

Mycobacteria Amplicon Sequencing Tool: automated resistance prediction and lineage classification for *Mycobacterium tuberculosis*

Idowu B. Olawoye,¹ Maxim Fedorov,¹ Robert A. Petit, III,² Jennifer L. Guthrie^{1,3}

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT The Mycobacteria Amplicon Sequencing Tool (<https://github.com/guthrielab/MAST>) is a modular Nextflow pipeline for antimicrobial resistance prediction and lineage classification of *Mycobacterium tuberculosis* from amplicon or whole-genome sequencing data sets. The workflow automates read processing, variant calling, and annotation to produce standardized, human-readable reports in DOCX and TSV format.

KEYWORDS bioinformatics, software, mycobacteria, antimicrobial resistance

Mycobacterium tuberculosis, the causative agent of tuberculosis, is responsible for an estimated 1.2 million deaths globally each year (1). Given its global burden and rising antimicrobial resistance (AMR), *M. tuberculosis* is routinely monitored through public health surveillance, with genome sequencing increasingly used to predict resistance and characterize circulating lineages. These data provide insight into phylogeographic distribution, transmissibility, and resistance patterns (2–4). In addition, targeted amplicon sequencing has recently been used to rapidly detect resistance-associated variants (5).

To support public health laboratories and researchers conducting targeted amplicon sequencing of *M. tuberculosis*, we have developed Mycobacteria Amplicon Sequencing Tool (MAST), a Nextflow-based pipeline that parallelizes the analysis of whole-genome and targeted amplicon sequencing data (6). The pipeline detects sequence variants and compares them against a catalog of 11,390 high-confidence variants and their associated drug-resistance phenotypes curated by the World Health Organization (2nd edition) (7, 8). It also applies a lineage barcode schema to assign isolates to 24 major lineages and sub-lineages, enabling standardized lineage identification alongside AMR prediction (3).

MAST is an open-source bioinformatics pipeline for the analysis of *M. tuberculosis* amplicon sequencing data and is available on GitHub. It uses conda dependencies such as Filtlong (9) for quality trimming, BWA-MEM (10), Pysam (11), and SAMtools (12) for read alignment and BAM manipulation, Freebayes (13) for variant calling, and a custom Python script to compare mutations against the drug-resistance and lineage databases as seen in Fig. 1. Acceptable input files include one or multiple single-end FASTQ files (Illumina or Nanopore sequence data) and a CSV file containing patient information; these inputs generate two output files:

1. A user-friendly MS Word (.docx) report summarizing the predicted drug-resistance profile, lineage assignment, and selected patient information provided by the user.
2. A tab-delimited summary file (.tsv) suitable for downstream parsing, integrating with other tools, or storage in a database.

We evaluated MAST against the widely used tool TB-Profiler (14) version 6.6.5, using 168 publicly available *M. tuberculosis* genomes selected from NCBI, restricting

Editor Patrick D. Schloss, University of Michigan, Ann Arbor, Michigan, USA

Address correspondence to Jennifer L. Guthrie, jennifer.guthrie@uwo.ca.

The authors declare no conflict of interest.

See the funding table on p. 3.

Editor's Note: This article was handled by Pat Schloss, who acted as a Guest Editor in consultation with the Editor in Chief, Irene Newton. Dr. Schloss's role as the Guest Editor of this article is outside of his duties as the chair of the Journals Committee.

