Journal of Clinical Microbiology

| Epidemiology | Full-Length Text

A tiled amplicon protocol for culture-free whole-genome sequencing of *M. tuberculosis* from clinical specimens

Chaney C. Kalinich,[1] Freddy L. Gonzalez,[2] Alice Osmaston,[3,4] Mallery I. Breban,[1] Isabel Distefano,[1] Candy Leon,[4] Jorge Coronel,[4] Grace Tan,[3] Valeriu Crudu,[5] Nelly Ciobanu,[5] Alexandru Codreanu,[5] Walter Solano,[4] Jimena Ráez,[4] Patricia Sheen,[4] Mirko Zimic,[4] Orchid M. Allicock,[1,6] Chrispin Chaguza,[1] Anne L. Wyllie,[1] Matthew Brandt,[1] Daniel M. Weinberger,[1,6,7] Benjamin Sobkowiak,[3,7] Ted Cohen,[1,7] Louis Grandjean,[3,4] Nathan D. Grubaugh,[1,2,6,7] Seth N. Redmond[1,6,7]

AUTHOR AFFILIATIONS See affiliation list on p. 12.

ABSTRACT Whole-genome sequencing of *Mycobacterium tuberculosis* can be a valuable tool for TB surveillance and treatment, providing insights into transmission patterns and comprehensive drug susceptibility testing. However, the slow growth of *M. tuberculosis* means traditional culture-based sequencing methods can take weeks to return results, which has limited the widespread adoption of these techniques and limited their use in clinical decision-making. Tiled amplicon sequencing is a fast, reliable, and cost-effective method of whole-genome sequencing that can be done directly on clinical specimens and has been implemented at scale in academic and public health laboratories across the world; it was the cornerstone of SARS-CoV-2 sequencing and has been adapted for a wide range of viral pathogens. However, similar methods are not yet available for far larger bacterial genomes. Extending this approach to *M. tuberculosis* would significantly reduce the cost, labor, and turnaround time for whole-genome sequencing. We designed a tiled amplicon panel consisting of 5,128 primers that covers the entire *M. tuberculosis* genome, the largest tiled amplicon sequencing panel we are aware of to date. Applying our amplicon panels to clinical samples of sputum, we show the ability to recover whole-genome bacterial sequences without the need for culture. The resulting sequence data can be used to determine *M. tuberculosis* lineage and reliably identify markers of drug resistance. Using this approach in clinical settings could reduce the time needed for comprehensive drug susceptibility testing from weeks to days and enable genomic epidemiology to be performed at scale, even in resource-limited settings.

IMPORTANCE We have developed and tested an amplicon panel, TB-seq, for the priority pathogen *Mycobacterium tuberculosis*, demonstrating recovery of near-full genomes directly from patient sputum, including mixed and low-concentration samples. This approach significantly reduces the turnaround time for this slow-growing bacterium while maintaining high accuracy in detecting clinically relevant mutations, including those associated with drug resistance. Given the global burden of tuberculosis and the critical need for faster diagnostic solutions, we believe our method has the potential to improve clinical decision-making and public health strategies.

KEYWORDS genomic epidemiology, amplicon sequencing, pathogen genomics, *Mycobacterium tuberculosis*

Tuberculosis (TB) is the world's leading cause of death from a single infectious agent, resulting in more than a million deaths per year (1). COVID-19-related disruptions in disease detection and treatment were responsible for approximately 700,000 excess TB deaths from 2020 to 2023 (1), and worldwide cases have continued to increase over the past 2 years (2).

Editor Christine Y. Turenne, University of Manitoba, Winnipeg, Manitoba, Canada

Address correspondence to Seth N. Redmond, seth.redmond@yale.edu.

Chaney C. Kalinich and Freddy L. Gonzalez contributed equally to this article. Author order was determined based on seniority.

The authors declare no conflict of interest.

See the funding table on p. 13.

Received 10 December 2025
Accepted 17 December 2025
Published 9 February 2026

However, the pandemic also saw impressive advances in the use of viral whole-genome sequencing (WGS) to identify viral lineages (3) and track disease spread on both global (4) and local scales (5), enabling real improvements in disease control (6).

Applying similar approaches to TB could make up for recent lost ground. Indeed, the application of WGS to inform TB control is far from a new idea, having been shown more than a decade ago to be able to reconstruct fine-scale transmission networks (7), and shows particular value in detecting superspreading (8), distinguishing recrudescence from reinfection (9), and has been used to characterize transmission dynamics at a national scale (10, 11). Perhaps more importantly, given the complexity of TB treatment regimes and the high global prevalence of drug resistance, WGS of *Mycobacterium tuberculosis* offers unique advantages for detecting antimicrobial resistance (12, 13), enabling tailored treatment regimens (14). Genotypic predictions of TB drug resistance have been shown to be highly accurate (15, 16).

While routine WGS of *M. tuberculosis* has been adopted for national TB surveillance in the UK, Italy, and the USA, it has not been implemented at scale outside of high-income countries (17, 18). This is, in large part, due to the challenges of reliably sequencing patient specimens: direct sequencing from patient sputum does not consistently produce sufficient data for resistance prediction or epidemiologic investigation (19, 20). Enrichment approaches such as hybrid capture have been developed to enable direct sequencing of *M. tuberculosis* from sputum (21), though high library preparation costs and a laborious protocol have prevented them from being widely adopted.

Reliable WGS can be achieved with the addition of a prior culturing step, but imposes delays of up to 6 weeks (22) and significantly increases the risk of airborne contagion: while sputum samples, having high viscosity and low bacterial load, can be handled in many local health centers, bacterial culture is typically handled in regional reference laboratories (23). The culturing process itself may also introduce discrepancies, with loss of diversity following culturing having been reported in some (24, 25), but not all, cases (26), a finding that may impair detection of heteroresistance.

Rapid drug susceptibility can be achieved by targeted assays such as nucleic acid amplification tests (NAATs) or line-probe assays (LPAs) within 1–2 days, but these are limited in the number of loci they can cover. NAATs work well for common rifampicin resistance but show poor sensitivity for second-line drugs (27), while LPAs offer high sensitivity to both first- and second-line drugs (28, 29) but perform poorly where secondary resistance mechanisms are at high prevalence (30). The World Health Organization (WHO) continues to recommend additional phenotyping via liquid culture assay for suspected extremely drug-resistant (XDR) samples (31). None of these approaches to drug susceptibility testing enable the bacterial lineage to be identified or transmission networks to be inferred.

While genotyping approaches such as genotyping restriction fragment length polymorphisms and variable number tandem repeats can identify TB lineages, both are hampered by the low genomic diversity of *M. tuberculosis*, meaning they are insufficient to infer transmission in many circumstances (22). In recent years, the WHO has also recommended assays that use a targeted amplicon approach to sequence resistance-associated loci for first- and second-line drugs (32). Encouragingly, these have been successfully implemented in resource-limited settings (33); however, the targeted approach is not suitable for assaying resistance markers outside of the targeted regions or for the discovery of novel resistance loci. In comparison, WGS is able to simultaneously assay all 450 SNPs that have shown association or preliminary association with drug resistance (16), making it particularly well suited for tailoring treatment regimens for highly drug-resistant TB (34).

The dramatic scale-up in WGS of viral pathogens was achieved using tiled amplicon sequencing. Initially developed for genomic surveillance during the 2016 Zika epidemic (35), where low viremia had precluded direct sequencing of clinical samples even where the infection had been confirmed, tiled amplicon sequencing uses multiplex PCR of tiled overlapping regions of a target genome to recover whole genomes from samples of low

concentration or complex microbial communities. This has proven particularly useful for sequencing remnant samples from diagnostic tests. Its use in the "ARTIC" protocol for sequencing SARS-CoV-2 (36) has led to its deployment in thousands of public health laboratories around the world, facilitating true global surveillance of viral dynamics (6). The ease, reliability, and low cost of amplicon sequencing have seen its adaptation to a broad range of viral pathogens both in respiratory disease (37, 38) and beyond (39, 40). Adapting these techniques to bacterial pathogens could allow rapid culture-free sequencing from low input volumes, at large scale, and with minimal infrastructure requirements.

We present here TB-seq, a tiled amplicon panel for *M. tuberculosis* and the first use of amplicon-based WGS for a bacterial pathogen. This panel is able to generate complete genome coverage from samples with minimal input concentrations without the need for bacterial culturing. We show that TB-seq can be used to identify clinically relevant phenotypes, such as antimicrobial susceptibility, within days of sample collection and can detect resistance loci that were not found by rapid diagnostics. We anticipate that this work will not only generate opportunities for genomic epidemiology of *M. tuberculosis* but will also provide a roadmap for the development of amplicon sequencing for other clinically important bacterial pathogens.

RESULTS

Amplicon sequencing enables recovery of *M. tuberculosis* whole genomes from clinical specimens without prior culturing

The TB-seq panel consists of 2,564 primer pairs generating two pools of 2 kb amplicons that are combined to achieve whole or near-whole coverage of the 4.4 Mb *M. tuberculosis* genome. Sequencing libraries were prepared using a standard amplicon sequencing workflow, with and without adding amplicon panel primers to allow comparison of amplified and unamplified DNA libraries. This approach was applied to cultured isolates, sourced from Moldova, and sputum samples from patients in Peru (Table S1).

We assessed the limits of detection for each amplicon panel by sequencing serial dilutions of six cultured isolates using both amplified and unamplified sequencing approaches. For *M. tuberculosis*, high genome coverage (>95%) was achieved in all amplified samples above 100 genome equivalents per microliter (GE/µL), compared to 10,000 GE/µL for unamplified samples (Fig. 1; Table S2).

Conversion of GE/µL in extracted DNA to sputum concentration is imprecise, with the concentration dependent on sample quality and extraction efficiency. For a specimen with 10^4 AFB/mL (at which concentration 60% of samples are smear-positive [41]), 1 mL of sputum eluted into 50 µL (as used in our extraction protocols) would result in a DNA concentration of $10^4 \div 50 = 2e^3$ GE/µL at 100% efficiency of extraction; we have used a $1e^3$ GE/µL (i.e., 50% efficiency) as our lower limit of extracted DNA concentration from smear-positive sputum. At this concentration, all of our dilution series achieved over 75% genome coverage using TB-seq (Fig. 1).

