Frequently Asked Questions (FAQs):

Genomic Sequencing for Drug Resistant Tuberculosis

General Questions

1) What is sequencing?

Sequencing is a laboratory procedure that determines the order of nucleotide in a fraction, or the whole genome (RNA or DNA), of an organism.

2) What is next-generation sequencing?

Next Generation Sequencing (NGS) is a high-throughput sequencing method used to determine the nucleotide sequence of a genome in a single biochemical reaction. NGS is performed by sequencing technologies that differ from Sanger sequencing. NGS technologies read the sequences of multiple fragments of DNA in parallel. These sequencing 'reads' are then assembled and mapped to a reference genome using software specially designed to manage biological sequencing data, known as 'bioinformatics software'. Once the sequencing reads are assembled and mapped, bioinformatics analyses can identify and investigate changes in the sequence of nucleotides that were read compared with the reference sequence. Such changes in mycobacterial DNA may be associated with antimicrobial resistance or other biological phenomena of interest to disease control programs.

3) What is the difference between Next Generation Sequencing (NGS) and Sanger sequencing?

The main difference between Sanger and Next Generation Sequencing (NGS) is the method they use to determine the sequence of nucleotides in a genome. Sanger sequencing is based on the incorporation of base-specific chain terminating dideoxy-nucleotides by DNA polymerase during in vitro DNA replication for downstream fluorescence detection. NGS, on the other hand, includes a series of technologies that are based on DNA fragmentation and library preparation followed by massive parallel sequencing using fluorescence or voltage detection depending on the type of NGS technology that is used. See **'What are the different types of NGS technologies?'** below for more information on how different NGS technologies work. NGS is high throughput, sequencing millions of DNA fragments at a time for post-sequencing fragment assembly, while Sanger sequencing is lower throughput, sequencing one, longer, DNA fragment at a time. In addition, NGS can detect changes in DNA at lower frequencies (or with greater sensitivity) and is becoming more cost effective as use of the technology expands around the world.

4) What does 'high throughput' sequencing mean?

'High-throughput sequencing' is a term used to describe technologies that sequence millions of DNA fragments simultaneously in a rapid manner. The distinction between NGS instrument throughputs is made based on the volume of data that they produce. This volume (or output)

ranges from a few megabytes (Mb) to 15 gigabytes (Gb; where each gigabyte equals 1000 megabytes) for “benchtop throughput” and up to 6,000 Gb for “high throughput” instruments.

5) How does ‘Whole Genome’ Sequencing (WGS) differ from ‘targeted’ Next Generation Sequencing (tNGS)?

Whole Genome Sequencing (WGS) uses the NGS method to determine the nucleotide sequence of an entire genome for a given organism at one time, while targeted NGS (tNGS) uses the NGS method to determine the nucleotide sequences of specific, purposefully selected regions of a genome (not the whole genome). tNGS, as the name suggests, is best suited for applications in which one already knows the genes or gene regions to be investigated.

While WGS offers comprehensive sequencing information that enables characterization of all genes within microorganisms (including viruses and bacteria), it also requires higher volumes of high-quality DNA that can only currently be obtained from TB culture isolates (not clinical samples). This means that turnaround time for WGS take longer, as DNA can be processed when liquid or solid cultures are properly grown. tNGS does not require as much high-quality DNA and can be performed directly from clinical samples such as sputum. Moreover, the targeted method produces less sequence data, which therefore requires less data storage and analytic processing capacity and time.

Disease program plans and priorities for data use and resource considerations are two primary factors that can help inform selection of WGS and/ or tNGS. The currently available NGS instruments can potentially run WGS or tNGS protocols, however, they differ in cost, batching capacity, and accuracy. Further details are reported in Table 1.1 of the WHO 2023 The use of next-generation sequencing for the surveillance of drug-resistant tuberculosis: an implementation manual.

6) What are the different types of ‘NGS technologies’?

Currently, there are two types of NGS technologies that differ by the length of genomic sequence ‘reads’ that they provide. These types include Short-Read Sequencing (SRS; second-generation) and Long-Read Sequencing (LRS; third generation). In contrast to SRS that produces short reads (50–600 base pairs) representing fragments of a genome, LRS has the capacity to produce long and fewer reads (>10 kilobases on average) across the same portion of a genome.

Short reads are accurate, cheaper to produce with fewer reagents, and are adequate when you have a reference genome. Long reads can improve de novo assembly of sequencing reads in the absence of a reference genome, in cases of mapping certainty, or when transcript isoform identification or detection of structural variants are desired by the user.

