Table of Contents

Advanced Genomic Techniques for Tuberculosis Diagnosis

Data analysis

Module developers

Elizabeth Streicher - Module lead and developer
Anzaan Dippenaar - Module developer and instructor

Learning Outcomes

By the end of this module, participants will be able to:

1. Comprehensively determine drug resistance profiles and identify the lineage of *Mtb* strains based on tNGS Illumina data and ONT WGS data.

ONT data analyis

Background

***Mtb* WGS Analysis with ONT Data:**

In the realm of *Mtb* research, WGS has emerged as a transformative tool for molecular epidemiology and drug resistance profiling. Particularly notable is Oxford Nanopore Technologies (ONT), a sequencing platform esteemed for being portable, having low-cost hardware and generating long reads that encapsulate comprehensive genomic information in a single sequence.

ONT WGS Analysis Pipeline: Custom Approach:

ONT WGS data analysis entails a series of pivotal steps to transform raw sequencing data into actionable insights. This custom pipeline generally encompasses procedures like quality assessment, aligning reads to a reference genome (*Mtb* H37Rv), calling variants, and annotating genomic changes. The advantage of ONT's long reads lies in their ability to disentangle complex genomic regions, repetitive sequences, and structural variations, which substantially enhances the precision of mutation identification and strain differentiation.

Introducing TB-Profiler: Automating WGS Analysis:

To streamline the analysis of ONT WGS data while accommodating both Illumina and ONT data formats, we will adopt the TB-Profiler pipeline. TB-Profiler is a purpose-built automated analysis tool, tailored to *Mtb* genomic data. Its role is to facilitate the WGS analysis process, from preliminary data preparation to deciphering strain types and predicting drug resistance profiles.

Key Steps in the TB-Profiler Pipeline:

1. **Genome Alignment:** The pipeline orchestrates the alignment of sequencing data to a reference *Mtb* genome to unearth variations and mutations.
2. **Variant Calling:** TB-Profiler's forte lies in identifying SNPs and other genetic variations within the aligned data.

3. **Strain Typing:** A pivotal objective of TB-Profiler is to decode the strain type of the *Mtb* isolate, offering insights into lineage and potential transmission dynamics.
4. **Drug Resistance Profiling**: TB-Profiler scrutinizes the genomic data, pinpointing mutations linked to drug resistance, crucial for shaping informed treatment strategies.
5. **Report Generation:** The pipeline culminates by generating comprehensive reports that encapsulate strain typing, drug resistance patterns, and other pertinent genomic attributes.

Bioinformatics on the Command Line and Introduction to Galaxy:

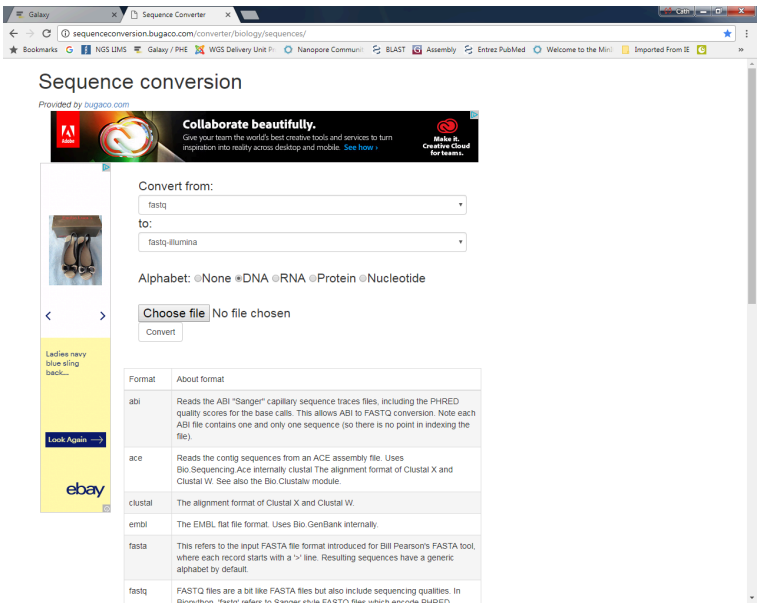
Traditionally, bioinformatics analyses are often executed via command line interfaces, where researchers input commands to navigate complex analyses. This approach, while powerful, requires familiarity with specific commands and tools, which can be a steep learning curve.

