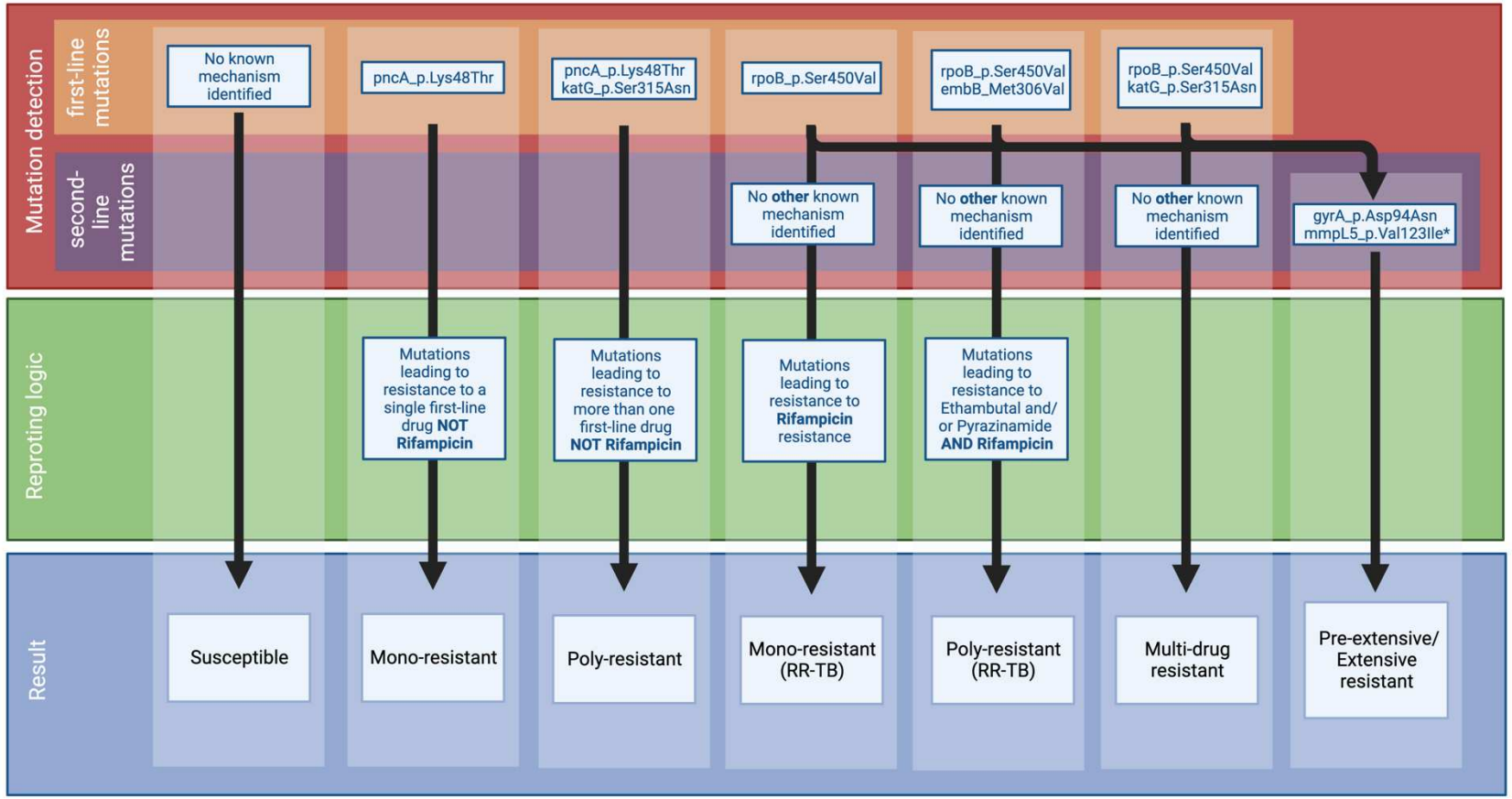


Figure 6 – Reporting of predicted drug resistance profile

Drug resistant profile is based on the WHO guidelines[21]. Note that Pre-extensive (Pre-XDR) and Extensive Drug Resistance (XDR) cannot yet be differentiated from genomic data (distinguished by bedaquiline and linezolid resistance) and are hence reported as a single group.