As example of SRS, Illumina (**1**) produces a commonly used “short-read” NGS technology that uses a within-procedure innovation called bridge amplification. Millions of clusters of fragmented and amplified DNA are clustered across a silicon chip with fluorescent tags

incorporated that are color-coded for each nucleotide. The fluorescent tags can be excited by a laser and the resultant fluorescence read from each nucleotide to determine the nucleotide sequences of the DNA fragments. The billions of sequencing 'reads' that are obtained vary in length between 50 – 600 nucleotides and are arranged against a complete reference sequence by bioinformatics data analysis solutions. The complete arrangement of reads, or assembled genome, is then used to determine the whole or targeted genome sequences, as well as to identify any differences in sequence between the assembled and reference genomes. Multiple Illumina instruments are available for short-read NGS including the iSeq 100, MiSeq, NextSeq, and NovaSeq, all with various infrastructure requirements and data output capacities.

In contrast to the Illumina-based SRS example, Oxford Nanopore Technologies (ONT) (**2**) produces a common LRS NGS technology. Also known as 'single-molecule' or 'nanopore' sequencing, this technology determines nucleotide sequences by passing a single long strand of DNA through a 'nanopore' in a flow cell that is within an ION sequencer. As the DNA strand passes through the flow cell, the ONT MinION and GridION instruments detect unique changes in electrical current that are generated by each of the four nucleotide bases (A, C, T, and G); thereby reading the sequence of the DNA strand. The MinION, with a one flow cell capacity, and GridION, with a five flow cells capacity, produce sequencing reads of 10-30 Kb or more.

Further details are reported in Annex 6 of the WHO 2023 The use of next-generation sequencing for the surveillance of drug-resistant tuberculosis: an implementation manual.

7) What are the advantages and disadvantages of NGS technologies?

Short-read sequencing is currently the most used form of NGS, enabling high depth and high-quality data for the lowest cost per base and wide variety of applications. Short-Read Sequencing (SRS) typically requires a reference sequence for genome assembly, which supports highly accurate sequencing read assembly and results. However, SRS may involve higher up-front equipment costs, and requires more frequent and complex instrument maintenance compared to certain (but not all) long-read sequencing devices.

In contrast, Long-Read Sequencing (LRS) use for drug-resistant TB has been less common and its accuracy to-date has been lower than short-read technology. However, LRS accuracy has improved in recent years, and therefore may offer comparable results for drug resistance prediction and genotypic analyses with potential lower costs per sample in the future (**3**). In addition, portable sequencing instruments such as the MinION are small, lightweight, and do not require the microscopic alignment and calibration maintenance steps required by short-read instrumentation. In addition, the long-read approach can help resolve challenges with accuracy of short-read sequence assembly, which may occur when sequencing reads for DNA fragments are located within areas of the genome that have complex structures or repetitive series of nucleotides.

Of note, both technologies require sufficient engineering, administrative, and human resources to support unidirectional molecular sample preparation and testing workflows, as well as bioinformatics and data storage needs.

8) What is the difference between the WHO-recommended tNGS solutions and other molecular WHO-recommended Molecular Diagnostics (mWRDs)?

Targeted Next Generation Sequencing (tNGS) solutions provide the complete nucleotide sequences of specific genes within the *Mycobacterium tuberculosis* (Mtb) genome, while other mWRDs use fluorescent or immunologic probes to report the presence or absence of specific, shorter nucleotide sequences within specified regions of Mtb genes. Because tNGS solutions can investigate entire genes instead of gene regions, they detect drug resistance-associated mutations in the Mtb genome more holistically compared to currently recommended mWRDs.

tNGS testing protocols also target a wider number of Mtb genes associated with resistance to more anti-TB drugs, including bedaquiline, clofazimine, linezolid, delamanid, and pretomanid which are not targeted by currently recommended mWRDs. However, tNGS tests are more expensive than mWRDs, take longer to produce results (days compared to hours), and target limited genes associated with resistance to the new and repurposed TB drugs based on our incomplete understanding of these targets. Therefore, WHO recommends use of these tNGS approaches as complementary or follow-on to mWRDs that should be used for initial diagnostic and drug susceptibility testing. Further details are reported in the WHO operational handbook on tuberculosis: module 3: diagnosis: rapid diagnostics for tuberculosis detection, 3rd ed.

9) What are some of the public health applications of NGS for tuberculosis?