To simplify this process, we will employ Galaxy, a user-friendly web-based platform designed to democratize bioinformatics analysis. Galaxy offers a suite of preinstalled tools and workflows, enhancing accessibility for researchers regardless of their bioinformatics background. Through Galaxy, we'll navigate complex analyses without needing to delve deeply into command line intricacies, providing a more intuitive and efficient approach for ONT WGS data analysis.

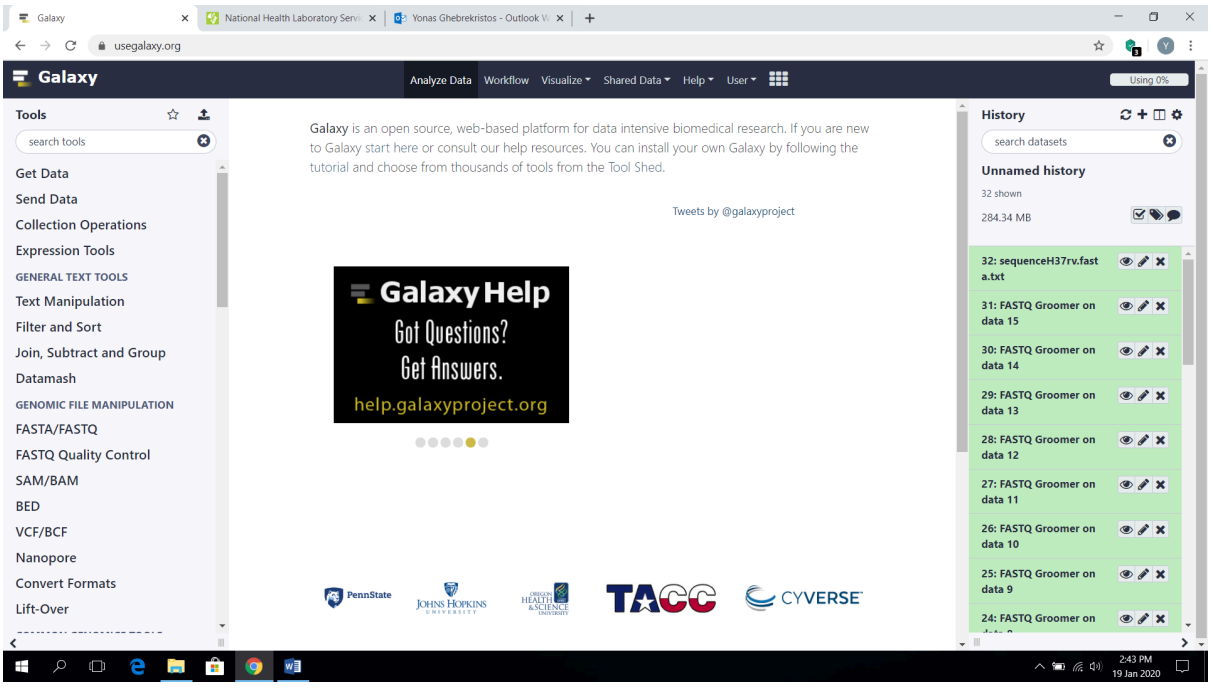
Converting ONT data to fastq

This conversion has already been done for the sake of time during the practical. Please refer to the steps below on how this can be done.

1. To convert your MinION fastq sequences to a suitable format for Galaxy analysis, navigate to sequenceconversion.bugaco.com in your browser. Choose the fastq files from the MinION pass folder and convert from fastq to fastq-illumina. These will be written to your download folder. The files will now be labelled sample (x).fastq-illumina, where x is the number corresponding to the order you converted them in the sequence converter.



2. Go to Galaxy and register if you haven't already done so at https://usegalaxy.org/
3. To upload your converted sequences into Galaxy, in the left-hand column, select Get Data then Upload File from your computer.