Received 22 January 2026

Accepted 24 March 2026

Published 27 April 2026

Copyright © 2026 Olawoye et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

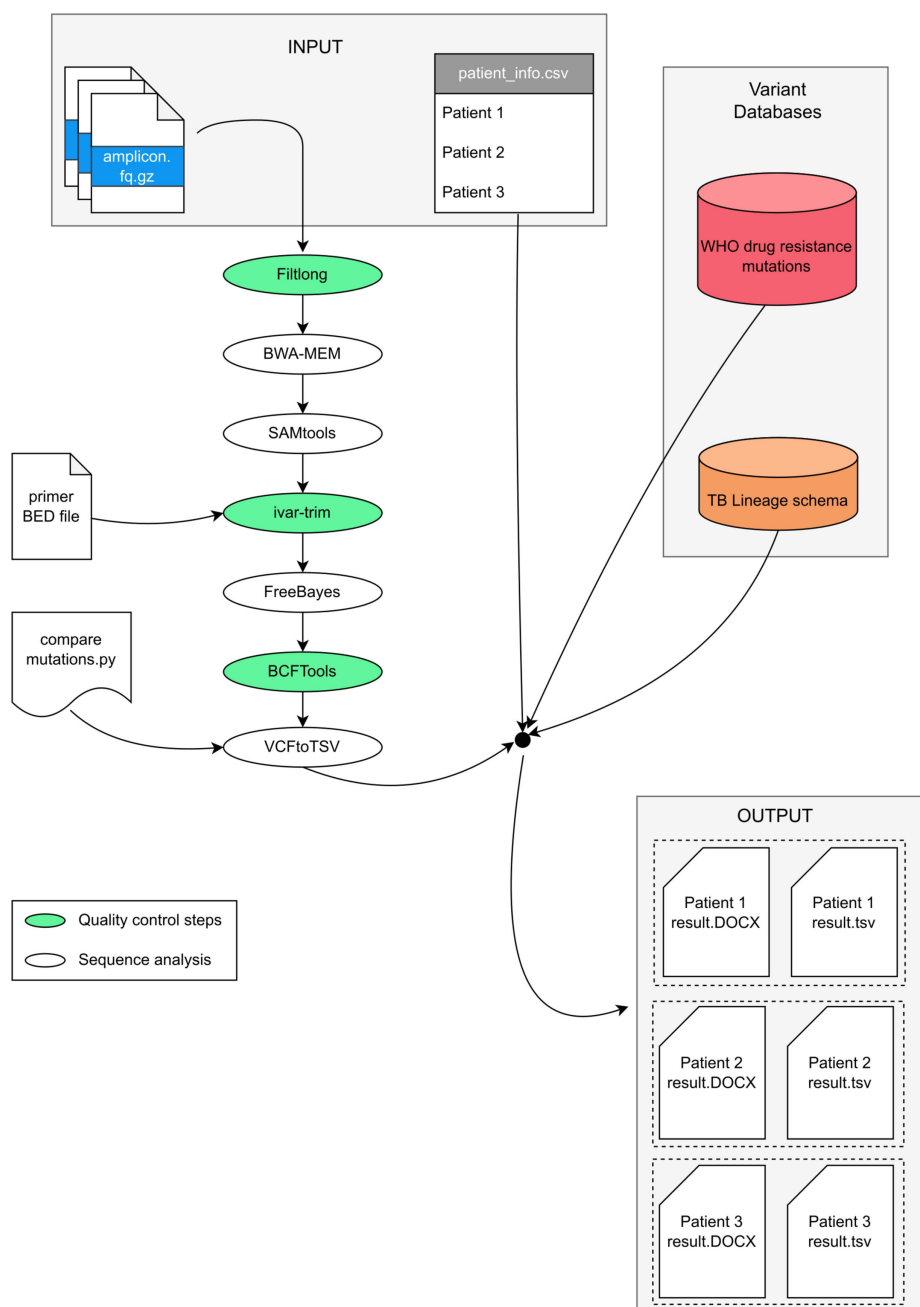


FIG 1 Schematic overview of the MAST workflow.

to genomes with an average sequencing depth of at least 30× and including isolates spanning major global lineages with both drug-susceptible and resistant profiles. Lineage predictions and AMR outputs produced by MAST were compared with TB-Profiler using default parameters. Lineage assignments were identical for all genomes (100% concordance). AMR predictions showed 99.7% concordance, with 350 of 351 drug-specific resistance calls matching those reported by TB-Profiler. The single discordant call corresponded to a variant that fell below MAST's reporting threshold due to filtering criteria, reflecting a difference in variant-handling logic rather than an incomplete mutation list (Table 1).

A key advantage of MAST is its ability to process both targeted amplicon and whole-genome sequencing data, while remaining accessible to users with diverse levels

TABLE 1 Comparison of TB-Profler and MAST demonstrating concordance in lineage assignment and antimicrobial resistance prediction

Category	MAST	TB-Profler	Concordance %
Lineage assignment	168	168	100
AMR prediction	350	351	99.7

of bioinformatics expertise. To support reproducibility and practical utility, we engaged domain experts to review the workflow and its applicability in routine settings. The pipeline will be maintained, supported, and openly available to the community.

ACKNOWLEDGMENTS

This work was supported by Western University's Frugal Biomedical Innovations Catalyst Grant (to I.B.O. and J.L.G.) and funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant Program (to J.L.G.) and an NSERC Undergraduate Student Research Award to M.F.