DNA was extracted from 60 sputum specimens with a range of acid-fast bacilli semi-quantitative measurements (e.g., 1+ to 3+). Several different DNA extraction methods were tested during optimization (methods A–E: detailed in Table S1, Figure S1b and c and Appendix S1). For the 10 sputum samples extracted with our optimized protocol (E), *M. tuberculosis*-positive sputum samples demonstrated dramatic increases in coverage for amplified samples when compared to their unamplified replicates (Fig. 1a through c). While only 2/10 of unamplified samples achieved more than 75% genome coverage, 9/10 of the amplified samples sequenced above this threshold, with 7 of those generating more than 95% coverage. The remaining sample achieved 33% coverage amplified and negligible coverage unamplified. Metagenomic sequencing indicated successful amplification from samples containing both commensal and pathogenic bacteria including *Streptococcus, Pseudomonas, Actinomycetes,* and *Schaalia* spp.

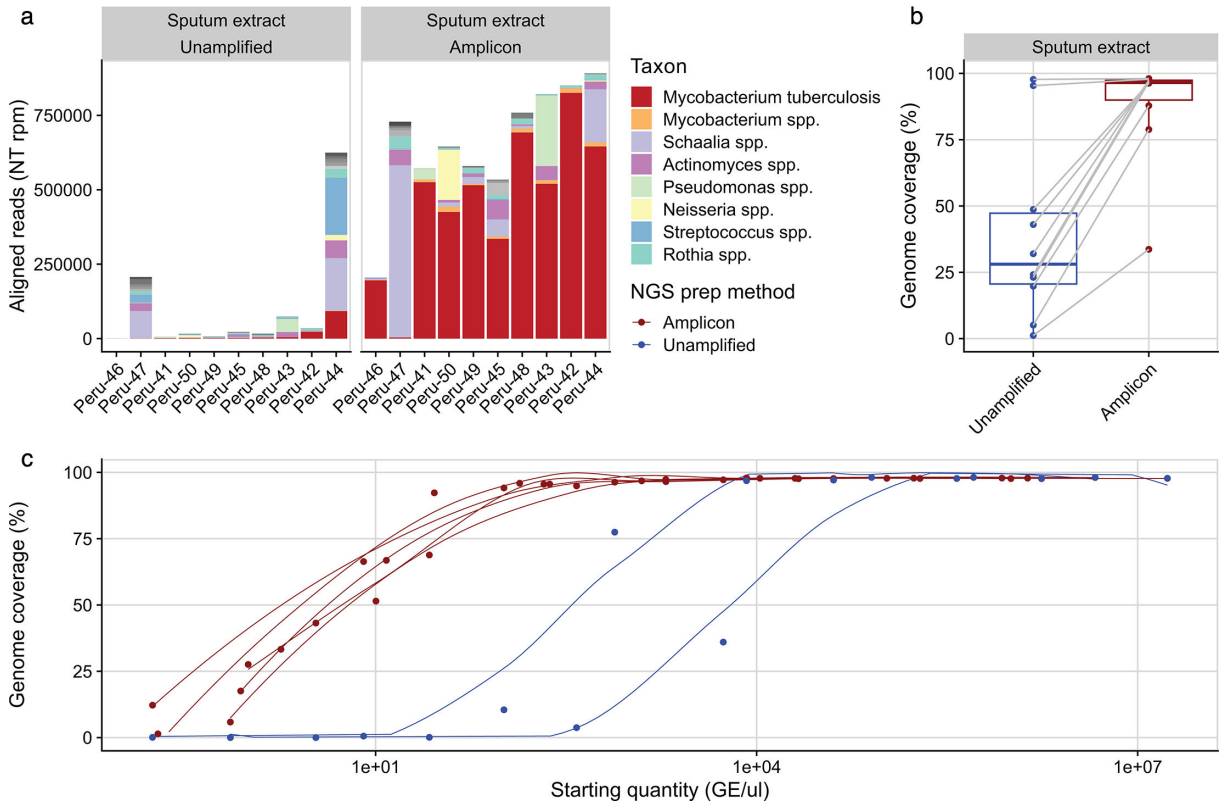

FIG 1 Tiled amplicon sequencing enables culture-free recovery of whole-genome sequences from *Mycobacterium tuberculosis* sputum specimens. We compared sequencing results for unamplified and amplified *M. tuberculosis* clinical specimens with regard to (a) microbial content via the CZID metagenomics pipeline and (b) average genome coverage. *M. tuberculosis*-positive sputum samples sequenced directly showed dramatic increases in genome coverage, with 8/10 samples generating more than 80% coverage after amplification with our protocol, and a ninth sample generating 78% coverage despite a significant infection with *Schaalia odontolytica*. Serial dilutions of DNA from cultured isolates (c) demonstrated that the amplicon scheme enables recovery of genomes with >95% coverage from all samples with ≥100 GE/µL, compared to ≥10,000 GE/µL without amplification.

The protocol is extremely cost-effective: while library preparation costs vary between sites, the per-sample cost for amplicon generation is just $19.07 per sample (Table S3).

Direct sputum sequencing enables phylogenetic classification and detects markers of antimicrobial resistance

M. tuberculosis lineages were called with the Mykrobe package (v.0.13.0) (42), which assigned all samples to lineage 2 (sublineage 2.2.1) or 4. Mykrobe performed equally well when calling lineages for all high-coverage (>75%) samples, regardless of whether they were derived from cell culture or sputum. We constructed maximum-likelihood (ML) phylogenies using IQ-TREE (43) including all sequenced specimens and the broad reference set of *M. tuberculosis* sequences used for primer design (Fig. S2 and Appendix S2). In all cases, the primary lineage predicted by Mykrobe aligned with lineages from an ML tree, though in some cases, secondary lineages were predicted based on minor variants that did not concord with the ML tree.

Sequencing data were of sufficiently high quality to produce a drug susceptibility prediction for all template dilutions from cultured isolates with at least 10 GE/µL starting quantity (Fig. 2a; Table S4). Seven samples had been previously phenotyped at diagnostic centers in Moldova, via growth on liquid media (BD BACTEC mycobacterial growth indicator tubes [MGIT]) or solid media (Löwenstein–Jensen) using approved protocols (44). A further sample was assayed for rifampicin resistance using GeneXpert rapid testing (Xpert MTB/Rif, Cepheid, Sunnyvale, CA, USA), and two of our samples were also sequenced without amplification, and drug susceptibility predicted. In the

FIG 2 Amplicon sequencing predicts TB antimicrobial resistance *in silico*. Drug susceptibility to 15 anti-TB drugs was predicted by Mykrobe on amplicon-sequenced data for 8 colony samples (at 0.01 template dilution) and 10 sputum samples using our optimized extraction protocol. Colony samples (a) showed strong concordance with prior phenotyping on both liquid and solid media, enabling detection of both MDR and XDR samples. Sputum-derived samples (b) were able to generate susceptibility predictions for 8/10 samples, including at least one predicted MDR isolate (Peru-48).

overwhelming majority of cases (42/47) where a resistance phenotype was available, our susceptibility predictions were consistent with phenotyping.

Four discordant results were found: we failed to predict resistance for ethambutol and streptomycin in sample 444-2588-18 and failed to predict resistance to streptomycin in sample 444-3403-18, despite growth on medium containing these drugs. Resistance to moxifloxacin was predicted for sample 111-2565-18 despite phenotypic susceptibility. We also identified heteroresistance for three first-line drugs in one sample (444-3403-18) was found to be phenotypically fully resistant; the effects of growth in drug-free medium during maintenance cannot be determined.

When comparing unamplified sequencing and TB-seq on these samples, all resistance predictions were identical. While coverage in resistance-associated genes was typically extremely high (Table S4), lower coverage was seen in all samples in the rpsL gene associated with streptomycin resistance, suggesting discordant predictions may derive from poor sequencing coverage; further modifications to the amplicon panel may be necessary to reliably predict resistance for streptomycin.

Predictions were internally consistent for all template dilutions above 100 GE/µL (with some variability between samples called as heteroresistant or with fixed resistance alleles), with the exception of streptomycin, suggesting the potential for robust susceptibility calling using TB-seq at lower bacterial loads. While a comprehensive assessment of the sensitivity of resistance prediction is beyond the scope of this study, our results do suggest that TB-seq may be useful for rapid susceptibility prediction in public health settings.

Of the 60 extracted sputum specimens with all protocols, sequencing data were robust enough to produce a drug susceptibility prediction for 53/60 sputum specimens (Fig. S1a). Of the seven specimens which failed, four (Yale-TB121, Yale-TB123, Yale-TB149, Yale-TB150) had starting quantities following extraction below 10 GE/µL. None of the other three (Yale-TB126, Yale-TB139, Yale-TB148) had GeneXpert results available for comparison. Eight out of 10 specimens that were extracted with the best-performing protocol (Protocol E; see Appendix S1), which included a NALC-NaOH treatment to deplete non-mycobacterial DNA, had adequate data to predict resistance (Fig. 2b). At least one sample was predicted to be resistant to isoniazid, rifampicin, and pyrazinamide and would be classed as MDR-TB. We do not have access to phenotypic susceptibility results for those sputum samples; however, genotypic-based predictions of drug resistance in *M. tuberculosis* are usually highly accurate for most drug classes (15).