Example applications for genomic sequencing of *M. tuberculosis* (Mtb) include (i) comprehensive prediction of drug resistance profiles based on the presence of known mutations and their associations with resistance to specific anti-TB drugs, (ii) surveillance of emergent or circulating strains of Mtb, and (iii) monitoring emergence and trends in drug resistance, or the impact of introducing new anti-TB drugs on drug resistance [WHO 2021 Guidance for the surveillance of drug resistance in tuberculosis].

Additional applications may include investigations of outbreaks to resolve transmission events and networks, confirmation of epidemiological links between TB cases that may share similar or the same strains of Mtb, as well as strain genotyping to determine the types and diversity of MTB lineages that are locally circulating (currently and over time) and can differ in their virulence, drug susceptibility profiles, or transmissibility. Of note, as is the case for many sequencing use cases, outbreak and epidemiological investigations require additional clinical, geographic, or other data to complement sequencing results.

10) How might national TB programs use sequence data to enhance detection of DR TB for clinical use or surveillance?

Since there is a strong correlation between molecular (genotypic) and culture-based (phenotypic) detection of *M. tuberculosis* drug resistance, high-confidence and accurate prediction of drug resistance profiles based on known genomic mutations is possible for most anti-TB drugs. Sequencing can determine resistance to many anti-TB drugs within one testing procedure and data may be used to: (i) confirm drug susceptibility testing (DST) results obtained from other molecular and phenotypic DST methods, (ii) detect resistance that may not have been targeted or detected by other molecular and culture-based DST, and (iii) identify novel mutations associated with drug resistance or heteroresistance. For example, low- and moderate-complexity molecular WHO-recommended rapid diagnostics (mWRDs) interrogate narrow sections of genes associated with drug resistance, and therefore will not detect mutations conferring resistance outside of these regions. Moreover, mutations in the Mtb genome that are associated with growth defects or result in low levels of drug resistance may not be detected by culture-based methods due to their limits of detection. Lastly, existing diagnostic algorithms and tests may not allow for molecular detection of resistance to new anti-TB drugs. Use of the WHO-recommend tNGS solutions for this purpose can provide ruled-in resistance detection with primary clinical samples (sputa) within days for clinical use, where program priorities and resources allow. In addition, sequencing data may be used to complement Mtb detection, lineage, and resistance data obtained by other methods for surveillance use. NGS may therefore be incorporated into testing algorithms to enhance DR TB surveillance as part of national TB drug resistance surveys as well as sentinel or routine surveillance efforts. For more information on how NGS can be used for DR TB clinical use please see the WHO consolidated guidelines on tuberculosis: module 3: diagnosis: rapid diagnostics for tuberculosis detection, 3rd ed. For surveillance use, please see the Guidance for the surveillance of drug resistance in tuberculosis: Sixth edition (who.int) and the WHO 2023 Use of next-generation sequencing for the surveillance of drug-resistant tuberculosis: an implementation manual.

11) What resources are required for implementation of NGS?

Financial resources and budgeting considerations for NGS implemenation should cover costs associated with any testing site infrastructure updates, personnel, instrumentation, equipment and their associated service and maintenance, tests, commodities and reagents, quality control and assurance mechanisms, and other systems costs required to incorporate sequencing into proposed diagnostic or surveillance-based algorithms (see the WHO 2023 Use of next-generation sequencing for the surveillance of drug-resistant tuberculosis: an implementation manual). NGS-related human resource needs include qualified, trained, and competant staff to perform laboratory testing and bioinformatic analyses, as well as IT support for data analytics, data storage, surveillance system connectivity or linkages, reporting capacity, and technical and quality performance monitoring. Of note, the WHO-recommended tNGS solutions for drug-resistant TB reduce the bioinformatics capacity that is needed for implementing sites, as the

data is processed and automatically analyzed for Mtb detection, lineage, and resistance patterns for reported targeted genes and anti-TB drugs on the servers of the tNGS product manufacturers. Automated views of results are available through web-based dashboards and data exports and aggretation are also supported. Manufacturers also work to ensure their resistance determinations align as closely as possible with the latest 'WHO Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance' so that results are standardized across testing sites. Lastly, sequencing implementation budgets should consider costs for integrating sequencing data into laboratory information and surveillance systems that are already established and operational for timely programmatic monitoring of testing performance and quality, as well as action for disease control.

12) What additional steps should be taken for quality-assured implementation of NGS?

