Public Health Ontario | Santé publique Ontario

TECHNICAL NOTES

Ontario Universal Typing of Tuberculosis Program by Whole Genome Sequencing

Updated: May 2026

Introduction

The Ontario Universal Typing of Tuberculosis Program by Whole Genome Sequencing (OUT-TB by WGS) was designed by Public Health Ontario (PHO) to assist public health units with tuberculosis (TB) case management, investigation, and surveillance.

As part of the OUT-TB Program, Whole Genome Sequencing (WGS) is performed on the first *Mycobacterium tuberculosis* (*M. tuberculosis*) complex isolate from each new case and repeated at three months if the cultures remain positive. The sequence data is then linked with laboratory data from Public Health Ontario's Laboratory Information Management System and case data from Ontario's Integrated Public Health Information System (iPHIS).

The combined data is presented in the OUT-TB by WGS interactive report and provides information on cases across public health units in Ontario which can be used to identify genetically related strains of *M. tuberculosis*, help confirm epidemiological links, and detect unsuspected transmission events.

Data Sources

- Whole genome sequencing results are extracted from the Public Health Ontario Whole Genome Sequencing Database.
- Specimen-specific data and drug susceptibility testing results are extracted from the Public Health Ontario Laboratory Information Management System.
- Case and epidemiological data for TB cases are extracted from the Integrated Public Health Information System.

Background

Mycobacterium tuberculosis Complex Lineages

Human-adapted *M. tuberculosis* complex (MTBC) is highly clonal and is classified into nine main phylogenetic groups, designated lineage 1 through lineage 9 (Table 1). These nine lineages demonstrate strong biogeographic associations that have been proposed to result from co-diversification with different human populations. Lineages 1-4, 7-9 are *M. tuberculosis* sensu stricto. Lineages 5 and 6 are largely limited to West Africa and are traditionally referred to as *M. africanum.* Other animal adapted strains from the MTBC (i.e., bovis, caprae, microti, mungi, orygis, pinnipedii, and suricattae) are also able to infect humans. The modern MTBC lineages 2, 3, and 4 are distributed worldwide while other lineages are restricted in their geographical spread.[1-3] Lineage 2 dominates in East Asia and is one of the most successful variants, with more than a quarter of global TB cases caused by this lineage.

Table 1: Lineage Designation of *Mycobacterium tuberculosis* Complex Variants

Complex member (variant)[4]	Lineage
M. tuberculosis var. tuberculosis	Lineage 1 – Indo-Oceanic Lineage 2 – East Asian / Beijing Lineage 3 – East-African-Indian Lineage 4 – Euro-American Lineage 7 – Horn of Africa / Ethiopia Lineage 8 – African Great Lakes region Lineage 9 – East Africa
M. tuberculosis var. africanum	Lineage 5, 6 – West Africa
M. tuberculosis var. bovis	La1 – bovis
M. tuberculosis var. bovis BCG	La1.2 – bovis BCG
M. tuberculosis var. caprae	La2 – caprae
M. tuberculosis var. microti	No designation
M. tuberculosis var. pinnipedii	No designation
M. tuberculosis var. canettii	Not applicable
M. tuberculosis var. mungi	No designation
M. tuberculosis var. orygis	La3 – orygis
M. tuberculosis var. suricattae	No designation
M. tuberculosis var. dassie	No designation
M. tuberculosis var. chimpanzee	No designation

Whole Genome Sequencing

WGS is a molecular sequencing method that looks into the makeup of an organism by reading most or all of its genetic code. MTBC possesses a DNA-based genome that is 4.4 million base pairs (Mb) in length and WGS is able to determine the order of nucleotides in approximately 90% of its genome. The OUT-TB Program uses WGS to provide an accurate genetic profile for each MTBC sample, enabling the ability to determine the genetic relatedness between samples.

This method provides greater resolution of genetically related clusters compared to polymerase chain reaction (PCR)-based genotyping methods such as Mycobacterial Interspersed Repetitive Units Variable Number Tandem Repeats (MIRU-VNTR) and spoligotyping, which when combined only looks at approximately 1% of the genome.