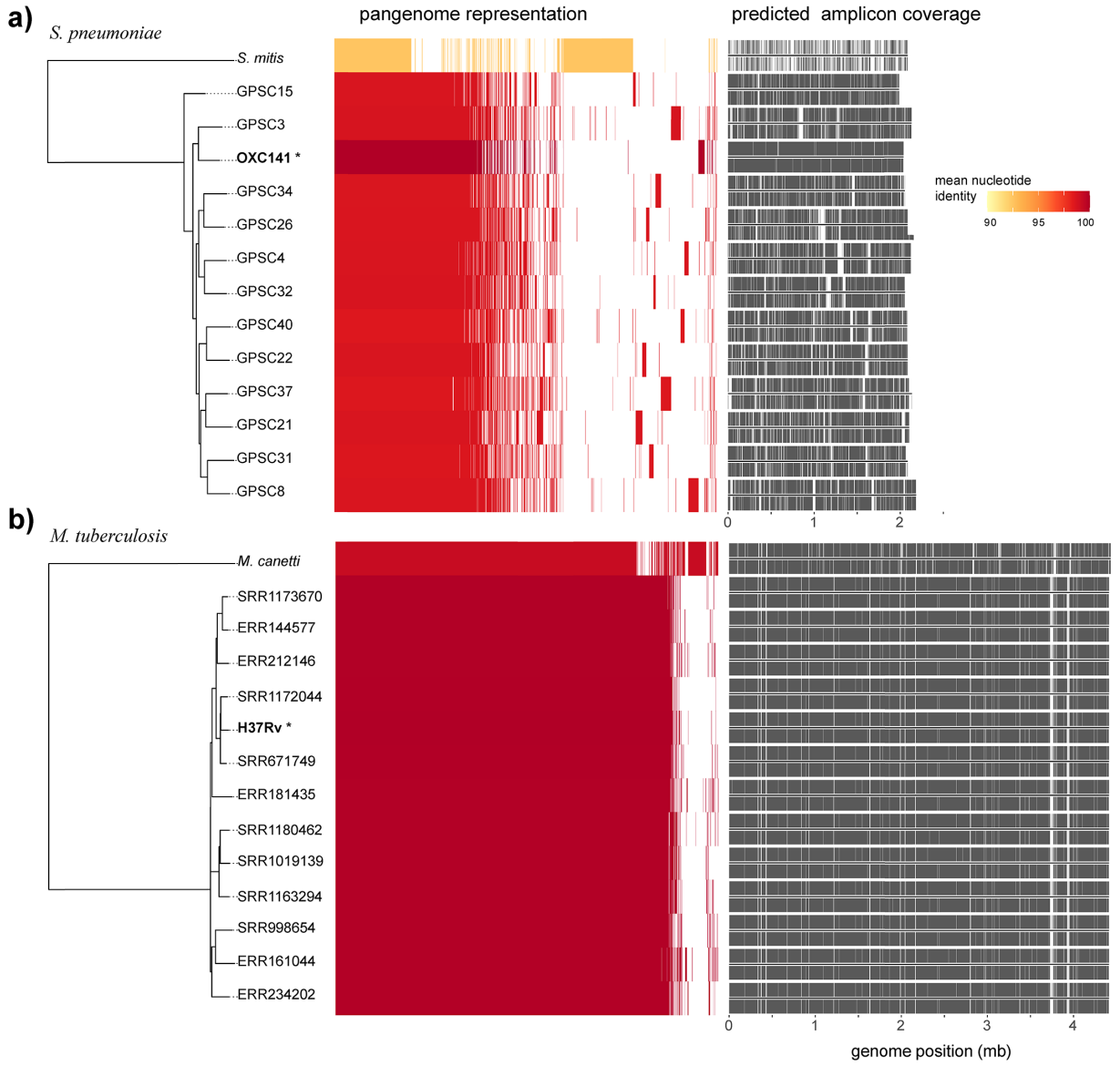

FIG 3 *In silico* modeling indicates broad applicability across diverse *Streptococcus pneumoniae* serotypes and *Mycobacterium tuberculosis* lineages. Pangenome representation of (a) *S. pneumoniae* whole-genome sequences (*n* = 13) and *Streptococcus mitis* outgroup (accession: AP023349) and (b) *M. tuberculosis* whole-genome sequences (*n* = 13) and *Mycobacterium canettii* outgroup (accession: NC_019950). Starred phylogenetic tree tips mark the reference sequences used for primer design. Shaded bar graphs (middle) denote genes shared among clades; the yellow-red color scale denotes average nucleotide identity. Predicted amplicon coverage (right) is shown in gray, with forward and reverse amplicon pairs displayed above and below the line. *M. tuberculosis* has a higher proportion of genes shared between clades and higher ANI than the open pangenome of *S. pneumoniae*, driving the higher predicted amplicon coverage. A list of the sequences used in this analysis can be found in Table S2.

In silico predictions indicate applicability of amplicon schemes to large bacterial genomes

To explore the limits of amplicon sequencing and assess performance on open and closed bacterial genomes, we assessed genome coverage *in silico* against a set of bacterial assemblies representative of global diversity.

We designed a comparable amplicon schema for *Streptococcus pneumoniae*, a prolifically recombinogenic pathogen with an open genome, and compared performance for both panels *in silico* across a representative set of closely related genomes (Fig. 3; Fig. S3b). *S. pneumoniae* has an average core genome size of 2,160,000 bp (45),

requiring 2,292 primers for full coverage, while *M. tuberculosis*, with an average genome size of 4,411,000 bp (46), required 5,128 primers to amplify the entire genome.

For each bacterial species, we selected a small number of genetically diverse sequences from the larger set of publicly available sequences to predict coverage in related strains (Fig. 3; Table S5a). As expected, predicted amplicon coverage was highest against the strains used as a reference for panel design (*Sp*:OXC141: 98.93%; *Mt*/H37Rv: 94.31%). *M. tuberculosis* coverage is reduced due to the omission of PE/PPE regions from the design, which accounts for up to 8%–10% of its genome (12). However, predicted coverage in *M. tuberculosis* remained high across all members of the *M. tuberculosis* complex, including all seven major lineages, *Mycobacterium bovis* (≥94.23%), and *Mycobacterium canettii* (89.44%). In contrast, coverage of *S. pneumoniae* fell sharply between Global Pneumococcal Sequence Clusters (GPSCs) (<82.00%–>88.00%) and in the *Streptococcus mitis* outgroup (32.18%). The range of average nucleotide identity was narrower across *M. tuberculosis* (99.98%–99.27%) compared to the broader variation observed among *S. pneumoniae* (98.76%–92.51%), reflecting the diversity in the GPSC lineages included. Pangenome size and similarity were also markedly different between the two species, with *M. tuberculosis* having a smaller relative pangenome (4,335 total genes and a mean genome size of 4,067 genes, a ratio of 1.07, based on the 13 genomes included) than *S. pneumoniae* (3,942 total genes and a mean genome size of 2,071 genes, a ratio of 1.90, based on the 13 genomes included). As a result, far more sharing of genes with the reference strain was observed for *M. tuberculosis* (4,002–4,050 shared genes) than *S. pneumoniae* (1,705–1,793 shared genes). Our findings suggest that amplicon panels are tolerant to increasing genetic distance but negatively affected by genome rearrangements.

Amplicon predictions for expected coinfections and commensals indicated limited enrichment of off-target species (Fig. S4a and b; Table S5b).

DISCUSSION

Tiled amplicon sequencing of pathogens has proven extremely useful for reconstructing disease spread and gaining insight into transmission patterns for a variety of viruses (47). The COVID-19 pandemic stimulated a global effort to adopt these methods and use genomics to track and monitor SARS-CoV-2; however, they have not previously been applied to the significantly larger and often more complex genomes of bacteria. We report the use of a tiled amplicon panel to sequence *M. tuberculosis,* generating whole-genome coverage from specimens with minimal input DNA and demonstrating the ability to identify lineages and markers of drug resistance. This is the first application of this technique to sequence a pathogenic bacterium, and the use of TB-seq in public health settings could have a major impact on TB control.

As of 2023, with *M. tuberculosis* reclaiming the title of most deadly single pathogen (1), there is a renewed focus on improving TB treatment and control. In particular, there are six vaccine candidates currently in phase III clinical trials (48) and continuing efforts to develop shorter and more effective treatment regimens (49). This may be an ideal time to apply tools used to control and monitor interventions for COVID-19 to TB (1).

Antimicrobial resistance is a critical issue in treating and controlling TB, due to the prevalence of resistance to first-line drugs and the length, cost, and complexity of treatment regimens (31). Despite the introduction of shorter regimens, the time taken to find an effective treatment can be long, and incomplete treatment remains a problem (50). For this reason, the WHO now recommends the use of targeted sequence-based diagnostics for rapid drug susceptibility testing for patients who are at high risk of, or have already experienced, treatment failure (32). However, designing such an assay is not simple; more than 40 separate loci, each containing numerous individual mutations, have been implicated in drug resistance (51), and uncertainty can be higher for new or second-line drugs (52). WGS works around these limitations of targeted amplicon sequencing. Unlike targeted sequencing approaches, data are being generated across the entire genome; therefore, as new genetic markers of resistance are discovered, these

can be added bioinformatically without adding new primers. As per-sample primer costs are low (Table S3), the required time, infrastructure, and costs for tiled amplicon sequencing are almost identical to targeted amplicons; the additional data generated through WGS can be used along with phenotypic drug susceptibility to expand our understanding of the genetic markers of drug resistance, especially for third-line or novel drugs, increasing the accuracy of predictions over time (12).

WGS obviates the need to design a targeted assay and can also return resistance predictions within days of a positive culture. However, the requirement of most existing WGS approaches to first culture a sample means that the overall sample-to-sequence turnaround time for *M. tuberculosis* is measured in weeks or months (53), and significant biases can be introduced during the culturing process itself (12). WGS without culture has not consistently produced data of high enough quality for resistance prediction or thorough epidemiologic investigation (19, 20), or has been limited to specimens with a high bacterial load (20, 54); subsequent efforts to implement WGS have relied on expensive techniques such as hybrid capture (21, 55). For this reason, routine WGS still relies on prior culturing in MGIT to ensure reliable whole-genome recovery, both in public health settings (56–59) and ongoing trials for clinical application (60).

Our technique recovers whole or near-whole genomes directly from sputum, enabling lineage classification and providing genome coverage of many first- and second-line resistance loci, something that was not possible with prior WGS approaches (19–21, 53, 61). For a notoriously slow-growing organism such as *M. tuberculosis*, eliminating the culturing step reduces the time from sample collection to genome from weeks to days, meaning genomic epidemiology could be used in real-time to inform outbreak investigations (10, 62) and public health measures to reduce spread (11, 63).

Consistent with the *in silico* modeling prediction, TB-seq did not show higher rates of amplicon dropout when faced with targets that had drifted from the H37Rv reference strain. However, comparison with similar results for *S. pneumoniae* is instructive, with significant coverage gaps emerging even in clades closely related to the OXC141 reference strain. This suggests a weaker applicability of the tiling amplicon approach in species that undergo significant levels of recombination and horizontal gene transfer, and the ratio of genome to pangenome size is likely to be a key metric for our ability to design an amplicon panel. This ratio is highly sensitive to the diversity of habitats in which the pathogen is found: free-living or commensal species, such as *S. pneumoniae*, gain particularly large pangenomes to enable adaptation to diverse environments, while intracellular pathogens, such as *M. tuberculosis*, show strong purifying selection, low effective population sizes, and low genome-to-pangenome ratios (64). Other intracellular pathogens such as *Yersinia pestis*, *Listeria monocytogenes*, *Legionella pneumophila*, *Chlamydia trachomatis, Treponema pallidum, and Neisseria gonorrhoeae* may be suitable targets.