Establishing and routinely monitoring and analyzing a comprehensive quality assurance program for NGS with clear performance and quality indicators is essential to ensure high-quality sequencing services are delivered. Quality assurance program elements should include technical and quality documentation, training and competency assessment for testing staff, instrument verification, equipment maintenance, method validation and verification, quality control, and post market validation of reagents with program oversight. In addition, as with any surveillance or clinical testing method, NGS procedures should be evaluated through enrolment in at least one external quality assurance program. The WHO global culture and drug susceptibility testing EQA program available to TB Supranational Reference Laboratories (SRLs) includes well-characterized strains with known sequencing profiles that may be used by SRLs and propagated for network NRLs for this purpose. In addition, the GLI Practical Guide for Laboratory Strengthening and the recently-released WHO 2023 Use of next-generation sequencing for the surveillance of drug-resistant tuberculosis: an implementation manual describe general quality assurance program recommendations, while the WHO manual specifically describes development of an quality assurance and quality control programs specific to sequencing.

13) How should turnaround time (TAT) and cost be considered when implementing NGS?

The threshold for acceptable turnaround times (TATs) may differ depending on the application of sequencing. For clinical use, TATs should be as short as possible (i.e., < 1-2 days) to provide actionable DST results for patient treatment, while longer TATs may be acceptable if NGS is employed for routine surveillance purposes or as part of a national TB drug resistance survey.

In terms of cost, factors such as the sequencing method, existing equipment availability and type, reagent consumption rate, and bioinformatic and IT requirements access and support, may all be influential. Cost and TAT for NGS can be inversely related in that cost may be reduced by batching samples; however, TATs may increase as a result. This highlights the

importance of disease programs clearly establishing priorities for sequencing data use that consider available and future resources prior to finalizing plans for sequencing technology selection and implementation.

14) Which laboratories are suitable to implement NGS for TB?

NGS requires a high-level of technical skill with specialized instrumentation, the appropriate laboratory infrastructure, practices, and protective equipment for safely working with TB cultures or clinical specimens, and availabilty of dedicated, unidirectional, space-separated, and contamination-free molecular workflows. Leveraging previous investments in sequencing expertise and intrastructure is often a cost-effective and efficient strategy for NGS implmentation. As such, centralized public health laboratories, including regional, national, or highly specialized public health or safe and fully operational TB culture reference laboratories should be considered as candidate TB NGS implementation sites.

15) What should be considered when assessing laboratory readiness for NGS?

Availability of essential laboratory components for sequencing, such as: infrastructure, equipment, biosafety, quality management systems, the technical experience of staff, training and compentancy assessment needs, inventory management, and the status of standardized technical and quality documentation required for pre-testing, testing-, and post-testing procedures should be evaluated. As with any readiness evaluation, gaps identified should be rectified ahead of implementation. A laboratory readiness assessment checklist is available in the WHO 2023 Use of next-generation sequencing for the surveillance of drug-resistant tuberculosis: an implementation manual.

16) What are the key considerations for coutries implementing NGS for DR TB surveillance?

National TB program and country priorities for DR TB surveillance should be used to inform sample and site selection strategies, placement of NGS into DST algorithms, and development of data analytic plans that support country goals and maximize data use for disease control. Additional country considerations may include budgeting to incorporate NGS into diagnostic algorithms, integration of sequencing data into existing surveillance systems, establishment of an NGS Technical Working Group (TWG) and updating Strategic National Plans to include NGS for DR TB surveillance.

17) What mechanisms are in place to better support low and middle income countries to have access to sequencing supplies?

The release of WHO policies on tNGS use for drug resistant TB clinical care and surveillance will enable the Global Drug Facility to include these products in their Diagnostics Catalogue for regular global supply at negotiated pricing (in progress). Furthermore, manufacturers of tNGS

products are establishing strategic partnerships with global and regional leaders involved in *in vitro* diagnostic supply provision to help support global access to their products.

18) Where can I find the latest global recommendations and resources to support implementation of NGS for drug-resistant tuberculosis (DR TB)?

WHO consolidated guidelines on tuberculosis: module 3: diagnosis: rapid diagnostics for tuberculosis detection, 3rd ed - This latest edition includes new recommendations on the use of a new class of diagnostic technologies: targeted next-generation sequencing (NGS) tests for the diagnosis of drug-resistant tuberculosis (TB). The recommendations provide a novel approach for the rapid detection of drug resistance to new anti-TB drugs using the latest technologies.