Ethical approval

Ethical approval was received from the University of Melbourne Human Research Ethics Committee (study number 1954615).

Acknowledgements

We thank all the laboratory staff at the Mycobacterial Reference Laboratory, Victorian Infectious Disease Reference Laboratory (VIDRL) and the Microbiological Diagnostic Unit Public Health Laboratory (MDU PHL) who undertook receipt, diagnostics, culture, DNA extraction and sequencing of the material used in this publication and the diagnostic laboratories who diligently provide samples for testing, and the Victorian TB Program and Department of Health Victoria who fund these services. We also thank the developers and maintainers of the open-source software used and cited in this publication.

Supplementary Data

Methods

Bacterial culture

Clinical samples collected for suspected TB are processed at diagnostic laboratories throughout the state of Victoria, Australia (population 6.71 million in 2022), where routine culture is performed using both broth (mycobacterial growth indicator tubes (MGIT)) and solid culture media. In Victoria, primary samples are cultured both in-house and within external laboratories. Samples with acid-fast bacilli detected are referred to the Mycobacterial Reference Laboratory (MRL) at the Victorian Infectious Diseases Reference Laboratory (VIDRL) for identification and phenotypic susceptibility testing. Broth cultures (MGIT) are sub-cultured onto solid media to provide sufficient material for downstream processes.

DNA extraction from MGITs and solid cultures

DNA was extracted from solid cultures and broth cultures (MGIT) as previously described[22] with minor modifications. In brief, 3 x 1µl loops of culture were resuspended in 700µL TE and heat killed at 95℃ for 15minutes. For MGITs, 1ml aliquots were heat killed followed by centrifugation and resuspension of the pellet into 700 µL TE. Cells were lysed through mechanical disruption and DNA precipitated with ethanol and sodium acetate followed by elution into EB buffer (QIAGEN).

Whole genome sequencing (WGS)

Extracted DNA from solid or broth (MGIT) cultures was transferred to the Microbiological Diagnostic Unit Public Health Laboratory (MDU PHL) for WGS, starting with NexteraXT (Illumina) library preparation according to manufacturer's instructions, then paired-end short-read sequencing on Illumina NextSeq500/550 platforms.

Phenotypic DST

Phenotypic drug susceptibility testing was set up for first line drugs in the BACTEC MGIT 960 system according to WHO guidelines[23]. If resistance at the critical concentrations (rifampicin 0.5µg/mL, isoniazid 0.1µg/mL, ethambutol 5.0µg/mL and pyrazinamide 100µg/mL) was detected, test was repeated and simultaneously second line drugs were set up (amikacin 1.0µg/mL, capreomycin 2.5µg/mL, ethionamide 5.0µg/mL, kanamycin 2.5µg/mL, ofloxacin 2.0µg/mL, moxifloxacin 0.25/1.0µg/mL and isoniazid 0.4µg/mL).

Reverification strategy

It is important to maintain the integrity of any process where the outcome is to be used in informing public health reporting and patient management. Many bioinformatics tools and databases are updated frequently and whilst it may be desirous to always have the most up to data versions, it is also important to ensure that no degradation of quality results occurs as a result. Therefore, it is important to have a robust reverification strategy in place to assess the impact of any changes or updates. Any updates to the underlying tools of tbtAMR, including TB-Profiler and its dependencies or new mutations, the potential impact will first be assessed, using the most appropriate dataset (Supplementary Figure 5).

1. If updates include changes to the methodology used to detect underlying detection of variants, the performance of tbtAMR to identify accurate sequence will be assessed using the simulated dataset described above.
2. If updates involve changes to the database, the performance of tbtAMR to predict AMR will be assessed using sequences from samples submitted to the Victorian MRL.
3. If any modifications impact the way TB-profiler is implemented within the tbtAMR pipeline, or the reporting logic of tbtAMR the sequences from sample submitted to Victorian MRL and public datasets where phenotypic and sequence data are available will be used to confirm that the results are consistent with the original validation results.