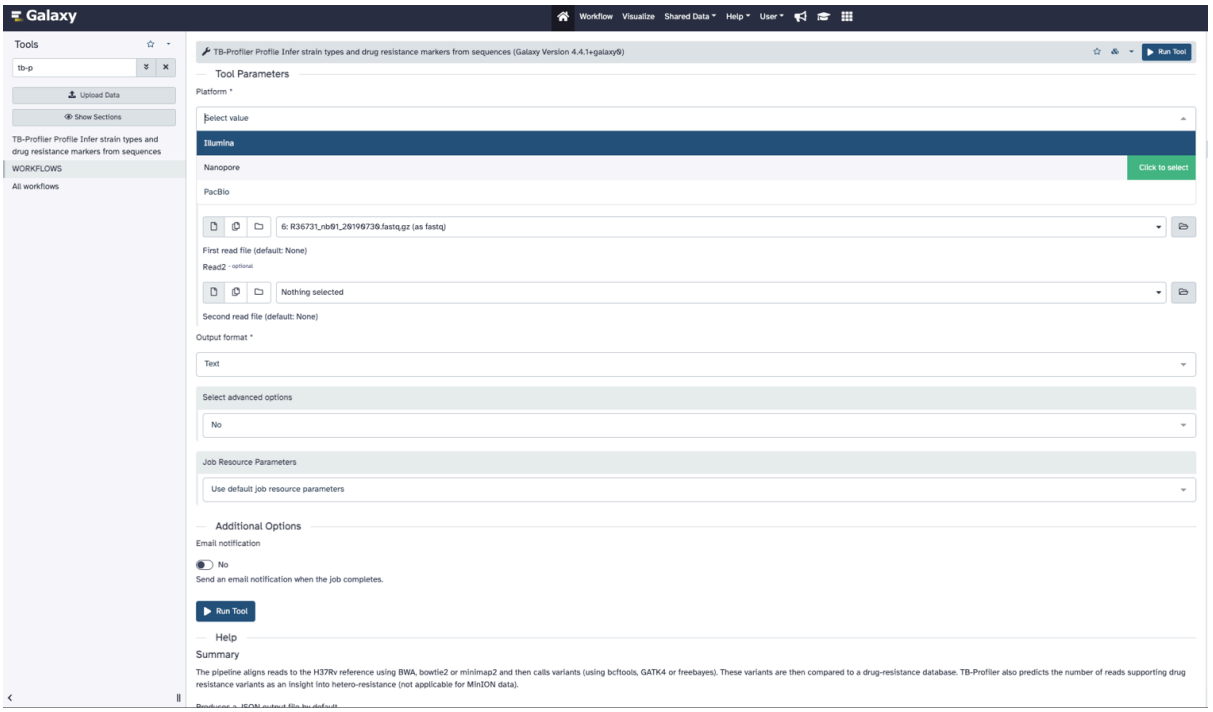


4. Select Choose local file and navigate to your downloaded sequences, select them all and select Start. Your files will now be uploaded and will appear in green on the right-hand column of your screen when they are ready (see above). Also download H37Rv on your desktop; this is the reference genome sequence you will be using to map your MinION sequences to.

a. Link to H37Rv reference genome: https://www.ebi.ac.uk/ena/browser/api/fasta/AL123456.3?download=true