WHO operational handbook on tuberculosis: module 3: diagnosis: rapid diagnostics for tuberculosis detection, 3rd ed – This latest edition replaces the one issued in 2021. The targeted next generation sequencing solutions endorsed by WHO for drug resistant tuberculosis detection are described in presented as a new class of technologies in this document.

"Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance" - This 2023 update to the original catalogue supports TB laboratories to interpret genomic sequencing results with a more recent reference dataset and may be used to guide development of new molecular drug susceptibility tests.

"The use of next-generation sequencing for the surveillance of drug-resistant tuberculosis: an implementation manual" – Released in 2023, this document provides practical guidance on planning and implementing NGS technology for characterization of drug resistance within the *Mycobacterium tuberculosis* complex, complementing WHO's Guidance for the surveillance of drug resistance in tuberculosis that was released in 2021.

Rapid Communication – Released by WHO in 2023, this document announced the use of targeted NGS solutions to aid in the detection of DR TB, in advance of the more detailed recommendations that included in the latest 2024 WHO consolidated guidelines on tuberculosis: module 3: diagnosis: rapid diagnostics for tuberculosis detection, 3rd ed.

Wet Lab Questions

19) What steps are included in an NGS workflow?

NGS for DR TB is classified by WHO as a high-complexity testing method with similar workflows between technologies and approaches. The NGS workflow consists of four main phases with multiple steps in each phase. These four phases include: (i) genomic DNA (gDNA) extraction from a culture isolate or clinical sample (depending on the type of technology used), (ii) library preparation, (iii) next generation sequencing, and (iv) bioinformatic analysis. The first phase, gDNA extraction, is completed using TB culture isolates for WGS, while tNGS may be performed from TB culture isolates or clinical samples (such as sputa). The turnaround time for the NGS workflow from sample processing to sequencing data is approximately 2-3 weeks for WGS (Mtb cultures) and 2-3 days for tNGS (patient samples). Bioinformatic analysis can be completed using custom analytics pipelines or open-source or proprietary software solutions. Both proprietary and open-source analytics solutions can be used to identify genomic variants and interrogate long or short read sequences for the public health applications described above. For more information on NGS workflows, please see the WHO 2023 Use of next-generation sequencing for the surveillance of drug-resistant tuberculosis: an implementation manual and the WHO/ GLI tNGS Information Sheets provided in Annex 4 of the 2024 WHO operational handbook on tuberculosis: module 3: diagnosis.

20) Can we use the same sequencing reagents with different instruments?

Usually, the reagent kits for DNA sequencing reactions are technology specific. However, the library preparation kits offered by a given manufacturer can often be used across several different NGS instruments from that manufacturer. Most DNA-purification kits are compatible with different NGS systems. There are also library prep kits suitable for sequencing on instruments of different manufacturers. Manufacturer guidance for each sequencing instrument should be consulted to review which reagents are compatible with the available technologies and workflows being considered for use at each testing site.

21) Which sample types can be used for TB sequencing?

In short, both sputum sediments and liquid and solid TB culture isolates can be used for TB sequencing. Sanger sequencing and WGS require the use of TB culture isolates, while the WHO-recommended tNGS solutions may be done using either sputum sediments or culture isolates.

Per WHO recommendations for use of the tNGS solutions for clinical reporting, DNA should be extracted from samples provided by people with bacteriologically confirmed and rifampicin-resistant pulmonary TB. Consult the latest WHO consolidated guidelines on tuberculosis: module 3: diagnosis: rapid diagnostics for tuberculosis detection, 3rd ed for details on patient and sample eligibility criteria for tNGS for clinical use.

22) In the algorithm for tNGS, is it required to collect new fresh specimens?

Collection of new specimens is not always necessary, as leftover sediment may be used for DNA extraction and tNGS, depending on the sequencing technology type and algorithm adopted.

Usually, one specimen is collected for mWRD testing. If sequencing and phenotypic testing are performed in the same laboratory, one additional specimen may be sufficient. If testing is performed in two laboratories, two specimens should be collected, and the sequencing and phenotypic testing should be conducted in parallel. As always, sample management, including safe packaging and timelines for referral should be considered against WHO biosafety and manufacturer sample management guidance when designing testing network strategies. The latest WHO operational handbook on tuberculosis: module 3: diagnosis: rapid diagnostics for tuberculosis detection, 3rd ed may be consulted for information on sample testing.

23) Is a protocol available to extract DNA from samples other than sputa for TB tNGS?