Any degradation of any performance metric compared to the original validation will be assessed to determine the impact on clinical reporting. Some discordances may result in improvements to prediction of AMR (increases in sensitivity and/or specificity etc). Whilst others may result in changes to interpretation, such as level of resistance or confidence in the prediction. In rare cases, updates could lead to a reduction in the confidence in results, such as failure to detect variants and/or incorrect genomic DST result, in which case the updates will be rejected, and the existing versions retained.

Cascade reporting from LIMS

Phylogenetic lineage and predicted drug resistance profiles are reported, consistent with the WHO classifications[21]. Pre-extensive drug resistance (pre-XDR) and XDR are reported together, as resistance to bedaquiline and linezolid cannot yet be easily inferred from genomic data. Rules for reporting can be defined by the user; in our lab, antimycobacterial drugs are reported in a cascade fashion. Second-

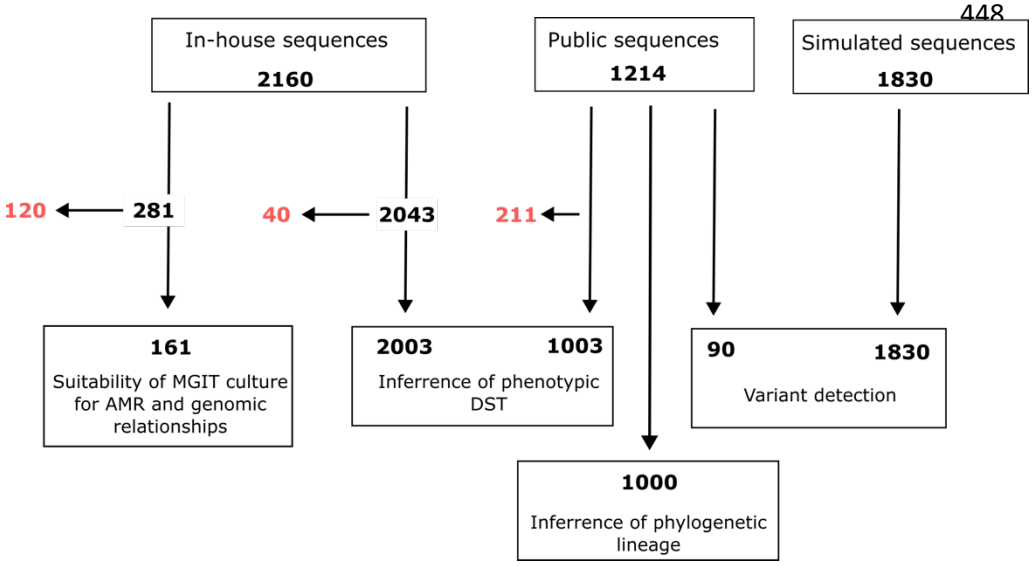
443 and third-line agents (where validated) are reported if MDR-TB, RR-TB or Pre-XDR/XDR-TB profiles are
444 identified.