Running TB-Profiler in Galaxy

1. In the "Tools" pane, search for TB-profiler



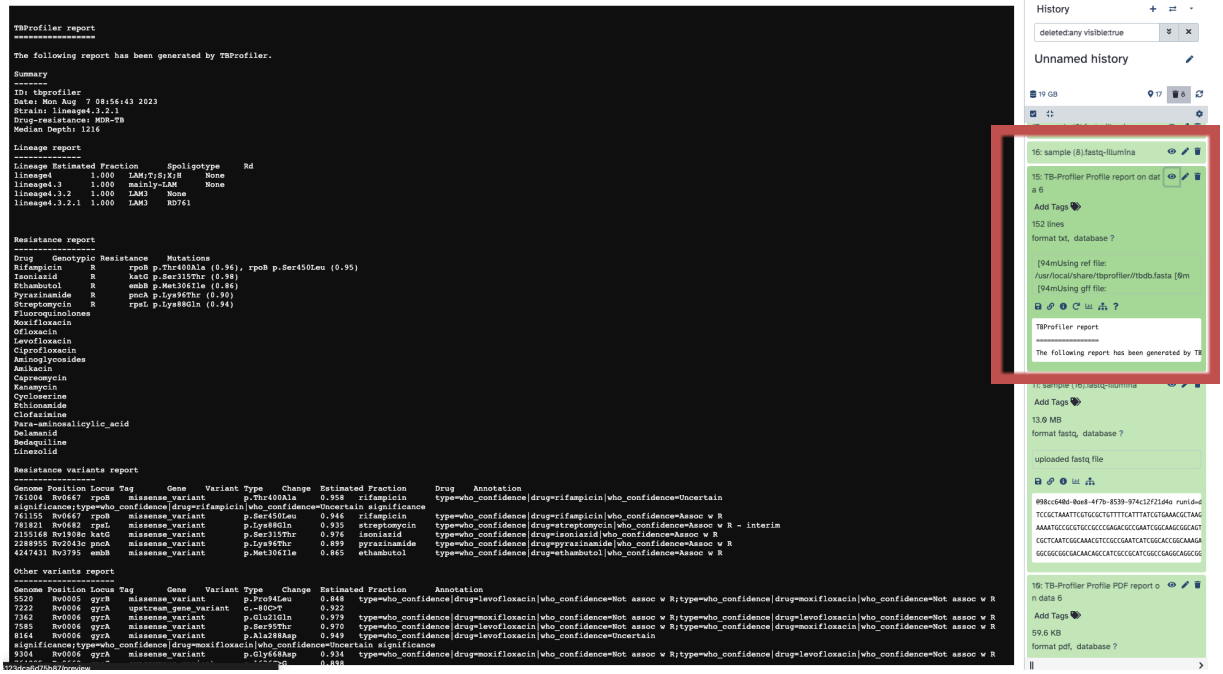
2. Choose the tool parameters appropriate to the analysis needed:
    a. Platform: Nanopore
    b. Input File Type: Single Fastq
    c. Output format: text (the report in text format is more detailed compared to the slimmed-down, easy to interpret PDF report)

The minimum depth and reporting allele frequencies can be customised under "Select advanced options" – 'Yes'

3. Click on "Run Tool" and wait or the analysis to finish

You can track the progress of the analysis in the pane on the right-hand side of the Galaxy window.

4. Click on the eye icon in the history pane under the "TB-profiler Profile report on data …" to view the TB-profiler report in text format



Interpreting TB-Profiler Reports and Output Files

TB-Profiler generates comprehensive reports that encapsulate critical genomic insights regarding *Mtb* isolates. These reports offer a wealth of information, aiding in the understanding of strain characteristics and drug resistance profiles. Here's a concise guide on how to interpret the TB-Profiler output files:

1. **Strain Typing:** TB-Profiler identifies the strain lineage of the *Mtb* isolate based on its genomic characteristics. This information sheds light on the evolutionary context and potential transmission dynamics of the isolate.
2. **Lineage-Specific Mutations:** The report highlights specific mutations associated with the identified lineage. These mutations are genetic markers that differentiate different *Mtb* lineages and offer insights into the evolutionary history.
3. **Drug Resistance Profiling:** One of the core features of TB-Profiler is its ability to predict drug resistance based on genomic variations. The report will detail mutations linked to resistance against various anti-TB drugs, enabling informed treatment decisions.
4. **Mutation Consequences:** The report often includes annotations explaining the functional impact of identified mutations. These annotations help elucidate how a mutation may affect the protein structure and function, potentially leading to drug resistance.
5. **Resistance Categories:** TB-Profiler categorizes drug resistance predictions into different levels, such as "high confidence" or "moderate confidence." This provides an indication of the reliability of the prediction.

6. **Resistance Patterns:** The report may outline the overall drug resistance pattern of the isolate, indicating which drugs the isolate is predicted to be resistant to and which ones it is susceptible to. (i.e MDR, XDR etc.)
7. **Additional Insights:** Depending on the specific version and features of TB-Profiler, the report may provide additional insights, such as information on virulence factors, potential transmission clusters, and more. These additional insights are typically generated when running TB-profiler on the command line with specialised parameters.