Single Nucleotide Polymorphisms

Single nucleotide polymorphisms (SNPs) arise from mutations resulting in the substitution of one nucleic acid for another at a specific position in the genome and they generally accumulate over time. The number of SNP differences between samples (SNP distance) can be used to determine if a case may have resulted from recent transmission. However, transmission cannot be ruled in or out based on matching or related genome sequences and requires interpretation within the context of epidemiological information.

The SNP distance between samples is determined by the method known as pairwise distance analysis. This method first compares each sample to a reference MTBC genome (H37Rv, NC_000962.3), and then the SNPs identified are compared between two samples in a pairwise manner and includes only high-quality SNPs that are present in both samples. It is important to recognize that samples considered matching by pairwise distance analysis may not have identical genomes. Furthermore, although SNP distances may be the same between samples for clusters containing more than two cases, the specific SNPs present in each sample may not be shared across all samples.

Below is an example of pairwise distance analysis performed on five cases using seven genome positions:

Figure 1a: Sequence Comparison

	Genome Position						
	10439	568917	1774532	2181447	2410979	3345489	3650739
Reference	A	T	C	G	A	A	T
Case 1	A	T	C	G	N	T	A
Case 2	A	T	C	G	T	T	A
Case 3	A	T	C	G	A	T	A
Case 4	A	A	C	T	A	T	A
Case 5	T	T	G	G	A	T	A

Figure 1b: Pairwise Distance Matrix

	Case 1	Case 2	Case 3	Case 4
Case 1				
Case 2	0			
Case 3	0	1		
Case 4	2	3	2	
Case 5	2	3	2	4

Genotyping

Genotyping refers to laboratory tests, including Mycobacterial Interspersed Repetitive Units (MIRU) Variable Number Tandem Repeats (VNTR) and Spoligotyping that permit the genetic comparison of the DNA from MTBC isolates from different cases of tuberculosis. When cases have genotypically identical isolates, it indicates that they may be part of the same TB disease transmission event. With the implementation of WGS, MIRU-VNTR and Spoligotyping are no longer available for prospective typing, however SM24 No. results for historical isolates are available on OUT-TB by WGS.

- Mycobacterial Interspersed Repetitive Units (MIRU) Variable Number of Tandem Repeats (VNTR) genotyping is a rapid PCR-based method used for genotyping of MTBC that looks at the number of repeat sequences in specific loci throughout the genome of each isolate. It has a higher discriminatory power than spoligotyping. The current international standard uses 24-loci (MIRU24) for the highest discriminatory power.
- Spoligotyping, or spacer oligonucleotide typing, is a rapid PCR-based method used for genotyping of MTBC. It looks for the presence or absence of spacers in a specific genome region which can be represented as a digital code known as the octal code. The overall discriminatory power of spoligotyping is less than that of MIRU-VNTR. However, it may provide strain lineage information and combined with MIRU-VNTR genotyping allows for improved discrimination between strains of many lineages.
- An SM24 No. is used to identify clusters of MTBC isolates with identical spoligotype (S) and MIRU24 (M24) genotypes. This number is unique to Ontario as there currently are no published international databases with strain designations using these combined methods. The SM24 No. is no longer available for prospective cases with WGS performed because the region of the genome targeted in MIRU-VNTR genotyping is not captured with WGS for analysis.

Inclusion Criteria

- The OUT-TB by WGS interactive report contains information about laboratory-confirmed, culture-positive cases of TB that are linked to case information in iPHIS. Clients with clinically confirmed or non-culture-positive cases are excluded. Clients with iPHIS records that did not meet the TB case definition are also excluded.
- Whole genome sequencing is performed on the first culture of MTBC isolated from a patient (DNA concentration of at least ≥1 ng/µL) and three months later if the cultures are still positive.
- Only isolates that are successfully sequenced and generate high quality sequence data (mean sequencing depth of ≥ 40x and genome completeness of ≥95%) are included.
- Only isolates linked to case information in iPHIS are included. Isolates are excluded if the linked iPHIS record did not meet the TB case definition (i.e., "DNM").