445

446

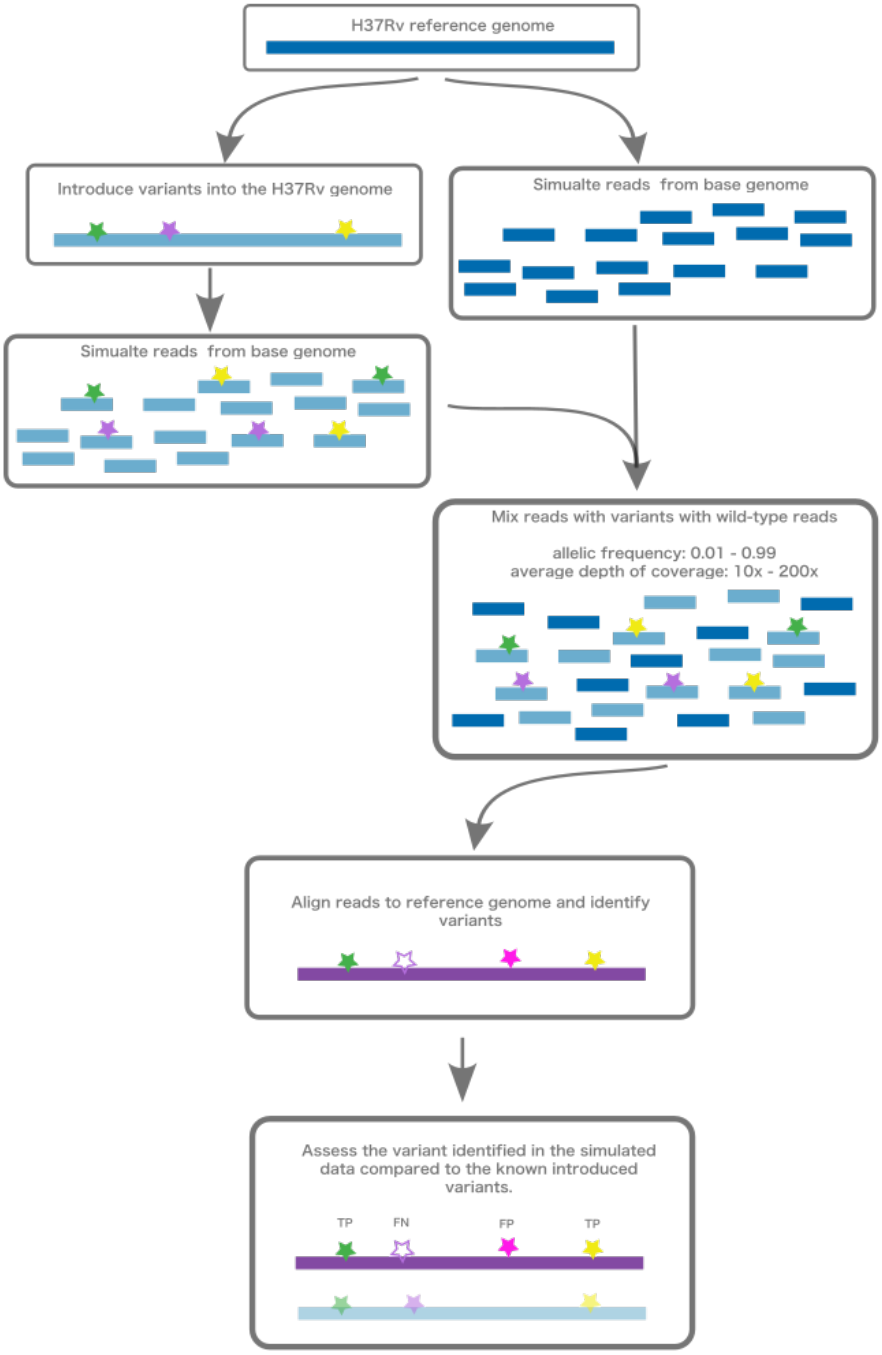
447 **Figures and tables**

448



Supplementary Figure 1 – Validation dataset

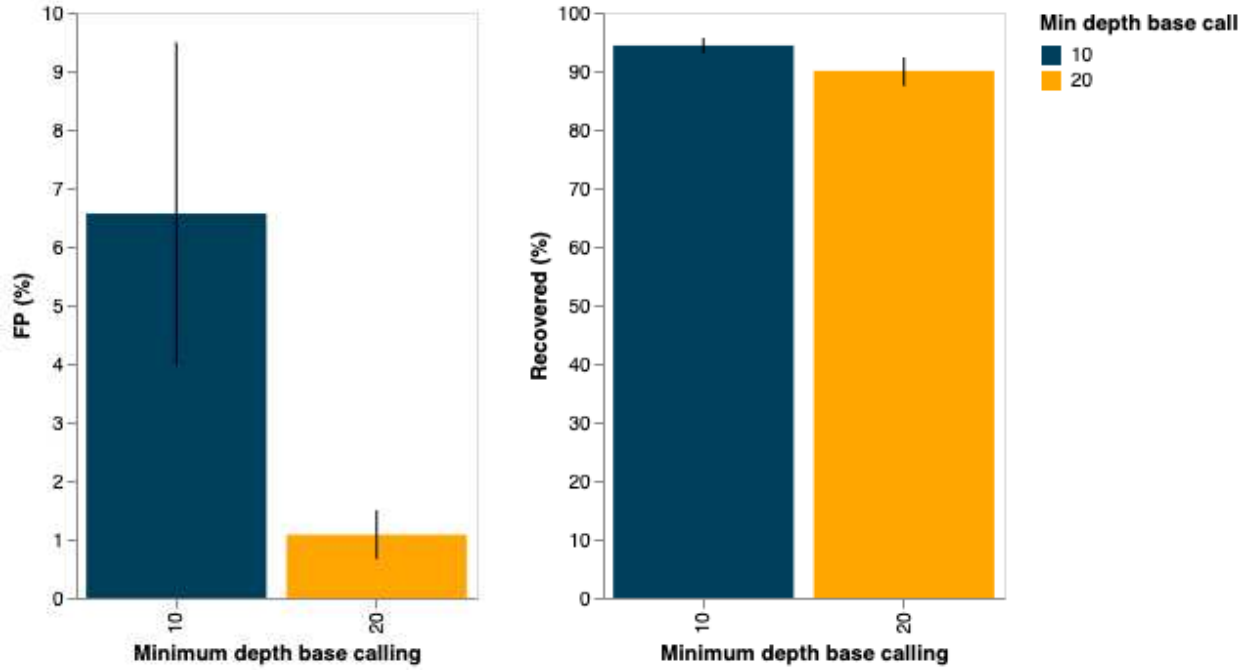
Data generated in-house at MDU, downloaded from public datasets and also simulated data were used to validate variant detection, appropriateness of sequences derived from MGIT and genomic DST. Red indicates the number of sequences excluded from inclusion due to the failure to meet quality requirements.