Interpretation Challenges:

It's important to note that while TB-Profiler provides valuable predictions and insights, interpretation may still present challenges. Mutations don't always guarantee drug resistance, and additional factors, such as gene expression and clinical history, must be considered. Additionally, new and emerging resistance-conferring mutations may not be well-characterized in the reference database used by TB-Profiler.

Integration with Clinical Context:

Interpreting TB-Profiler reports is most effective when done in conjunction with clinical context. Collaborating with medical professionals who understand the patient's history, treatment regimen, and overall health is crucial to making informed decisions based on the genomic insights provided by TB-Profiler.

OPTIONAL: INSTRUCTIONS TO RUN TB-Profiler ON THE COMMAND LINE:

TB-profiler is already installed on the virtual box.

First activate the Conda environment to enable TB-profiler:

```
conda activate tb-profiler
```

To get usage information in TB-profiler, run the program without any arguments:

```
tb-profiler
```

To get usage information on a specific tool within TB-profiler, run the program, specify the tool, and use the -h argument to access the help page:

```
tb-profiler profile -h
```

If in doubt, refer to the TB-profiler manual online on the Github page.

Create a directory where you want the TB-profiler output to be stored:

```
mkdir /path/to/my_TBP_output_directory
```

Change directories to the output directory that you have just created:

```
cd /path/to/my_TBP_output_directory
```

- To profile one isolate run the following command:

```
tb-profiler profile --nanopore sampleID.fastq.gz --prefix
sampleID -t 8 --txt
```

Refer to the manual for additional arguments. The arguments provided here include:

- --nanopore : specify the type of WGS data (default is Illumina).
- -p or -prefix : prefix, this is the name of the output file and should correspond to the isolate ID.
- -t : the number of threads to use on the server and translates to the computational resources that will be assigned to perform the task.
- - -txt : include the results in a plain text format, as opposed to the default output in Jason format.

To run TB-profiler on multiple isolates, either create a bash script listing the same command described above for each isolate that you wish to analyse on a new line, or for advanced users, use a for loop command like this:

```
for file in *.fastq.gz; do

    tb-profiler profile --nanopore $file --prefix
${file%.fastq.gz}_output -t 8 –txt

done
```

Interpretation of Xpert assay results

Overview

The Xpert MTB/RIF and Xpert MTB/XDR assays are automated molecular tests designed to detect the presence of Mycobacterium tuberculosis (MTB) complex DNA and identify mutations associated with drug resistance directly from clinical specimens. The MTB/RIF assay targets the rpoB gene, which is associated with rifampicin resistance, while the MTB/XDR assay extends this capability to additional drugs by targeting multiple genetic regions.

Xpert MTB/RIF assay interpretation:

1. **MTB Detection:**
    - **Positive Result:** Indicates the presence of MTB complex DNA in the sample.
    - **Negative Result:** Indicates the absence of detectable MTB complex DNA in the sample.
2. **Rifampicin Resistance (RRDR region of *rpoB* gene):**
    - **Detected:** Mutations associated with rifampicin resistance are present. This suggests the MTB strain is likely resistant to rifampicin.
    -

- **Not Detected:** No mutations associated with rifampicin resistance are present. This suggests the MTB strain is likely susceptible to rifampicin.
- **Indeterminate:** The assay did not yield a conclusive result for rifampicin resistance. Repeat testing or additional diagnostic methods may be required.

Xpert MTB/XDR assay interpretation:

1. **MTB Detection:**
   - **Positive Result:** Indicates the presence of MTB complex DNA in the sample.
   - **Negative Result:** Indicates the absence of detectable MTB complex DNA in the sample.
2. **Drug Resistance Detection:**
   - **Rifampicin Resistance (RRDR region of rpoB gene):**
     - **Detected:** Mutations associated with rifampicin resistance are present.
     - **Not Detected:** No mutations associated with rifampicin resistance are present.
     - **Indeterminate:** The assay did not yield a conclusive result for rifampicin resistance.
   - **Isoniazid, Fluoroquinolones, and Second-Line Injectables:**
     - **Detected:** Mutations associated with resistance to these drugs are present.
     - **Not Detected:** No mutations associated with resistance to these drugs are present.
     - **Indeterminate:** The assay did not yield a conclusive result for resistance to these drugs.