Globally recommended standard operating procedures for DNA extraction and purification from specimens other than sputum for downstream NGS are not currently available. The manufacturers of the WHO-approved tNGS solutions suggest their validated procedures in the corresponding user manuals.

Researchers are trying to identify the best performing extraction procedures. The main challenge with optimization of these procedures is the paucibacillary nature of extrapulmonary sample types, such as biopsy material. While, targeted sequencing for DR TB using stool samples has been described (Rapid molecular diagnostics of tuberculosis resistance by targeted stool sequencing | Genome Medicine | Full Text (biomedcentral.com)) further research and evidence will be needed to inform review and global policy updates.

24) Are there sequencing reagents that can be used to sequence both RNA and DNA?

Usually, the reagent kits used for sequencing can be used to sequence DNA and cDNA (complementary DNA produced by reverse transcription of messenger RNA, or mRNA). However, the library preparation kits required to prepare genomic material for downstream NGS differ for DNA and RNA.

25) What is the minimum concentration of DNA needed for sequencing?

The DNA input concentration is test-specific, with the range provided by each manufacturer for efficient library preparation and sequencing reaction steps. However, two examples are highlighted below.

For Illumina DNA library preparation: A current manufacturer protocol recommends DNA inputs of 1–500 nanogram (ng) or higher. For small genomes, such those from microbes, the DNA input amount may be reduced as low as 1 ng.

For ONT DNA library preparation: A current Rapid sequencing DNA V14 protocol recommends an input of 100 ng gDNA and. However, a current Rapid Barcoding Kit 24 V14 protocol recommends an input of 50 ng gDNA. Both protocols are optimized for samples which contain long fragments of DNA (>30 kb).

26) What is the difference between a NanoDrop spectrophotometer and a Qubit Fluorometer? Are both needed for next generation sequencing or can only one be used?

NanoDrop (ThermoFisher Scientific) spectrophotometers (and other UV absorbance methods and instruments) are used to measure the amount of total nucleic acids in a sample. These differ from fluorometers, such as the Qubit (ThermoFisher Scientific), which precisely quantify the specific amount of double-stranded DNA. This dsDNA quantification method enables quick detection of samples with both low and high amounts of dsDNA, and is designed to specifically distinguish dsDNA from ssDNA, RNA, protein, and free nucleotides.

27) Is it important for laboratories preparing to implement NGS to complete a new method validation first?

Yes. Completing a new method validation using a panel of well-characterized Mtb isolates or clinical primary samples will help ensure all steps involved in sample preparation, manipulation, and NGS – as well as the test kits, supplies, commodities, and equipment – are validated to accurately and precicely detect the presence and absence of genomic variants and assure quality sequencing results are obtained by specific sites and testers. New method validation data will also be required for international accreditation of sequencing procedures used for clinical reporting.

28) What quality control (QC) checks should be in place when performing NGS?

Quality control should be applied at all steps through the NGS procedure, as would be done for introduction, implementation, and monitoring of any surveillance or clinical method. Specific to NGS, quality control checks should minimally be performed to ensure appropriate quality of: primary samples and gDNA extractions, library preparations, sequencing runs, and post-sequencing data analytical assembly metrics.

Additional Wet Lab Resources for Next Generation Sequencing of DR TB

- WHO/ GLI Practical manual on tuberculosis laboratory strengthening, 2022 update – This document provides the latest practical guidance on use of newly recommended diagnostics as well as guidance in key technical areas.
- Illumina GenoScreen Deeplex® Myc-TB Assay with LP and Indexes - The Illumina and GenoScreen Deeplex® Myc-TB Combo Kit is a targeted next generation sequencing panel for the identification of *Mycobacterium tuberculosis* complex (MTBC) strains, mycobacterial species identification, genotyping, and prediction of drug resistance mutations of MTBC strains.
- TBseq Ultra 3.0- MTB/NTM identification & Drug resistance genes testing-Shengting Medical Technology Company Limited - TBseq Ultra 3.0- MTB/NTM identification & Drug resistance genes detection.
- Oxford Nanopore Technologies Products – Lists sequencing preparation kits and devices and provides links to region and country contacts.

Data Analysis Questions

29) Are data storage requirements different for sequencing than for other mWRDs?