Supplementary Figure 2 – Generation and analysis of simulated reads

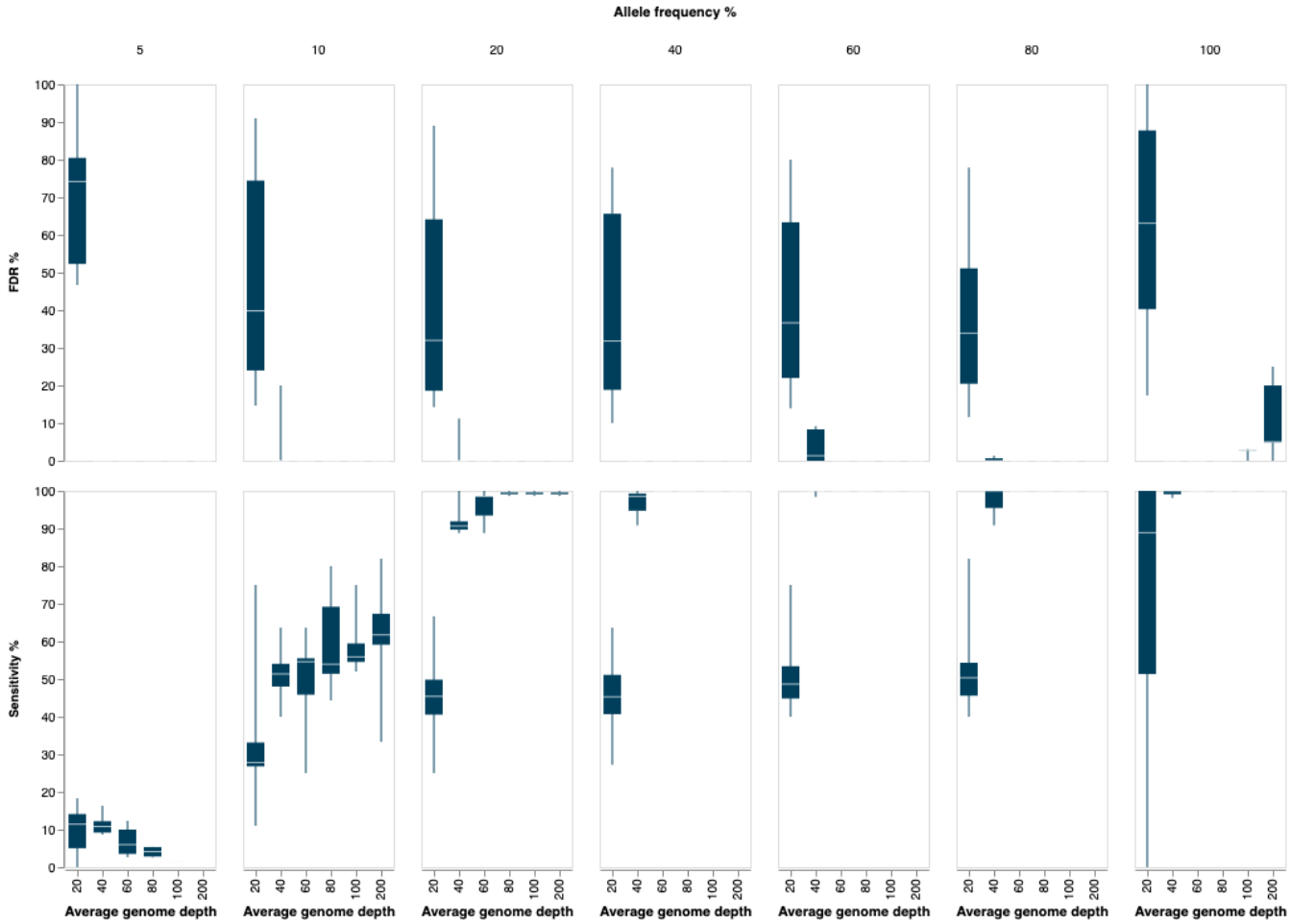
Simulated reads were generated to mimic the potential mixed alleles observed in Mtb sequencing. Variants were introduced (or not) into the H37rv reference at known positions and then mixed together at different proportions across a range of coverages. The simulated sequences were then analysed using the tbtamr pipeline and the recovery