Methods

Testing Algorithm and Methods

- Tuberculosis cases are microbiologically confirmed through a combination of real-time PCR testing (BD MAX™ MDR-TB Assay), culture-based methods, and additional molecular or biochemical testing to differentiate *M. tuberculosis* complex from non-tuberculous mycobacteria (NTM). For more details on the diagnostic testing algorithm, please refer to Mycobacterium – MTBC and Rifampin/Isoniazid Resistance PCR and Mycobacterium – Culture.
- Phenotypic drug susceptibility is performed using the BACTEC™ MGIT 960 method and whole genome sequencing is used to provide genotypic predictions of antimicrobial resistance, lineage assignment, and MTBC complex member (variant) identification. For more information on drug susceptibility testing, please refer to Mycobacterium – TB Susceptibility Testing.
- Whole genome sequencing is performed using the Nextera XT Library Preparation Kit on Illumina Sequencing Platforms.
- PHO's TB genomics Pipeline uses a Nextflow based pipeline for analyzing whole genome sequencing data from *M. tuberculosis* genomes.
    - Samples undergo a series of sequential processes, including initial quality control assessments, read trimming (fastp v0.23.2), taxonomic classification and contaminant detection (kraken v2.1.2), dehosting of sequencing reads (BWA v0.7.17; Samtools v19.2) and mapping to *M. tuberculosis* reference genome H37Rv NC_000962.3 (BWA v0.7.17; Samtools v19.2).
    - Variants are called with the Genomic Analysis Toolkit (GATK v4.4.0.0) and annotated for genotypic antimicrobial resistance predictions with TBProfiler (v6.6.6).
    - Samples that pass quality control assessments are used as a baseline to identify closest historic samples based on Single Nucleotide Polymorphism (SNP) distances.
- Isolates are linked to cases in iPHIS using manual lookup of first name, last name, health card number, date of birth, and patient address.

Data Caveats

Whole Genome Sequencing

- As whole genome sequencing requires sufficient organism DNA to produce reliable results (concentration of at least ≥1 ng/µL), additional growth of the TB culture is required. There is approximately a 3–4-week time interval between the laboratory diagnosis of tuberculosis (positive culture identified) and the isolate being added to OUT-TB by WGS. Therefore, results from recent weeks may be incomplete.
- For isolates with sequencing results, genetic matches between cases have been classified according to the number of SNP differences as follows:
  - Matching: 0–3 SNPs
  - Closely related: 4–10 SNPs
  - Related: 11–15 SNPs
  - Unrelated: ≥16 SNPs
- A difference of ±2 SNPs is within acceptable variation for SNP distances determined by WGS.
- A closely related or related match means that the genome is highly similar but not an exact match. This may be due to variations in detection of SNPs which can be influenced by the quality and genome completeness of the sample. This may also be due to small changes in the isolate's DNA since transmission (minor mutation events), or it may be a strain that circulates commonly (endemic) that is not the result of direct transmission. While whole genome sequencing can help to inform case management and contact investigation, the genetic relatedness information should be reviewed within the context of the clinical and epidemiological case information for a complete "picture" of possible transmission.
- Matching and closely related are relative to the searched case and may not be related to other cases on the list.
- Only isolates that have a pairwise SNP distance of 15 or less are included in the SNP distance matrix figure.
- Low quality or ambiguous positions are removed from the analysis which can affect the single nucleotide polymorphism (SNP) distance for a specific sample and cluster assignment. Cluster nomenclature is dynamic. Cluster naming and assignment may change as more samples are sequenced and added or removed from analysis.
- Only SNPs at positions that have high quality data in a pair wise comparison are included in the analysis. Samples considered matching by pairwise distance analysis may not have identical genomes. Furthermore, although SNP distances may be the same between samples for clusters containing more than two cases, the specific SNPs present in each sample may not be shared across all samples.
- MIRU-VNTR and Spoligotyping are no longer available for prospective typing with implementation of WGS, however SM24 No. results for historical isolates are available on OUT-TB by WGS.