Key Points for Interpretation:

- **SPC (Sample Processing Control):** Validates the sample processing. If the SPC fails, the test is invalid, and a new sample should be tested.
- **PCC (Probe Check Control):** Ensures the integrity of the assay reagents. If the PCC fails, the test is invalid.

Clinical Context:

Interpretation of Xpert results should be done in conjunction with clinical evaluation and other diagnostic findings. Positive results for MTB detection and drug resistance should guide appropriate treatment decisions. Indeterminate results require follow-up testing.

Interpretation of Deeplex Myc-TB results

Overview

The Deeplex Myc-TB assay is a targeted next-generation sequencing (tNGS) test that provides comprehensive data on drug resistance mutations, strain typing, and phylogenetic information directly from clinical samples. It covers 15 anti-TB drugs and analyzes multiple genomic regions associated with resistance.

Deeplex Myc-TB data interpretation

1. **Drug Resistance Profiling:**
    - **Detected Mutations:** The report will list mutations associated with resistance to the 15 anti-TB drugs. Each mutation's presence indicates potential resistance to the corresponding drug.
        - **High-Confidence Mutations:** These mutations are well-characterized and strongly associated with drug resistance.
        - **Moderate-Confidence Mutations:** These mutations are less well-characterized but are still associated with potential resistance.
    - **Not Detected:** Indicates that no resistance-associated mutations were found in the analyzed regions for specific drugs.
    - **Heteroresistance:** Detection of both susceptible and resistant alleles within the same sample, indicating mixed bacterial populations.
2. **Strain Typing:**
    - **Lineage and Sublineage:** The report provides information on the lineage and sublineage of the MTB strain, helping to understand the strain's evolutionary context and potential epidemiological significance.
    - **Spoligotyping:** The assay includes data on spoligotype patterns, which can offer additional phylogenetic insights.
3. **Phylogenetic Markers:**
    - **Identification:** Deeplex Myc-TB opportunistically assesses phylogenetic markers that fall within the regions analyzed for drug resistance. This allows determination of the Mtb lineage and possibly sublineage.
    - **Accuracy:** While the assay can determine spoligotypes and lineages with fair accuracy, it doesn't offer the same high discriminatory power as whole-genome sequencing.
4. **Report Generation:**
    - **Detailed Analysis:** The report provides a detailed analysis of all detected mutations, including their potential impact on protein function and drug resistance.
    - **Confidence Levels:** Predictions are categorized by confidence levels, helping to assess the reliability of the resistance predictions.

Key Points for Interpretation:

- **Clinical Relevance:** Interpret the detected mutations in the context of clinical data and patient history. Some mutations may not confer complete resistance but could indicate reduced susceptibility.
- **Additional Testing:** In cases where results are ambiguous or unexpected, additional testing or confirmatory methods may be necessary.
- **Bioinformatics Support:** Use the Deeplex-Myc TB online data analysis platform for comprehensive interpretation, leveraging the included analysis tokens.

Clinical Context:

Deeplex Myc-TB provides extensive data that should be interpreted alongside clinical findings and other diagnostic results. The detailed resistance profile and strain typing information can guide effective treatment regimens and inform public health strategies.

Practical exercise

Clinical scenario:

You are a clinical microbiologist employed at a regional laboratory. The medical superintendent of a rural district hospital seeks your expertise due to concerns about a nurse in the medical ward who has recently been diagnosed with multi-drug-resistant tuberculosis (MDR-TB) through GeneXpert (Xpert MTB/RIF and Xpert MTB/XDR) tests. The superintendent suspects that the nurse might have contracted drug-resistant TB within the hospital premises, given the presence of multiple patients with drug-resistant TB in the ward over the past months. After examining ward records, it was revealed that four patients diagnosed with varying degrees of drug-resistant TB through Xpert testing had been admitted. *Mtb* cultures (from sputum specimens) were obtained from all these patients, and the isolated strains of *Mtb* are accessible. Your role involves using advanced molecular techniques to investigate whether this scenario indicates a potential instance of hospital-acquired infection and to assess the relatedness of the strains.