For bacteria with large pangenomes, designing tiling amplicon schemes will rely upon databases of previously sequenced genomes to determine circulating genetic diversity and to map horizontal exchange and genome rearrangements. The improving availability of vast databases of assembled genomes (65), as well as methods for bacterial genome graph construction (66) and recombination-aware graph aligners (67), suggests this could be a viable approach for high-priority pathogens.

Within our TB-seq panel, gaps remain in our coverage of the highly repetitive PE/PPE regions. While these are frequently omitted from *M. tuberculosis* analyses, increasing evidence of functions in host-cell invasion (68) and importance for vaccine design (69) suggests inclusion of these regions in future iterations of this amplicon panel would be an important improvement.

As all clinical specimens used were smear-positive, the usability of this technique with smear-negative (or paucibacillary) specimens was not determined. Approximately 60% of specimens with a bacterial load of 10^4 AFB/mL are smear-positive (41); while the precise limit of detection of smear microscopy is dependent on sample quality, quantity, and user experience, our estimated DNA yield for such samples (1e3 GE/µL)

is comfortably above the limit of applicability determined by serial dilution, suggesting that amplicon sequencing is likely to be successful with at least a subset of smear-negative specimens. We successfully recovered genomes classified as "low" by GeneXpert Ultra that are likely to consist of "scanty" smear-positive samples (70). Yet, with a limit of detection as low as 15 AFB/µL (71), GeneXpert Ultra will uncover smear-negative samples that cannot be sequenced with TB-seq, and direct sequencing of sputum is unlikely to be suitable as a diagnostic in the near future.

In the near term, our expected use case for this technology is to sequence confirmed, smear-positive cases for the purposes of genomic epidemiology. However, due to the importance of drug resistance to effective TB treatment, we wished to assess whether tiled amplicon sequencing could plausibly determine drug susceptibility. For the majority of our specimens (53/60), tiled amplicon sequencing from uncultured sputum was able to make accurate drug susceptibility predictions for all first- and second-line drugs (Fig. S1). Barriers to clinical application are necessarily higher (72): if diagnostics and resistance prediction are to be used to tailor treatment regimens, a comprehensive performance assessment will be necessary, including spike-ins with calibrated numbers of organisms and large numbers of drug-resistant samples. It will also be important to assess performance in a range of other likely scenarios: paucibacillary infections; mixed *M. tuberculosis* strains; mixed *M. tuberculosis* and nontuberculous mycobacteria; mixed resistant and susceptible strains. This would be a major undertaking: previous sensitivity assessments for sequence-based diagnostics have required many thousands of samples (73), and the use of sputum or live cultures would incur additional logistical challenges and significant biosafety risks.

Yet despite this complex landscape, prior experience suggests this technique could have a significant impact. The widespread use of tiled amplicon sequencing for pathogen genomics during the COVID-19 pandemic has ensured that this method is trusted, understood, and easily implemented in academic and public health laboratories worldwide. As the focus now turns to adapting this capacity to other public health threats (6), it is important to prioritize the development of tools for global priority pathogens that can be implemented in the regions suffering the greatest burden. The method presented here uses the same workflow as used by thousands of public health labs to sequence SARS-CoV-2, utilizing an off-the-shelf commercial sequencing library preparation kit with the principal change of swapping out primer pools used to generate amplicons. While the total cost of the primer pool is almost $10,000 at current prices, only a small amount is used per reaction, and the per-sample cost for library preparation is just $19.07 (Table S3). Genomic surveillance of *M. tuberculosis* has demonstrated capacity to guide TB interventions in high-income countries (8, 9); the reductions in cost and turnaround time afforded by tiled amplicon sequencing could enable this to be implemented in LMICs with high TB burden. Just four countries (India, Bangladesh, Indonesia, Democratic Republic of the Congo) account for over half of all TB deaths; all four have prior in-country experience with amplicon sequencing of SARS-CoV-2 (74–77), suggesting a ready capacity for tiled amplicon sequencing of *M. tuberculosis*. Extensive use of alternative sequencing methods such as Oxford Nanopore in these regions (74, 76, 78) suggests adaptation to cheaper and more portable sequencing platforms may further increase surveillance capacity.

If successfully implemented in a clinical environment, the capacity for culture-free *M. tuberculosis* sequencing to provide rapid lineage determination and resistance profiling could be transformative—not only by increasing rates of successful TB treatment at the patient level (79), but by preventing the further transmission of MDR-TB at the population level (80, 81). With several new TB vaccines in phase III trials, WGS can also provide lineage-specific estimates of vaccine efficacy and early signals of vaccine escape. Optimizing the use of available antimicrobials and vaccines, both old and new, is vital to improve TB control, and culture-free WGS can be used to ensure the right interventions are reaching the right populations in a timely manner.

MATERIALS AND METHODS

TB-seq primer design

Raw reads for *M. tuberculosis* sequences were downloaded from a previously described globally representative data set (13) (n = 489) (Appendix S2) from the European Nucleotide Archive at the European Molecular Biology Laboratory-European Bioinformatics Institute. The reference genome (H37Rv; accession NC_000962.3) was obtained from the National Center for Biotechnology Information (NCBI) GenBank.

Variants were called against the reference using Snippy, and a time-resolved ML tree was built using our variant call file, along with sample data generated from Augur (v.22.4.0) (82), IQ-TREE (v.2.23) (83), and TreeTime (v.0.10.1) (84). Representative sequences (n = 6) were selected from across this tree using Parnas (v.0.1.4) (85), to cover >50% of the expected overall diversity. We used these representatives to create an *M. tuberculosis* core genome assembly using Snippy (v.4.6.0) (https://github.com/tseemann/snippy).

Tiled primer schemes (target amplicon size 2 kb) were designed for the full H37Rv reference genome (NCBI accession NC_000962.3) using PrimalScheme (v.1.4.1) (35). PE/PPE regions (conserved proline-glutamine/proline-proline-glutamine domains and hypervariable C-termini (86) and sites with known resistance-related polymorphisms were hard-masked to prevent primers from being designed at these loci. Primer sequences are included in Appendix S3.

SP-seq primer design

For comparison to TB-seq, an amplicon schema was designed for *S. pneumoniae*, which is highly recombinogenic and frequently reshuffles its genome via horizontal gene transfer. Because of the open nature of its pangenome, we targeted primers only in the core genomic regions shared across several strains (Appendix S2).

We downloaded all available *S. pneumoniae* serotype 3 contigs (n = 490) (Appendix S2) from the Global Pneumococcal Sequencing (GPS) database (87) as of 2 February 2023. These sequences were assembled into a reference-guided "metaconsensus sequence" based on the alignment of genomes from multiple strains mapped to the serotype 3 reference. PrimalScheme (35) was run on this metaconsensus sequence as described for TB-seq.

Primer pooling

Both primer sets were synthesized by Integrated DNA Technologies (Coralville, IA, USA), desalted, and in IDTE pH 8.0, at concentrations of 200 µM and 100 µM for *M. tuberculosis* and *S. pneumoniae*, respectively. Primer pairs in each set were divided into two pools, to separate overlapping tiled primers, and combined in equal proportions to create working primer solutions. These solutions were not diluted any further prior to use in the PCR reaction.

Clinical specimens

M. tuberculosis clinical samples consisted of DNA extracted from positive solid or liquid cultures from sputum and DNA extracted directly from sputum specimens. Extracts from culture consisted of remnant specimens from a study site in Moldova, where sputum specimens were tested at a number of diagnostic centers in Moldova by microscopy, GeneXpert MTB/Rif, and culture; positive cultures were sent to the National TB Reference Laboratory in Chisinau for extraction by the cetyltrimethylammonium bromide method, as described previously (11). Extracts from sputum consisted of specimens collected in Peru after routine diagnostics confirmed the presence of *M. tuberculosis*. In order to test the efficiency of different methods for extracting DNA from sputum, each specimen was split into two and processed with two different protocols. A total of 30 unique sputum specimens were processed with two protocols each, and a total of six different protocols

were tested. A full list of all *M. tuberculosis* samples and the extraction methods used can be found in Table S1, and a detailed description of extraction methods can be found in Appendix S1.

M. tuberculosis quantification

Following extraction, DNA was quantified using a *M. tuberculosis*-complex-specific RT-qPCR probe-based assay using a previously designed primer and probe set (88) and a custom-designed gBlock gene fragment for use as a positive control, all synthesized by Integrated DNA Technologies (Newark, NJ, USA). Lyophilized primers and probes were resuspended in nuclease-free water to achieve a working stock concentration of 10 µM. The lyophilized gBlock was resuspended in nuclease-free water following manufacturer's instructions (89), and serially diluted to generate a standard curve that was assayed alongside the clinical samples. All oligonucleotide sequences can be found in Table S6.

RT-qPCR was performed using the Luna Universal Probe One-Step RT-qPCR Kit (New England Biolabs, Ipswich, MA, USA), following the manufacturer's recommendations for reaction composition (NEB): 10 µL Luna Universal Probe qPCR Master Mix, 0.8 µL forward primer (10 µM), 0.8 µL reverse primer (10 µM), and 0.4 µL probe; with 3 µL nuclease-free water and 5 µL template DNA, for a total reaction volume of 20 µL. Each RT-qPCR plate included a negative template control and eight gBlock 10-fold serial dilutions with known copy numbers of the target DNA region from 1×10^7 to 1×10^0. All assays were run on a CFX 96 thermal cycler (Bio-Rad Laboratories, Hercules, CA, USA) and interpreted using the CFX Maestro Software for CFX Instruments (Bio-Rad). Genome copy number was calculated by the CFX Maestro Software using the Cq values for the 10-fold serial dilution standards to calculate reaction efficiency and generate a standard curve of Cq vs log-concentration. The standard curve was used to calculate genome copy numbers for experimental samples.