Sequencing produces more data than other mWRDs. NGS data require immediate, short-term, and long-term data storage solutions. Immediate solutions include instrument-based data storage (an Illumina MiSeq can save several sequencing runs of data) and computer hard drives. Short-term storage solutions may include continued use of computer and external hard drives of sufficient data capacity and/ or local servers. Additionally, the following solutions offer both short- and long-term storage of small and large amounts of NGS data with variable associated costs: Illumina BaseSpace, Google Cloud Platform (GCP) and Amazon Web Services (AWS). For more information on NGS data management and storage, please see the The use of next-generation sequencing for the surveillance of drug-resistant tuberculosis: an implementation manual.

30) Are there any free-of charge, automated, bioinformatic tools or training courses available for analysis of *M. tuberculosis* NGS data?

a. Options for free, automated, publicly available bioinformatic pipelines include but are not limited to:

MTBSeq – For analysis of Illumina WGS data from MTBC isolates for mapping, variant calling, detection of resistance, and comparative analysis to calculate phylogenetic tree.
Github: https://github.com/ngs-fzb/MTBseq_source

TB Profiler – Analyzes Illumina or Nanopore WGS data from *M. tuberculosis* to predict drug resistance and type strains (lineage)
Github: https://github.com/jodyphelan/TBProfiler
Web Version: https://tbdr.lshtm.ac.uk/

Mykrobe – Analyzes Illumina or Nanopore WGS data from *M. tuberculosis*, *S. aureus*, *S. sonnei*, *S. typhi* for drug resistance prediction
Github: https://github.com/Mykrobe-tools/mykrobe
Website: https://www.mykrobe.com/

PhyResSE – Analyzes Illumina or Ion Torrent WGS data from *M. tuberculosis* to predict drug resistance and type strains (lineage)
Website: https://bioinf.fz-borstel.de/mchips/phyresse/

ResFinder – Identifies antimicrobial resistance genes from bacterial sequences
Github: GitHub - cadms/resfinder

genTB – Translational genomics tool for M. tuberculosis that analyses sequence data to predict and map drug resistance and mutations
Github: https://github.com/farhat-lab/gentb-site
Website: Translational Genomics of Tuberculosis (harvard.edu)

MAGMA - The MAGMA pipeline for comprehensive genomic analyses of clinical Mycobacterium tuberculosis samples

Github: https://github.com/TORCH-Consortium/MAGMA/

b. Free, online webinars and hands-on tutorials and training for analyzing *M. tuberculosis* NGS data using the Galaxy platform can be found here, Mycobacterium tuberculosis NGS made easy: data analysis step-by-step (gallantries.github.io) including mapping and variant calling using Illumina short-read sequences, cluster analysis, development of phylogenetic trees.

31) Data security?

IT requirements when analyzing and storing sequencing data from NGS applications should include appropriate data security measures such as access control, development of standardized procedures and systems for data security, sharing and ensuring confidentiality, regardless of the data storage and analysis solution adopted (e.g., local servers, cloud).

32) Internet Speed?

An internet network and a stable, high-speed internet connection (at least 10 Mbps upload speed) are critical components of sequencing data transmission and analysis. Certain storing solutions such as cloud-based solutions are only accessible via a stable, high-speed internet connection during upload and download. Slow and routinely disrupted internet services may impact data availability for analysis and result use.

33) How can the WHO Mutations Catalogue be used as reference for countries implementing next generation sequencing?

Sequencing programs may compare the mutations they identify that may be associated with TB drug resistance to those listed and classified within the latest version of the Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance, 2nd ed. Based on the breadth of the WHO dataset and standardized, vetted analytics approach, it is advisable that publicly available bioinformatic or in-house pipelines incorporate or consider the WHO Mutations Catalogue for result interpretation. However, when reviewing DR TB mutation results, each setting should account for any known local *M. tuberculosis* lineage and drug-resistant strain differences that may differ or add information to global analyses presented by the WHO Mutations Catalogue.

Lastly, the existing WHO-recommended tNGS solutions for clinical DR TB testing interrogate the Catalogue of mutations in Mycobacterium tuberculosis complex and their association with drug resistance, 2nd ed to identify mutations associated with resistance or susceptibility to anti-tuberculosis drugs.

34) Which reference genome should be used for next generation sequencing phylogenetic analyses?

Sequencing data may be used to investigate and measure strain relatedness through phylogenetic analyses. Bioinformatic analytic pipelines often map genomic sequencing reads to the *M. tuberculosis* reference strain H37Rv genome and then call those that vary in sequence as genomic variants. These pipelines may also produce tables of genomic nucleotide differences between the reference and variant genomes as Single Nucleotide Polymorphisms (SNPs), Insertions/ deletions (Indels), or frameshifts. This is the classical approach for identifying variants that can be used in relatedness, or phylogenetic, analyses that visually represent the strain data based on the number of differences between them.