AUTHOR AFFILIATIONS

¹Department of Microbiology & Immunology, University of Western Ontario, London, Ontario, Canada

²Wyoming Department of Health, Wyoming Public Health Laboratory, Cheyenne, Wyoming, USA

³Public Health Ontario, Toronto, Ontario, Canada

AUTHOR ORCIDs

Robert A. Petit, III  <http://orcid.org/0000-0002-1350-9426>

Jennifer L. Guthrie  <http://orcid.org/0000-0001-8565-203X>

FUNDING

Funder	Grant(s)	Author(s)
Natural Sciences and Engineering Research Council of Canada		Jennifer L Guthrie Maxim Fedorov
Western University		Jennifer L Guthrie Idowu B. Olowoye

AUTHOR CONTRIBUTIONS

Idowu B. Olowoye, Conceptualization, Formal analysis, Methodology, Software, Validation, Writing – original draft, Writing – review and editing | Maxim Fedorov, Methodology, Software, Writing – review and editing | Robert A. Petit, III, Methodology, Validation, Writing – review and editing | Jennifer L. Guthrie, Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review and editing

DATA AVAILABILITY

MAST is available on GitHub under the GNU General Public License version 3.0 at <https://github.com/guthrielab/MAST>. An example data set to test the pipeline is available on Zenodo (DOI: <https://doi.org/10.5281/zenodo.19041451>). The workflow itself is archived on Zenodo and accessible via DOI: <https://doi.org/10.5281/zenodo.19166350>.

REFERENCES

- World Health Organization. 2026. Global tuberculosis report. World Health Organization. <https://www.who.int/publications/i/item/9789240101531>.
- Meehan CJ, Goig GA, Kohl TA, Verboven L, Dippenaar A, Ezewudo M, Farhat MR, Guthrie JL, Laukens K, Miotto P, et al. 2019. Whole genome sequencing of *Mycobacterium tuberculosis*: current standards and open

- issues. *Nat Rev Microbiol* 17:533–545. <https://doi.org/10.1038/s41579-019-0214-5>
3. Napier G, Campino S, Merid Y, Abebe M, Woldeamanuel Y, Aseffa A, Hibberd ML, Phelan J, Clark TG. 2020. Robust barcoding and identification of *Mycobacterium tuberculosis* lineages for epidemiological and clinical studies. *Genome Med* 12:114. <https://doi.org/10.1186/s13073-020-00817-3>
 4. Walker TM, Ip CLC, Harrell RH, Evans JT, Kapatai G, Dedicoat MJ, Eyre DW, Wilson DJ, Hawkey PM, Crook DW, Parkhill J, Harris D, Walker AS, Bowden R, Monk P, Smith EG, Peto TEA. 2013. Whole-genome sequencing to delineate *Mycobacterium tuberculosis* outbreaks: a retrospective observational study. *Lancet Infect Dis* 13:137–146. [https://doi.org/10.1016/S1473-3099\(12\)70277-3](https://doi.org/10.1016/S1473-3099(12)70277-3)
 5. Murphy SG, Smith C, Lapiere P, Shea J, Patel K, Halse TA, et al. 2023. *Direct Detection of Drug-Resistant Mycobacterium Tuberculosis Using Targeted next Generation Sequencing*. *Front Public Health*. Available from: <https://www.frontiersin.org/journals/public-health/articles/10.3389/fpuh.2023.1206056/full>
 6. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. 2017. Nextflow enables reproducible computational workflows. *Nat Biotechnol* 35:316–319. <https://doi.org/10.1038/nbt.3820>
 7. World Health Organization. 2023. Catalogue of mutations in *Mycobacterium tuberculosis* complex and their association with drug resistance. Available from: <https://www.who.int/publications/i/item/9789240082410>
 8. Walker TM, Miotto P, Köser CU, Fowler PW, Knaggs J, Iqbal Z, Hunt M, Chindelevitch L, Farhat M, Cirillo DM, et al. 2022. The 2021 WHO catalogue of *Mycobacterium tuberculosis* complex mutations associated with drug resistance: a genotypic analysis. *Lancet Microbe* 3:e265–e273. [https://doi.org/10.1016/S2666-5247\(21\)00301-3](https://doi.org/10.1016/S2666-5247(21)00301-3)
 9. Wick R. 2025. Rrwick/Filtlong. Available from: <https://github.com/rrwick/Filtlong>
 10. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*
 11. pysam-developers. 2025. Pysam-developers/pysam. Available from: <https://github.com/pysam-developers/pysam>
 12. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The sequence alignment/map format and samtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
 13. Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. *arXiv*
 14. Verboven L, Phelan J, Heupink TH, Van Rie A. 2022. TBProfiler for automated calling of the association with drug resistance of variants in *Mycobacterium tuberculosis*. *PLoS One* 17:e0279644. <https://doi.org/10.1371/journal.pone.0279644>