Amplicon sequencing

DNA libraries were prepared using the Illumina COVIDSeq DNA Prep Kit (Illumina, San Diego, CA, USA) with the included SARS-CoV-2 primers replaced with primer pools for *M. tuberculosis*, as previously described (90). Template DNA was amplified in two separate multiplex PCR reactions, with each primer pool containing non-overlapping sections of the tiled primer scheme. After amplification, PCR products were combined in equal parts and subjected to a 3 minute tagmentation. Tagmented amplicons were purified with a bead cleanup, followed by library amplification with IDT for Illumina Unique Dual Indexes (Illumina, San Diego, CA, USA). Individual indexed libraries were pooled together in equal proportions and underwent a double-ended bead cleanup, selecting for DNA fragments between 300 and 600 bp. The purified library pool was quantified using the 1× dsDNA High-Sensitivity Assay Kit on the Qubit 4 Fluorometer (Thermo Fisher, Waltham, MA, USA), and fragment distribution was verified using the dsDNA High-Sensitivity kit on the 2100 Agilent Bioanalyzer Instrument (Agilent, Santa Clara, CA, USA). Pooled libraries were sequenced in 150 bp paired-end reads at the Yale Center for Genome Analysis on an Illumina NovaSeq (Illumina, San Diego, CA, USA), with an average of 10 million reads per library.

Unamplified libraries were prepared as described above, with the primer pools in the first PCR reaction omitted and replaced with nuclease-free water.

Alignments and calling

Reads were aligned to the *M. tuberculosis* H37Rv reference using BWA-MEM (v.2.2.1) (91) and SAMtools (v.1.15.1) (92). Amplicon sequencing data were filtered (using defaults; Q > 20 over a sliding window of 4, minimum read length of 50% of the average length). *M. tuberculosis* primer sequences were trimmed using iVar (v.1.4.2) (93). Metagenomic sequences were trimmed and filtered for quality and length (<100 bp) using Trim Galore (v.0.6.10) (94). Variants were called and filtered (Phred score Q > 10 and read depth >10)

using BCFtools (v.1.21) (95). Read subsampling, depth, and coverage were calculated using SAMtools (92). Raw reads were directly submitted to the CZID mNGS Illumina pipeline (96) for microbial composition characterization within samples. Further data analyses and visualizations were carried out in RStudio (v.2024.04.2+764) (97) using the tidyverse suite (v.2.0.0) (98).

Off-target amplification prediction

For both the *M. tuberculosis* and *S. pneumoniae* primer schemes, off-target amplification was assessed *in silico* against a set of related genomes. We first downloaded *S. pneumoniae* complete genome assemblies (n = 35) (87) representing 62% of the GPS database lineages. Lineages containing serotype 3 (i.e., GPSC 12) were excluded to demonstrate off-target amplification, though the reference genome for serotype 3 (i.e., OXC 141) was retained (Fig. S3b). We similarly selected a subset of *M. tuberculosis* sequences (n = 76) from our initial set used to develop the *M. tuberculosis* primers (Fig. S3a). For each species, we further compiled a genome cluster consisting of the reference genome used for primer design, 12 diverse strains representing various lineages, and an outgroup (Table S5a) using Parnas (v.0.1.4). The pangenome for each cluster was calculated using Roary (v.3.13.0) (99) and an ML phylogeny was constructed using FastTree (v.2.1.11) (100). Average nucleotide identity was calculated between our references and all other genomes in the cluster using FastANI (v.1.34) (101). Off-target amplification was inferred by primer alignment using Bowtie (v.1.3.1) (102); amplicons were predicted for any properly oriented amplicon pairs within 2,200 bp.

Following sequencing of clinical samples, off-target amplicons from our TB panel only were predicted against a range of co-infections and commensal bacteria found within our patient samples; a list of reference genomes used in this analysis is given in Table S5b.

Serotyping, lineage assignment, and resistance prediction

Mykrobe (42) was used to both assign lineages and predict resistance using the built-in panel 202309 for *M. tuberculosis* (103). Mykrobe generates calls for both fixed resistant and heteroresistant samples (those that are heterozygous at the resistance-associated locus; with an expected sensitivity of >90% at an allele frequency of 8% or above [104]). A time-resolved ML tree was built using our variant call file, along with sample data generated from Augur (v.22.4.0), IQ-TREE (v.2.23), and TreeTime (v.0.10.1) (Fig. S2). Tree visualizations were done using Auspice (v.2.57.0).

ACKNOWLEDGMENTS

This publication was supported by the Office of Advanced Molecular Detection, Centers for Disease Control and Prevention, through Cooperative Agreement Number CK22-2204. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention. This publication was made possible by the New England Pathogen Genomics Center of Excellence (US CDC NU50CK000629); the National Heart, Lung, and Blood Institute of the National Institutes of Health; and the Richard K. Gershon Endowed Medical Student Fellowship at Yale University School of Medicine. Research reported in this publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number 1S10OD030363-01A1. L.G. was funded by a Wellcome Trust Career Development Award (226007/Z/22/Z) and the US National Institutes of Health R01 (5R01AI146338-02). A.O. was supported by an Institute of Child Health Child Health Research PhD fellowship.

AUTHOR AFFILIATIONS

[1]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, Connecticut, USA

[2]Department of Ecology and Evolutionary Biology, Yale University, New Haven, Connecticut, USA
[3]Department of Infection, Immunity, and Inflammation, Institute of Child Health, University College London, London, England
[4]Universidad Peruana Cayetano Heredia, Lima, Peru
[5]Institute of Phthisiopneumology, Chisinau, Moldova
[6]Yale Institute for Global Health, Yale University, New Haven, Connecticut, USA
[7]Public Health Modeling Unit, Yale School of Public Health, New Haven, Connecticut, USA

AUTHOR ORCIDs

Patricia Sheen http://orcid.org/0000-0002-7118-9301
Mirko Zimic http://orcid.org/0000-0002-7203-8847
Orchid M. Allicock http://orcid.org/0000-0002-6570-5453
Daniel M. Weinberger http://orcid.org/0000-0003-1178-8086
Seth N. Redmond http://orcid.org/0000-0003-2653-760X

FUNDING

Funder	Grant(s)	Author(s)
National Institutes of Health	1S10OD030363-01A1	Matthew Brandt
Wellcome Trust	226007/Z/22/Z	Louis Grandjean
National Institutes of Health	5R01AI146338-02	Louis Grandjean
Centers for Disease Control and Prevention	CK22-2204	Nathan D. Grubaugh
National Institutes of Health	Medical Student Fellowship	Chaney C. Kalinich
Centers for Disease Control and Prevention	NU50CK000629	Nathan D. Grubaugh
Great Ormond Street Institute of Child Health	PhD Fellowship	Alice Osmaston
Yale University	Richard K. Gershon Endowed Medical Student Fellowship	Chaney C. Kalinich

AUTHOR CONTRIBUTIONS

Chaney C. Kalinich, Formal analysis, Software | Freddy L. Gonzalez, Formal analysis, Visualization | Alice Osmaston, Data curation, Formal analysis, Resources | Mallery I. Breban, Data curation, Methodology | Isabel Distefano, Investigation | Candy Leon, Data curation, Resources | Jorge Coronel, Data curation, Resources | Grace Tan, Data curation, Resources | Valeriu Crudu, Data curation, Formal analysis, Resources | Nelly Ciobanu, Data curation, Formal analysis | Alexandru Codreanu, Data curation, Resources | Walter Solano, Data curation, Resources | Jimena Ráez, Data curation, Resources | Patricia Sheen, Data curation, Resources | Mirko Zimic, Data curation, Resources | Orchid M. Allicock, Data curation, Resources | Chrispin Chaguza, Formal analysis, Methodology | Anne L. Wyllie, Formal analysis, Resources | Matthew Brandt, Data curation, Software | Daniel M. Weinberger, Conceptualization | Benjamin Sobkowiak, Formal analysis, Software | Ted Cohen, Conceptualization, Resources, Supervision | Louis Grandjean, Formal analysis, Resources, Supervision | Nathan D. Grubaugh, Conceptualization, Supervision | Seth N. Redmond, Conceptualization, Formal analysis, Software, Supervision, Visualization

DATA AVAILABILITY

Pipelines for data processing and drug susceptibility prediction, and code for analysis and publication figures, are available on Github (https://github.com/cck42/YSPH_TBseq/releases/tag/v2.0.2). All raw sequences used in this study are available on SRA under BioProject PRJNA1280504 (https://www.ncbi.nlm.nih.gov/bioproject/1280504). While

stocks are available, aliquots of the TB-seq primer pool can be sent to academic or public health partners for testing on request.

ETHICS APPROVAL

All specimens were de-identified, remnant specimens used previously for diagnostic testing or IRB-approved human subjects research, in accordance with Yale University IRB-exempt protocol #2000033281. *M. tuberculosis*-positive specimens from Moldova were remnant specimens collected from study participants enrolled and sampled in accordance with protocol #2000023071 approved by Yale University Human Investigations Committee and the Ethics Committee of Research of the Phthisiopneumology Institute, Moldova. *M. tuberculosis* specimens from Peru were remnant specimens collected from study participants enrolled in accordance with protocol #204749 approved by the Institutional Committee on Research Ethics at Cayetano Heredia University, Peru.

ADDITIONAL FILES

The following material is available online.

Supplemental Material

Supplemental material (JCM01823-25-S0001.docx). Figures S1 to S4, Tables S1 to S6, and Appendices S1 and S3.
Appendix S2 (JCM01823-25-S0002.xlsx). TB-seq and *S. pneumoniae* pangenome assembly and primer design reference sequences.