However, as H37Rv is a *M. tuberculosis* lineage 4 strain, its use as a reference for other lineages may be inappropriate due to gene content differences between lineages and lineage-specific nucleotide sequences not found in other strains. To address these challenges, alternative approaches have been explored, including use of (i) a pangenome as a reference strain that incorporates the entire gene pool of MTBC lineages, (ii) an inferred ancestral genome representative of the MTBC population, or (iii) ad hoc lineage-specific reference genomes, depending on analysis objectives (reference: Whole genome sequencing of Mycobacterium tuberculosis: current standards and open issues | Nature Reviews Microbiology). In general, any replacement of H37Rv as the reference genome should be assessed by in silico studies across data sets and clinical settings before use for clinical DR TB strain classification or reporting purposes.

35) Is it mandatory for the sequencing personnel to be bioinformaticians as well?

Depending on the NGS intended use, applications and workload, programmes or laboratories implementing NGS should develop terms of reference and position descriptions related to bioinformatics analyses.

In short, molecular testing personnel are typically not required to be bioinformaticians. However, molecular biologists involved in sequencing should receive basic training on post-sequencing processes. These processes may include how to demultiplex and convert raw sequencer data files to FASTQ format for downstream analysis and how to use user-friendly or commercial NGS data analysis tools. Depending on the specific context, these same individuals could be trained specifically on using bioinformatics tools.

In contrast, bioinformatics officers are typically trained on text-based command line interfaces, such as UNIX or Linux operating systems, and on use of software programs commonly associated with NGS. These individuals will be best positioned to adapt existing bioinformatics pipelines for NGS sequence data analysis to meet local needs and identify the best computation tools and analytical methods available to manage the sequencing data (i.e., handling, storage). It is helpful if these individuals also receive basic training in genomic sequencing technologies

and testing procedures so that they are fully versed in the methods used within their programs to produce the data they are analyzing.

A NGS facility dealing with large amounts of data generated by NGS (particularly WGS) should have bioinformaticians in the staff to guide and provide oversight for sequencing data management within the program.

Glossary of Terms

From the The use of next-generation sequencing for the surveillance of drug-resistant tuberculosis: an implementation manual.

Amplicon – A specific fragment of DNA that is replicated millions of times using the polymerase chain reaction (PCR) amplification process. Amplicons are the starting material for library preparation during targeted next-generation sequencing (tNGS).

Bioinformatics pipeline – A series of analyses, algorithms and programs executed in a predefined sequence to process, analyze, interpret and report NGS results.

Breadth of coverage –The percentage of the intended reference genome for which the genomic positions (bases) were sequenced with minimal selected coverage.

Depth of coverage – The number of sequencing reads at a given position in the sequenced genome. The more reads there are, the more confidence there is that the sequence is accurate.

DNA library – The processed sample material that serves as the input material for NGS. A DNA library is obtained by fragmenting and sorting DNA to obtain fragments of a predefined length and attaching oligonucleotide adaptors to the ends of the fragments to enable tNGS or whole genome sequencing (WGS).

FASTQ format – A text-based, raw data format for storing nucleotide sequence information and corresponding quality scores (i.e., Phred scores).

Next-generation sequencing – A high-throughput sequencing method used to determine the nucleotide sequence of a genome in a single biochemical reaction. NGS is performed by non-Sanger-based sequencing technologies that can sequence multiple DNA fragments in parallel; the sequences are then assembled and mapped to a reference genome using bioinformatics analyses.

NGS laboratory workflow – A series of laboratory procedures required to generate raw NGS data from a sample. The workflow typically includes sample processing, DNA extraction, fragmentation or amplification, and preparation and sequencing of the library.

Phred score – A read editing program called Phred assigns a quality score to each base identified during sequencing, which is equivalent to the probability of error for that base. All NGS manufacturers use Phred scores as the measure of sequence quality reporting.

Reference genome – The validated and published sequence of a known genome, gene, or artificial DNA construct. A sequence produced by an NGS instrument may be aligned to a reference sequence to assess NGS accuracy or to find nucleotide changes (mutations).

Sanger sequencing – Technique for DNA sequencing based on the incorporation of chain terminating dideoxy-nucleotides by DNA polymerase during in vitro DNA replication.