of the introduced variant was assessed. A TP result was observed where a introduced variant was recovered, a FP result was observed where a variant was reported that was not known to be present, and a FN result was observed where an introduced variant was not recovered

450
451
452



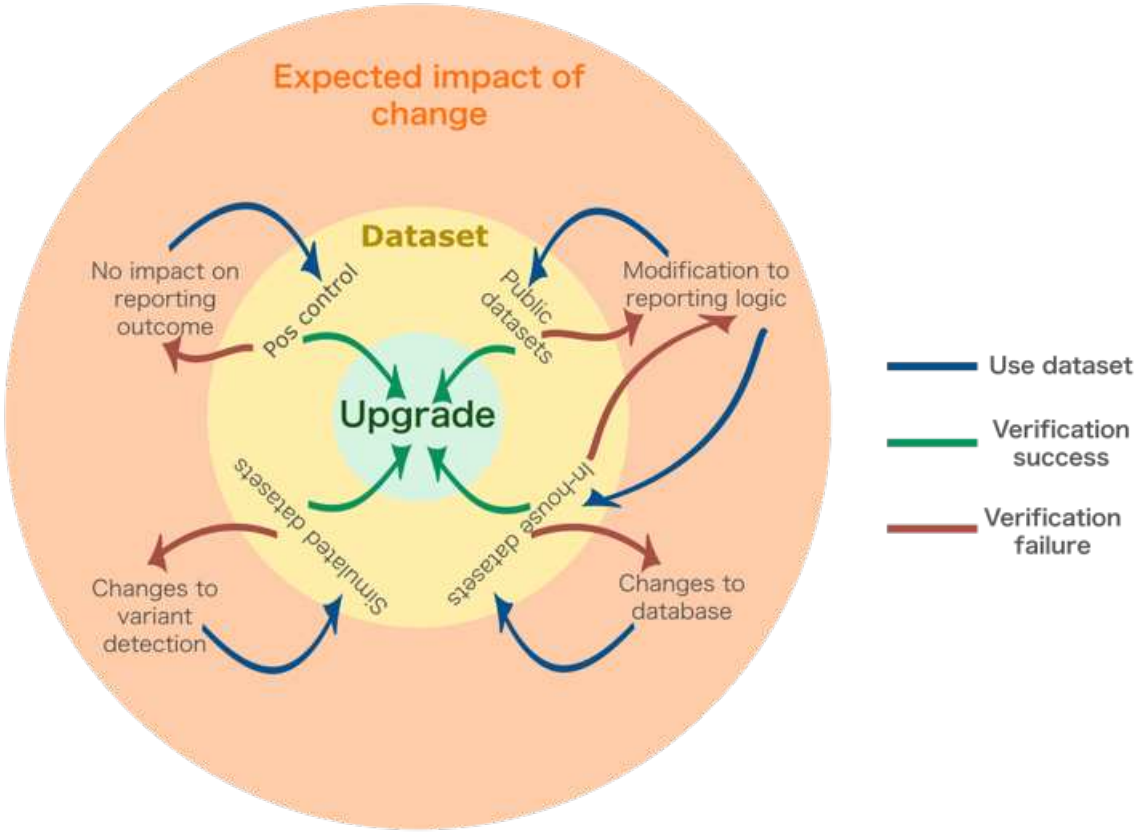
Supplementary Figure 3 – Impact of minimum read depth on performance of tbtAMR to recover SNPs from Mtb sequences