REFERENCES

1. World Health Organization. 2024. Global tuberculosis report 2024. World Health Organization, Genève, Switzerland
2. World Health Organization. 2023. Global tuberculosis report. World Health Organization
3. O'Toole Á, Scher E, Underwood A, Jackson B, Hill V, McCrone JT, Colquhoun R, Ruis C, Abu-Dahab K, Taylor B, Yeats C, du Plessis L, Maloney D, Medd N, Attwood SW, Aanensen DM, Holmes EC, Pybus OG, Rambaut A. 2021. Assignment of epidemiological lineages in an emerging pandemic using the pangolin tool. Virus Evol 7:veab064. https://doi.org/10.1093/ve/veab064
4. Worobey M, Pekar J, Larsen BB, Nelson MI, Hill V, Joy JB, Rambaut A, Suchard MA, Wertheim JO, Lemey P. 2020. The emergence of SARS-CoV-2 in Europe and North America. Science 370:564–570. https://doi.org/10.1126/science.abc8169
5. Francis RV, Billam H, Clarke M, Yates C, Tsoleridis T, Berry L, Mahida N, Irving WL, Moore C, Holmes N, Ball JK, Loose M, McClure CP. 2022. COVID-19 Genomics UK (COG-UK) consortium. J Infect Dis 225:10–18. https://doi.org/10.1093/infdis/jiab483
6. Hill V, Githinji G, Vogels CBF, Bento AI, Chaguza C, Carrington CVF, Grubaugh ND. 2023. Toward a global virus genomic surveillance network. Cell Host Microbe 31:861–873. https://doi.org/10.1016/j.chom.2023.03.003
7. Gardy JL, Johnston JC, Ho Sui SJ, Cook VJ, Shah L, Brodkin E, Rempel S, Moore R, Zhao Y, Holt R, Varhol R, Birol I, Lem M, Sharma MK, Elwood K, Jones SJM, Brinkman FSL, Brunham RC, Tang P. 2011. Whole-genome sequencing and social-network analysis of a tuberculosis outbreak. N Engl J Med 364:730–739. https://doi.org/10.1056/NEJMoa1003176
8. Lee RS, Proulx J-F, McIntosh F, Behr MA, Hanage WP. 2020. Previously undetected super-spreading of *Mycobacterium tuberculosis* revealed by deep sequencing. eLife 9:e53245. https://doi.org/10.7554/eLife.53245
9. Hatherell H-A, Didelot X, Pollock SL, Tang P, Crisan A, Johnston JC, Colijn C, Gardy JL. 2016. Declaring a tuberculosis outbreak over with genomic epidemiology. Microb Genom 2:e000060. https://doi.org/10.1099/mgen.0.000060
10. 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
11. Yang C, Sobkowiak B, Naidu V, Codreanu A, Ciobanu N, Gunasekera KS, Chitwood MH, Alexandru S, Bivol S, Russi M, Havumaki J, Cudahy P, Fosburgh H, Allender CJ, Centner H, Engelthaler DM, Menzies NA, Warren JL, Crudu V, Colijn C, Cohen T. 2022. Phylogeography and transmission of *M. tuberculosis* in Moldova: a prospective genomic analysis. PLoS Med 19:e1003933. https://doi.org/10.1371/journal.pmed.1003933
12. Cohen KA, Manson AL, Desjardins CA, Abeel T, Earl AM. 2019. Deciphering drug resistance in *Mycobacterium tuberculosis* using whole-genome sequencing: progress, promise, and challenges. Genome Med 11:45. https://doi.org/10.1186/s13073-019-0660-8
13. Manson AL, Cohen KA, Abeel T, Desjardins CA, Armstrong DT, Barry CE III, Brand J, Chapman SB, Cho S-N, Gabrielian A, et al. 2017. Genomic analysis of globally diverse *Mycobacterium tuberculosis* strains provides insights into the emergence and spread of multidrug resistance. Nat Genet 49:395–402. https://doi.org/10.1038/ng.3767
14. Dowdy DW, Theron G, Tornheim JA, Warren R, Kendall EA. 2017. Of testing and treatment: implications of implementing new regimens for multidrug-resistant tuberculosis. Clin Infect Dis 65:1206–1211. https://doi.org/10.1093/cid/cix486
15. The CRyPTIC Consortium. 2022. Genome-wide association studies of global *Mycobacterium tuberculosis* resistance to 13 antimicrobials in 10,228 genomes identify new resistance mechanisms. PLoS Biol 20:e3001755. https://doi.org/10.1371/journal.pbio.3001755
16. World Health Organization. 2023. The use of next-generation sequencing for the surveillance of drug-resistant tuberculosis: an implementation manual. Available from: https://www.who.int/publications/i/item/9789240078079
17. Satta G, Lipman M, Smith GP, Arnold C, Kon OM, McHugh TD. 2018. *Mycobacterium tuberculosis* and whole-genome sequencing: how close are we to unleashing its full potential? Clin Microbiol Infect 24:604–609. https://doi.org/10.1016/j.cmi.2017.10.030
18. Cannas A, Butera O, Mazzarelli A, Messina F, Vulcano A, Parracino MP, Gualano G, Palmieri F, Di Caro A, Nisii C, Fontana C, Girardi E. 2024. Implementation of whole genome sequencing of tuberculosis isolates

in a referral center in Rome: six years' experience in characterizing drug-resistant TB and disease transmission. Antibiotics (Basel) 13:134. https://doi.org/10.3390/antibiotics13020134
19. Doughty EL, Sergeant MJ, Adetifa I, Antonio M, Pallen MJ. 2014. Culture-independent detection and characterisation of *Mycobacterium tuberculosis* and *M. africanum* in sputum samples using shotgun metagenomics on a benchtop sequencer. PeerJ 2:e585. https://doi.org/10.7717/peerj.585
20. Votintseva AA, Bradley P, Pankhurst L, Del Ojo Elias C, Loose M, Nilgiriwala K, Chatterjee A, Smith EG, Sanderson N, Walker TM, Morgan MR, Wyllie DH, Walker AS, Peto TEA, Crook DW, Iqbal Z. 2017. Same-day diagnostic and surveillance data for tuberculosis via whole-genome sequencing of direct respiratory samples. J Clin Microbiol 55:1285–1298. https://doi.org/10.1128/JCM.02483-16
21. Brown AC, Bryant JM, Einer-Jensen K, Holdstock J, Houniet DT, Chan JZM, Depledge DP, Nikolayevskyy V, Broda A, Stone MJ, Christiansen MT, Williams R, McAndrew MB, Tutill H, Brown J, Melzer M, Rosmarin C, McHugh TD, Shorten RJ, Drobniewski F, Speight G, Breuer J. 2015. Rapid whole-genome sequencing of *Mycobacterium tuberculosis* isolates directly from clinical samples. J Clin Microbiol 53:2230–2237. https://doi.org/10.1128/JCM.00486-15
22. Roetzer A, Diel R, Kohl TA, Rückert C, Nübel U, Blom J, Wirth T, Jaenicke S, Schuback S, Rüsch-Gerdes S, Supply P, Kalinowski J, Niemann S. 2013. Whole genome sequencing versus traditional genotyping for investigation of a *Mycobacterium tuberculosis* outbreak: a longitudinal molecular epidemiological study. PLoS Med 10:e1001387. https://doi.org/10.1371/journal.pmed.1001387
23. World Health Organization (WHO). 2012. Tuberculosis laboratory biosafety manual. Genève, Switzerland
24. Nimmo C, Shaw LP, Doyle R, Williams R, Brien K, Burgess C, Breuer J, Balloux F, Pym AS. 2019. Whole genome sequencing *Mycobacterium tuberculosis* directly from sputum identifies more genetic diversity than sequencing from culture. BMC Genomics 20:389. https://doi.org/10.1186/s12864-019-5782-2
25. Shockey AC, Dabney J, Pepperell CS. 2019. Effects of host, sample, and *in vitro* culture on genomic diversity of pathogenic mycobacteria. Front Genet 10:477. https://doi.org/10.3389/fgene.2019.00477
26. Mariner-Llicer C, Goig GA, Torres-Puente M, Vashakidze S, Villamayor LM, Saavedra-Cervera B, Mambuque E, Khurtsilava I, Avaliani Z, Rosenthal A, Gabrielian A, Shurgaia M, Shubladze N, García-Basteiro AL, López MG, Comas I. 2024. Genetic diversity within diagnostic sputum samples is mirrored in the culture of *Mycobacterium tuberculosis* across different settings. Nat Commun 15:7114. https://doi.org/10.1038/s41467-024-51266-0
27. Penn-Nicholson A, Georghiou SB, Ciobanu N, Kazi M, Bhalla M, David A, Conradie F, Ruhwald M, Crudu V, Rodrigues C, Myneedu VP, Scott L, Denkinger CM, Schumacher SG, Consortium XXT. 2022. Detection of isoniazid, fluoroquinolone, ethionamide, amikacin, kanamycin, and capreomycin resistance by the Xpert MTB/XDR assay: a cross-sectional multicentre diagnostic accuracy study. Lancet Infect Dis 22:242–249. https://doi.org/10.1016/S1473-3099(21)00452-7
28. Gao Y, Zhang Z, Deng J, Mansjö M, Ning Z, Li Y, Li X, Hu Y, Hoffner S, Xu B. 2018. Multi-center evaluation of GenoType MTBDRsl line probe assay for rapid detection of pre-XDR and XDR *Mycobacterium tuberculosis* in China. Journal of Infection 77:328–334. https://doi.org/10.1016/j.jinf.2018.06.014
29. Ejo M, Van Deun A, Nunn A, Meredith S, Ahmed S, Dalai D, Tumenbayar O, Tsogt B, Dat PT, Ha DTM, et al. 2021. Effectiveness of GenoType MTBDR*sl* in excluding TB drug resistance in a clinical trial. Int J Tuberc Lung Dis 25:839–845. https://doi.org/10.5588/ijtld.21.0212
30. Conkle-Gutierrez D, Kim C, Ramirez-Busby SM, Modlin SJ, Mansjö M, Werngren J, Rigouts L, Hoffner SE, Valafar F. 2022. Distribution of common and rare genetic markers of second-line-injectable-drug resistance in *Mycobacterium tuberculosis* revealed by a genome-wide association study. Antimicrob Agents Chemother 66:e0207521. https://doi.org/10.1128/aac.02075-21
31. World Health Organization. 2022. WHO consolidated guidelines on tuberculosis: module 4: treatment - drug-resistant tuberculosis treatment, 2022 update. World Health Organization, Geneva
32. World Health Organization. 2024. WHO consolidated guidelines on tuberculosis. Module 3: diagnosis – rapid diagnostics for tuberculosis detection. World Health Organization
33. de Araujo L, Cabibbe AM, Mhuulu L, Ruswa N, Dreyer V, Diergaardt A, Günther G, Claassens M, Gerlach C, Utpatel C, Cirillo DM, Nepolo E, Niemann S. 2023. Implementation of targeted next-generation sequencing for the diagnosis of drug-resistant tuberculosis in low-resource settings: a programmatic model, challenges, and initial outcomes. Front Public Health 11:1204064. https://doi.org/10.3389/fpubh.2023.1204064
34. Grobbel H-P, Merker M, Köhler N, Andres S, Hoffmann H, Heyckendorf J, Reimann M, Barilar I, Dreyer V, Hillemann D, Kalsdorf B, Kohl TA, Sanchez Carballo P, Schaub D, Todt K, Utpatel C, Maurer FP, Lange C, Niemann S. 2021. Design of multidrug-resistant tuberculosis treatment regimens based on DNA sequencing. Clin Infect Dis 73:1194–1202. https://doi.org/10.1093/cid/ciab359
35. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, Oliveira G, Robles-Sikisaka R, Rogers TF, Beutler NA, et al. 2017. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc 12:1261–1276. https://doi.org/10.1038/nprot.2017.066
36. 2024. Artic Network. Available from: https://artic.network/viruses/sars-cov-2. Retrieved 30 Aug 2024.
37. Langedijk AC, Lebbink RJ, Naaktgeboren C, Evers A, Viveen MC, Greenough A, Heikkinen T, Stein RT, Richmond P, Martinón-Torres F, et al. 2020. Global molecular diversity of RSV - the "INFORM RSV" study. BMC Infect Dis 20:450. https://doi.org/10.1186/s12879-020-05175-4
38. Tulloch RL, Kok J, Carter I, Dwyer DE, Eden J-S. 2021. An amplicon-based approach for the whole-genome sequencing of human metapneumovirus. Viruses 13:499. https://doi.org/10.3390/v13030499
39. Vogels CBF, Hill V, Breban MI, Chaguza C, Paul LM, Sodeinde A, Taylor-Salmon E, Ott IM, Petrone ME, Dijk D, et al. 2024. DengueSeq: a pan-serotype whole genome amplicon sequencing protocol for dengue virus. BMC Genomics 25:433. https://doi.org/10.1186/s12864-024-10350-x
40. Chen NFG, Chaguza C, Gagne L, Doucette M, Smole S, Buzby E, Hall J, Ash S, Harrington R, Cofsky S, et al. 2023. Development of an amplicon-based sequencing approach in response to the global emergence of mpox. PLoS Biol 21:e3002151. https://doi.org/10.1371/journal.pbio.3002151
41. Pfyffer GE. 2015. *Mycobacterium*: general characteristics, laboratory detection, and staining procedures, p 536–569. *In* In manual of clinical microbiology. ASM Press, Washington, DC, USA.
42. Hunt M, Bradley P, Lapierre SG, Heys S, Thomsit M, Hall MB, Malone KM, Wintringer P, Walker TM, Cirillo DM, et al. 2019. Antibiotic resistance prediction for *Mycobacterium tuberculosis* from genome sequence data with Mykrobe. Wellcome Open Res 4:191. https://doi.org/10.12688/wellcomeopenres.15603.1
43. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. 2020. IQ-TREE 2:new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol 37:1530–1534. https://doi.org/10.1093/molbev/msaa015
44. World Health Organization. 2017. Technical manual for drug susceptibility testing of medicines used in the treatment of tuberculosis. Available from: https://www.who.int/publications/i/item/9789241514842
45. Tettelin H, Nelson KE, Paulsen IT, Eisen JA, Read TD, Peterson S, Heidelberg J, DeBoy RT, Haft DH, Dodson RJ, et al. 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science 293:498–506. https://doi.org/10.1126/science.1061217
46. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry CE III, et al. 1998. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. Nature 393:537–544. https://doi.org/10.1038/31159
47. Gardy JL, Loman NJ. 2018. Towards a genomics-informed, real-time, global pathogen surveillance system. Nat Rev Genet 19:9–20. https://doi.org/10.1038/nrg.2017.88
48. Lai R, Ogunsola AF, Rakib T, Behar SM. 2023. Key advances in vaccine development for tuberculosis—success and challenges. NPJ Vaccines 8:158. https://doi.org/10.1038/s41541-023-00750-7
49. Lienhardt C, Dooley KE, Nahid P, Wells C, Ryckman TS, Kendall EA, Davies G, Brigden G, Churchyard G, Cirillo DM, et al. 2024. Target regimen profiles for tuberculosis treatment. Bull World Health Organ 102:600–607. https://doi.org/10.2471/BLT.24.291881
50. Tseng S-Y, Huang Y-S, Chang T-E, Perng C-L, Huang Y-H. 2021. Hepatotoxicity, efficacy and completion rate between 3 months of isoniazid plus rifapentine and 9 months of isoniazid in treating latent tuberculosis infection: a systematic review and meta-analysis. J Chin