Aim:

To compare and analyze the results from three different molecular assays (Xpert, ONT, Deeplex Myc-TB) to investigate the potential hospital-acquired transmission of MDR-TB and to assess the genetic relatedness of the *Mtb* strains involved.

Complete the following table with the results obtained from the respective assays performed:

Sample ID	Drug	Xpert MTB/RIF	Xpert MTB/XDR	Deeplex Myc-TB	WGS (ONT MinION)
Nurse	Rifampicin				
	Isoniazid				
	FQ	N/A			
	SLIDs	N/A			
Patient 1	Rifampicin				
	Isoniazid				
	FQ	N/A			
	SLIDs	N/A			
Patient 2	Rifampicin				
	Isoniazid				
	FQ	N/A			
	SLIDs	N/A			
Patient 3	Rifampicin				
	Isoniazid				
	FQ	N/A			
	SLIDs	N/A			
Patient 4	Rifampicin				
	Isoniazid				

	FQ	N/A			
	SLIDs	N/A			

Discussion Points

1. **Drug Resistance Profiles:**
    - Compare the resistance profiles of the nurse's isolate and the patients' isolates. Are there any common mutations?
2. **Strain Relatedness:**
    - Assess the genetic relatedness of the isolates using the data from WGS. Do the phylogenetic markers indicate a common source or transmission link?

3. **Technique Comparison:**
    - Discuss the advantages and limitations of each assay (Xpert, Deeplex, WGS). Which method provided the most comprehensive data?
4. **Hospital-Acquired Infection:**
    - Based on the combined data, is there evidence to suggest that the nurse contracted MDR-TB from the patients in the ward? Provide a rationale for your conclusion.

Drug resistance in MTB:

Summary of the most common genes and mutations associated with Mtb drug resistance

Drug	Gene	Region of gene in genome	Most common mutations conferring resistance	Region of mutation
Rifampicin (RIF)	*rpoB*	759807-763325	**RRDR (codons 507-533)** 516 GAC-GTC 526 CAC-GAC 531 TCG-TTG	**761081-761162** 761109-761111 761139-761141 761154-761156
	RRDR sequence:	ggc acc agc cag ctg agc caa ttc atg gac cag aac aac ccg ctg tcg ggg ttg acc cac aag cgc cga ctg tcg gcg ctg		
Isoniazid (INH)	*katG*	2153889-2156111 (Transcribed on the negative strand)	315 AGC-ACC	21555169-2155167
	inhA promotor	1673415-1673439	C -15 T T -8 A	1673425 1673432
Fluoroquinolones (FQs)	*gyrA*	7302-9818	**QRDR (codons 74-113)** 90 GCG-GTG 94 GAC-GGC 95 AGC-ACC polymorphism that does not confer resistance	**7522-7625** 7569-7571 7581-7583 7584-7586
	QRDR sequence:	gcc cgg tcg gtt gcc gag acc atg ggc aac tac cac ccg cac ggc gac gcg tcg atc tac gac *agc* ctg gtg cgc atg gcc cag ccc tgg tcg ctg cgc tac ccg		
Aminoglycosides (AMI, KAN, SM)	*rrs*	1471846-1473382	A1401G	1473247

Pyrazinamide (PZA)	*pncA*	2288681-2289241 (Transcribed on the negative strand)	No hotspot – whole gene. Expert rules: nonsynonymous variants in *pncA* interpreted as PZA-R	NA
Ethambutol (EMB)	*embB*	4246514-4249810	306 ATG-ATC/GTG	4247429-4247431
Bedaquiline (BDQ)	*Rv0678* *atpE* *pepQ*	778990-779487 1461045-1461290 2859300-2860418	Not enough data	NA
Linezolid (LZD)	*rrl* *rplC*	1473658-1476795 800809-801462	Not enough data	NA