To assess the impact of minimum read depth for base calling, the percentage of A. FP and B. Introduced SNPs recovered by tbtAMR were calculated across a range of average genome depths from 10X to 200X at an allelic frequency of 100%.



Supplementary Figure 4 – Impact of genome depth and allelic frequency on performance of tbtAMR to recover SNPs from Mtb sequences

The FDR % (upper panel) and Sensitivity % (lower panel) of SNP recovery was assessed by varying the allelic frequency from 5 to 100 % and the average genome depth from 20 – 200X, whilst maintaining a minimum read depth of 20X for base calling.



Supplementary Figure 5 – Reverification process

Potential impact of proposed changes are shown in the outer circle and can include changes to the database, variant detection, modifications to reporting logic. Different datasets, shown in the middle circle, are required to address the different impact. If the result of the verification is deemed a success, then the update may proceed, however if the verification fails, then the update will not proceed and cause of failure established.

Supplementary Table 1 – Generation of confidences and resistance levels for tbtAMR database

Odds ratio	Evidence for low-level resistance	Reported parameters	
		Confidence	Resistance level
OR≥10	NA	High	Resistant
1≤ OR <10	Yes	High	Low-level resistant
1≤ OR <10	No	Moderate	Resistant
OR<1	NA	Unconfirmed	Resistant

474 **Supplementary Table 2 – Definitions of quality metrics used**

Metric	Definition
FDR	$\frac{False\ Positive}{True\ Positive + False\ Positive}$
PPV %	$\frac{True\ Positive}{True\ Positive + False\ Positive}\ x\ 100$
NPV %	$\frac{True\ Negative}{True\ Negative + False\ Negative}\ x\ 100$
Sensitivity %	$\frac{True\ Positive}{True\ Positive + False\ Negative}\ x\ 100$
Specificity %	$\frac{True\ Negative}{True\ Negative + False\ Positive}\ x\ 100$
Accuracy %	$\frac{True\ Positive + True\ Negative}{True\ Positive +\ True\ Negative + False\ Negative + False\ Positive}\ x\ 100$

475

476

Supplementary Table 3 – Discordant Phylogentic lineage

Phylogenetic lineages reported by tbtAMR were compared to those published, with only 1 of the 8 discordant results not being due the detection of mixed lineages by tbtAMR.