Med Assoc 84:993–1000. https://doi.org/10.1097/JCMA.0000000000000605
51. Walker TM, Miotto P, Köser CU, Fowler PW, Knaggs J, Iqbal Z, Hunt M, Chindelevitch L, Farhat MR, Cirillo DM, et al. 2022. The 2021 WHO catalogue of *Mycobacterium tuberculosis* complex mutations associated with drug resistance: a genotypic analysis. The Lancet Microbe 3:e265–e273. https://doi.org/10.1016/S2666-5247(21)00301-3
52. Kadura S, King N, Nakhoul M, Zhu H, Theron G, Köser CU, Farhat M. 2020. Systematic review of mutations associated with resistance to the new and repurposed *Mycobacterium tuberculosis* drugs bedaquiline, clofazimine, linezolid, delamanid and pretomanid. J Antimicrob Chemother 75:2031–2043. https://doi.org/10.1093/jac/dkaa136
53. Pankhurst LJ, Del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, Fermont JM, Gascoyne-Binzi DM, Kohl TA, Kong C, Lemaitre N, Niemann S, Paul J, Rogers TR, Roycroft E, Smith EG, Supply P, Tang P, Wilcox MH, Wordsworth S, Wyllie D, Xu L, Crook DW, Group C-T. 2016. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. Lancet Respir Med 4:49–58. https://doi.org/10.1016/S2213-2600(15)00466-X
54. Goig GA, Cancino-Muñoz I, Torres-Puente M, Villamayor LM, Navarro D, Borrás R, Comas I. 2020. Whole-genome sequencing of *Mycobacterium tuberculosis* directly from clinical samples for high-resolution genomic epidemiology and drug resistance surveillance: an observational study. Lancet Microbe 1:e175–e183. https://doi.org/10.1016/S2666-5247(20)30060-4
55. Lozano N, Lanza VF, Suárez-González J, Herranz M, Sola-Campoy PJ, Rodríguez-Grande C, Buenestado-Serrano S, Ruiz-Serrano MJ, Tudó G, Alcaide F, Muñoz P, García de Viedma D, Pérez-Lago L. 2021. Detection of minority variants and mixed infections in *Mycobacterium tuberculosis* by direct whole-genome sequencing on noncultured specimens using a specific-DNA capture strategy. mSphere 6:e0074421. https://doi.org/10.1128/mSphere.00744-21
56. Lam C, Martinez E, Crighton T, Furlong C, Donnan E, Marais BJ, Sintchenko V. 2021. Value of routine whole genome sequencing for *Mycobacterium tuberculosis* drug resistance detection. Int J Infect Dis 113 Suppl 1:S48–S54. https://doi.org/10.1016/j.ijid.2021.03.033
57. Soetaert K, Ceyssens P-J, Boarbi S, Bogaerts B, Delcourt T, Vanneste K, De Keersmaecker SCJ, Roosens NHC, Vodolazkaia A, Mukovnikova M, Mathys V. 2022. Retrospective evaluation of routine whole genome sequencing of *Mycobacterium tuberculosis* at the Belgian National reference center, 2019. Acta Clin Belg 77:853–860. https://doi.org/10.1080/17843286.2021.1999588
58. Olaru ID, Patel H, Kranzer K, Perera N. 2018. Turnaround time of whole genome sequencing for mycobacterial identification and drug susceptibility testing in routine practice. Clin Microbiol Infect 24:659. https://doi.org/10.1016/j.cmi.2017.10.001
59. Shea J, Halse TA, Lapierre P, Shudt M, Kohlerschmidt D, Van Roey P, Limberger R, Taylor J, Escuyer V, Musser KA. 2017. Comprehensive whole-genome sequencing and reporting of drug resistance profiles on clinical cases of *Mycobacterium tuberculosis* in New York state. J Clin Microbiol 55:1871–1882. https://doi.org/10.1128/JCM.00298-17
60. Van Rie A, De Vos E, Costa E, Verboven L, Ndebele F, Heupink TH, Abrams S, SMARTT team, Fanampe B, Van der Spoel Van Dyk A, Charalambous S, Churchyard G, Warren R. 2022. Sequencing mycobacteria and algorithm-determined resistant tuberculosis treatment (SMARTT): a study protocol for a phase IV pragmatic randomized controlled patient management strategy trial. Trials 23:864. https://doi.org/10.1186/s13063-022-06793-w
61. McNerney R, Clark TG, Campino S, Rodrigues C, Dolinger D, Smith L, Cabibbe AM, Dheda K, Schito M. 2017. Removing the bottleneck in whole genome sequencing of *Mycobacterium tuberculosis* for rapid drug resistance analysis: a call to action. Int J Infect Dis 56:130–135. https://doi.org/10.1016/j.ijid.2016.11.422
62. Bryant JM, Schürch AC, van Deutekom H, Harris SR, de Beer JL, de Jager V, Kremer K, van Hijum S, Siezen RJ, Borgdorff M, Bentley SD, Parkhill J, van Soolingen D. 2013. Inferring patient to patient transmission of *Mycobacterium tuberculosis* from whole genome sequencing data. BMC Infect Dis 13:110. https://doi.org/10.1186/1471-2334-13-110
63. Guerra-Assunção JA, Crampin AC, Houben R, Mzembe T, Mallard K, Coll F, Khan P, Banda L, Chiwaya A, Pereira RPA, McNerney R, Fine PEM, Parkhill J, Clark TG, Glynn JR. 2015. Large-scale whole genome sequencing of *M. tuberculosis* provides insights into transmission in a high prevalence area. eLife 4:e05166. https://doi.org/10.7554/eLife.05166
64. Bobay L-M, Ochman H. 2018. Factors driving effective population size and pan-genome evolution in bacteria. BMC Evol Biol 18:153. https://doi.org/10.1186/s12862-018-1272-4
65. Hunt M, Lima L, Shen W, Lees J, Iqbal Z. 2024. AllTheBacteria - all bacterial genomes assembled, available and searchable. bioRxiv.
66. Noll N, Molari M, Shaw LP, Neher RA. 2023. *PanGraph*: scalable bacterial pan-genome graph construction. Microb Genom 9:001034. https://doi.org/10.1099/mgen.0.001034
67. Avila Cartes J, Bonizzoni P, Ciccolella S, Della Vedova G, Denti L, Didelot X, Monti DC, Pirola Y. 2024. RecGraph: recombination-aware alignment of sequences to variation graphs. Bioinformatics 40:btae292. https://doi.org/10.1093/bioinformatics/btae292
68. Qian J, Chen R, Wang H, Zhang X. 2020. Role of the PE/PPE family in host-pathogen interactions and prospects for anti-tuberculosis vaccine and diagnostic tool design. Front Cell Infect Microbiol 10:594288. https://doi.org/10.3389/fcimb.2020.594288
69. D'Souza C, Kishore U, Tsolaki AG. 2023. The PE-PPE family of *Mycobacterium tuberculosis*: proteins in disguise. Immunobiology 228:152321. https://doi.org/10.1016/j.imbio.2022.152321
70. Martin-Higuera MC, Rivas G, Rolo M, Muñoz-Gallego I, Lopez-Roa P. 2023. Xpert MTB/RIF Ultra CT value provides a rapid measure of sputum bacillary burden and predicts smear status in patients with pulmonary tuberculosis. Sci Rep 13:1591. https://doi.org/10.1038/s41598-023-28869-6
71. Chakravorty S, Simmons AM, Rowneki M, Parmar H, Cao Y, Ryan J, Banada PP, Deshpande S, Shenai S, Gall A, Glass J, Krieswirth B, Schumacher SG, Nabeta P, Tukvadze N, Rodrigues C, Skrahina A, Tagliani E, Cirillo DM, Davidow A, Denkinger CM, Persing D, Kwiatkowski R, Jones M, Alland D. 2017. The new Xpert MTB/RIF ultra: improving detection of *Mycobacterium tuberculosis* and resistance to rifampin in an assay suitable for point-of-care testing. mBio 8:e00812-17. https://doi.org/10.1128/mBio.00812-17
72. Denkinger CM, Schumacher SG, Gilpin C, Korobitsyn A, Wells WA, Pai M, Leeflang M, Steingart KR, Bulterys M, Schünemann H, Glaziou P, Weyer K. 2019. Guidance for the evaluation of tuberculosis diagnostics that meet the World Health Organization (WHO) target product profiles: an introduction to WHO process and study design principles. J Infect Dis 220:S91–S98. https://doi.org/10.1093/infdis/jiz097
73. Allix-Béguec C, Arandjelovic I, Bi L, Beckert P, Bonnet M, Bradley P, Cabibbe AM, Cancino-Muñoz I, Caulfield MJ, Chaiprasert A, et al. 2018. Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. N Engl J Med 379:1403–1415. https://doi.org/10.1056/NEJMoa1800474
74. Ntagereka PB, Oyola SO, Baenyi SP, Rono GK, Birindwa AB, Shukuru DW, Baharanyi TC, Kashosi TM, Buhendwa J-P, Bisimwa PB, Kusinza AB, Basengere RA, Mukwege D. 2022. Whole-genome sequencing of SARS-CoV-2 reveals diverse mutations in circulating Alpha and Delta variants during the first, second, and third waves of COVID-19 in South Kivu, east of the Democratic Republic of the Congo. Int J Infect Dis 122:136–143. https://doi.org/10.1016/j.ijid.2022.05.041
75. Khairnar K, Tomar SS. 2024. COVID-19 genome surveillance: a geographical landscape and mutational mapping of SARS-CoV-2 variants in central India over two years. Virus Res 344:199365. https://doi.org/10.1016/j.virusres.2024.199365
76. Cowley LA, Afrad MH, Rahman SIA, Mamun MMA, Chin T, Mahmud A, Rahman MZ, Billah MM, Khan MH, Sultana S, Khondaker T, Baker S, Banik N, Alam AN, Mannoor K, Banu S, Chowdhury A, Flora MS, Thomson NR, Buckee CO, Qadri F, Shirin T. 2021. Genomics, social media and mobile phone data enable mapping of SARS-CoV-2 lineages to inform health policy in Bangladesh. Nat Microbiol 6:1271–1278. https://doi.org/10.1038/s41564-021-00955-3
77. Saha S, Tanmoy AM, Hooda Y, Tanni AA, Goswami S, Sium SMA, Sajib MSI, Malaker R, Islam S, Rahman H, Anik AM, Sarker N, Islam MS, Ghosh K, Sarkar PK, Bipul MRA, Ahmed SS, Shahidullah M, Saha SK. 2021. COVID-19 rise in Bangladesh correlates with increasing detection of B.1.351 variant. BMJ Glob Health 6:e006012. https://doi.org/10.1136/bmjgh-2021-006012
78. Jony MHK, Alam AN, Nasif MAO, Sultana S, Anwar R, Rudra M, Rahman M, Rahman M, Qadri F, Shirin T. 2024. Emergence of SARS-CoV-2 Omicron sub-lineage JN.1 in Bangladesh. Microbiol Resour Announc 13:e0013024. https://doi.org/10.1128/mra.00130-24
79. Koo H-K, Min J, Kim HW, Lee J, Kim JS, Park JS, Lee S-S. 2020. Prediction of treatment failure and compliance in patients with tuberculosis. BMC Infect Dis 20:622. https://doi.org/10.1186/s12879-020-05350-7

80. Liu Q, Zhu J, Dulberger CL, Stanley S, Wilson S, Chung ES, Wang X, Culviner P, Liu YJ, Hicks ND, Babunovic GH, Giffen SR, Aldridge BB, Garner EC, Rubin EJ, Chao MC, Fortune SM. 2022. Tuberculosis treatment failure associated with evolution of antibiotic resilience. Science 378:1111–1118. https://doi.org/10.1126/science.abq2787
81. van der Werf MJ, Langendam MW, Huitric E, Manissero D. 2012. Multidrug resistance after inappropriate tuberculosis treatment: a meta-analysis. Eur Respir J 39:1511–1519. https://doi.org/10.1183/09031936.00125711
82. Huddleston J, Hadfield J, Sibley TR, Lee J, Fay K, Ilcisin M, Harkins E, Bedford T, Neher RA, Hodcroft EB. 2021. Augur: a bioinformatics toolkit for phylogenetic analyses of human pathogens. JOSS 6:2906. https://doi.org/10.21105/joss.02906
83. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300
84. Sagulenko P, Puller V, Neher RA. 2018. TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol 4:vex042. https://doi.org/10.1093/ve/vex042
85. Markin A, Wagle S, Grover S, Vincent Baker AL, Eulenstein O, Anderson TK. 2023. PARNAS: objectively selecting the most representative taxa on a phylogeny. Syst Biol 72:1052–1063. https://doi.org/10.1093/sysbio/syad028
86. Ates LS. 2020. New insights into the mycobacterial PE and PPE proteins provide a framework for future research. Mol Microbiol 113:4–21. https://doi.org/10.1111/mmi.14409
87. Gladstone RA, Lo SW, Lees JA, Croucher NJ, Tonder AJ, Corander J, Page AJ, Marttinen P, Bentley LJ, Ochoa TJ, et al. 2019. International genomic definition of pneumococcal lineages, to contextualise disease, antibiotic resistance and vaccine impact. EBioMedicine 43:338–346. https://doi.org/10.1016/j.ebiom.2019.04.021
88. Goig GA, Torres-Puente M, Mariner-Llicer C, Villamayor LM, Chiner-Oms Á, Gil-Brusola A, Borrás R, Comas Espadas I. 2020. Towards next-generation diagnostics for tuberculosis: identification of novel molecular targets by large-scale comparative genomics. Bioinformatics 36:985–989. https://doi.org/10.1093/bioinformatics/btz729
89. Adebiyi MG. 2024. How to resuspend, store, and improve the longevity of your IDT Gene Fragments. *Integrated DNA Technologies*. Available from: https://www.idtdna.com/pages/education/decoded/article/how-to-resuspend-store-and-improve-the-longevity-of-your-idt-gene-fragments
90. Chen NFG, Gagne L, Doucette M, Smole S, Buzby E, Hall J, Ash S, Harrington R, Cofsky S, Clancy S, Kapsak CJ, Sevinsky J, Libuit K, Chaguza C, Grubaugh ND, Park DJ, Gallagher GE, Vogels CBF. 2022. Monkeypox virus multiplexed PCR amplicon sequencing (PrimalSeq). Protocols.io. Available from: https://doi.org/10.17504/protocols.io.5qpvob1nbl4o/v4
91. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv
92. 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
93. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, Tan AL, Paul LM, Brackney DE, Grewal S, Gurfield N, Van Rompay KKA, Isern S, Michael SF, Coffey LL, Loman NJ, Andersen KG. 2019. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol 20:8. https://doi.org/10.1186/s13059-018-1618-7
94. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet j 17:10. https://doi.org/10.14806/ej.17.1.200
95. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. Gigascience 10:giab008. https://doi.org/10.1093/gigascience/giab008
96. Lu D, Kalantar KL, Chu VT, Glascock AL, Guerrero ES, Bernick N, Butcher X, Ewing K, Fahsbender E, Holmes O, Hoops E, Jones AE, Lim R, McCanny S, Reynoso L, Rosario K, Tang J, Valenzuela O, Mourani PM, Pickering AJ, Raphenya AR, Alcock BP, McArthur AG, Langelier CR. 2024. Simultaneous detection of pathogens and antimicrobial resistance genes with the open source, cloud-based, CZ ID pipeline. bioRxiv. https://doi.org/10.1101/2024.04.12.589250
97. Allaire J. 2012. RStudio: integrated development environment for R. Boston, MA.
98. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen T, Miller E, Bache S, Müller K, Ooms J, Robinson D, Seidel D, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the tidyverse. JOSS 4:1686. https://doi.org/10.21105/joss.01686
99. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, Fookes M, Falush D, Keane JA, Parkhill J. 2015. Roary: rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693. https://doi.org/10.1093/bioinformatics/btv421
100. Price MN, Dehal PS, Arkin AP. 2009. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. Mol Biol Evol 26:1641–1650. https://doi.org/10.1093/molbev/msp077
101. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9
102. Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25. https://doi.org/10.1186/gb-2009-10-3-r25
103. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. 2022. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res 50:D785–D794. https://doi.org/10.1093/nar/gkab776
104. Bradley P, Gordon NC, Walker TM, Dunn L, Heys S, Huang B, Earle S, Pankhurst LJ, Anson L, de Cesare M, et al. 2015. Rapid antibiotic-resistance predictions from genome sequence data for *Staphylococcus aureus* and *Mycobacterium tuberculosis*. Nat Commun 6:10063. https://doi.org/10.1038/ncomms10063