Accession	tbtAMR lineage	Published lineage[16]
ERR2513557	lineage1;lineage4	lineage1
ERR067636	lineage2;lineage4	lineage4
SRR6339653	lineage4	lineage1
ERR2510523	lineage3;lineage4	lineage3
ERR067732	lineage2;lineage4	lineage4
ERR2515255	lineage2;lineage4	lineage2
ERR2512421	lineage3;lineage4	lineage3

References

1 Bennett J, Dolin R, Blaser MJ. Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases, 9th edn. .

2 TB guidelines. https://www.who.int/publications/digital/global-tuberculosis-report-2021/featured-topics/tb-guidelines (accessed March 5, 2022).

3 Carrol K, Pfaller M, Landry M. Manual of Clinical Microbiology, 12th edn. Wiley, 2019.

4 Donnan EJ, Marais BJ, Coulter C, *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 (2018)* 2023; **47**. DOI:10.33321/cdi.2023.47.8.

5 Hunt M, Bradley P, Lapierre SG, *et al.* Antibiotic resistance prediction for Mycobacterium tuberculosis from genome sequence data with Mykrobe. *Wellcome Open Res* 2019; **4**: 191.

6 Phelan JE, O'Sullivan DM, Machado D, *et al.* Integrating informatics tools and portable sequencing technology for rapid detection of resistance to anti-tuberculous drugs. *Genome Medicine* 2019; **11**: 41.

7 ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads | Microbiology Society. https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000131 (accessed July 4, 2023).

8 Walker TM, Miotto P, Köser CU, *et al.* The 2021 WHO catalogue of Mycobacterium tuberculosis complex mutations associated with drug resistance: a genotypic analysis. *The Lancet Microbe* 2022; **3**: e265–73.

9 14:00-17:00. ISO 15189:2022. ISO. https://www.iso.org/standard/76677.html (accessed July 4, 2023).

10 TBProfiler/tbprofiler at master · jodyphelan/TBProfiler. GitHub. https://github.com/jodyphelan/TBProfiler (accessed Oct 24, 2022).

11 Mykrobe. 2022; published online March 1. https://github.com/Mykrobe-tools/mykrobe (accessed March 5, 2022).

12 Huang W, Li L, Myers JR, Marth GT. ART: a next-generation sequencing read simulator. *Bioinformatics* 2012; **28**: 593–4.

13 McTavish EJ. TreeToReads. 2022; published online Sept 14. https://github.com/snacktavish/TreeToReads (accessed Nov 16, 2022).

14 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.

15 Sukumaran J, Holder MT. DendroPy: a Python library for phylogenetic computing. *Bioinformatics* 2010; **26**: 1569–71.

16 Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. *New England Journal of Medicine* 2018; **379**: 1403–15.

17 Ezewudo M, Borens A, Chiner-Oms Á, *et al.* Integrating standardized whole genome sequence analysis with a global Mycobacterium tuberculosis antibiotic resistance knowledgebase. *Sci Rep* 2018; **8**: 15382.

18 Coll F, Phelan J, Hill-Cawthorne GA, *et al.* Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis. *Nat Genet* 2018; **50**: 307–16.

19 Coeck N, de Jong BC, Diels M, *et al.* Correlation of different phenotypic drug susceptibility testing methods for four fluoroquinolones in Mycobacterium tuberculosis. *J Antimicrob Chemother* 2016; **71**: 1233–40.

20 Bradley P, Gordon NC, Walker TM, *et al.* Rapid antibiotic-resistance predictions from genome sequence data for Staphylococcus aureus and Mycobacterium tuberculosis. *Nat Commun* 2015; **6**: 10063.

21 WHO Consolidated Guidelines on Tuberculosis, Module 4: Treatment - Drug-Resistant Tuberculosis Treatment. https://www.who.int/publications-detail-redirect/9789240007048 (accessed March 5, 2022).

22 Votintseva AA, Pankhurst LJ, Anson LW, *et al.* Mycobacterial DNA Extraction for Whole-Genome Sequencing from Early Positive Liquid (MGIT) Cultures. *J Clin Microbiol* 2015; **53**: 1137–43.

23 World Health Organization. Technical manual for drug susceptibility testing of medicines used in the treatment of tuberculosis. Geneva: World Health Organization, 2018 https://iris.who.int/handle/10665/275469.