- For historical isolates that have not undergone whole genome sequencing, identical matches can be identified using the SM24_No, which is a PHO-assigned unique cluster number indicating identical genotype using spoligotyping pattern matching and exact 24-locus Mycobacterial Interspersed Repetitive Unit Variable Number of Tandem Repeats (MIRU-VNTR) methods. Only identical matches will receive an SM24_No, related matches cannot be identified for historical isolates.
- To allow for comparison between historical and prospective isolates, TB-Insight was used to translate between spoligotype data and lineage.
    - For historical isolates that were not sequenced, genotyping-derived spoligotyping data was translated to lineage and complex ID.
- MTBC WGS was developed with its performance characteristics determined by PHO for clinical testing and public health surveillance. It has not been cleared or approved by Health Canada.

Laboratory Information Management System

- TB drug resistance categories are based on the Canadian Tuberculosis Standards – 8th Edition[5]:
    - Mono-resistant TB is defined as resistance to only 1 of the 4 first-line drugs (isoniazid, rifampin, pyrazinamide, and ethambutol).
    - Polydrug-resistant TB is defined as resistance to at least two first-line drugs without resistance to rifampin.
    - Multi-drug-resistant TB (MDR-TB) is defined as resistance to isoniazid and rifampin with or without resistance to other first line anti-TB drugs.
    - Pre-XDR-TB is now defined as MDR-TB with additional resistance to any fluoroquinolone (moxifloxacin, levofloxacin, gatifloxacin, and ofloxacin).
    - Extensively-drug-resistant TB (XDR-TB) is defined as pre-XDR-TB with additional resistance to bedaquiline or linezolid.
    - Other drug resistance is defined as any atypical resistance profile that does not meet the definition of the standard TB drug resistance categories (e.g., susceptible to first-line anti-TB drugs but resistance to second-line TB drugs).
- PHO's testing algorithm does not test for bedaquiline susceptibility, instead isolates are referred to the National Microbiology Laboratory for testing. Therefore, the TB drug resistance category for a small number of isolates may be misclassified, although this is rare.
- As of April 2026, testing for the second-line drugs capreomycin, kanamycin, streptomycin, and para-aminosalicylic acid is no longer performed at PHO. Instead, isolates are referred to the National Microbiology Laboratory for testing on request only.
- Received date is based on the date the sample was received by the PHO laboratory.

Integrated Public Health Information System

- iPHIS is a dynamic disease reporting system that allows ongoing updates to previously entered data. As a result, data extracted from iPHIS represent a snapshot at the time of extraction and may differ from previous or subsequent reports.
- Data reported between 2020 and 2022 should be interpreted with caution. Both testing and iPHIS data entry practices were likely impacted by the COVID-19 pandemic response.

- Diagnosis refers to the type of TB disease. This is based on the iPHIS disease description and grouped into the following categories:
  - Non-respiratory
  - Respiratory
  - Respiratory - pulmonary
- Postal codes are first obtained from the iPHIS “time of illness address” where available. If this information is not available in iPHIS, it will then default to the most recent iPHIS address, then the patient postal code on the laboratory requisition, then the submitter postal code, and finally if none is available, PHOL– Toronto.
- Birth country combines iPHIS Client Immigration Birth Country and Born in Canada status. If birth country is missing, the information was not available/collected.

References

1. Ngabonziza J, Loiseau C, Marceau M, et al. A sister lineage of the *Mycobacterium tuberculosis* complex discovered in the African Great Lakes region. Nat Commun. 2020;11(1):2917. Available from: https://doi.org/10.1038/s41467-020-16626-6

2. Napier G, Campino S, Merid Y, et al. Robust barcoding and identification of mycobacterium tuberculosis lineages for epidemiological and clinical studies. Genome Med. 2020;12(1):114. Available from: https://doi.org/10.1186/s13073-020-00817-3

3. Coscolla M, Gagneux S, Menardo F, et al. Phylogenomics of *Mycobacterium africanum* reveals a new lineage and a complex evolutionary history. Microb Genom. 2021;7(2):000477. Available from: https://doi.org/10.1099/mgen.0.000477

4. Riojas MA, McGough KJ, Rider-Riojas CJ, Rastogi N, Hazbón MH. Phylogenomic analysis of the species of the Mycobacterium tuberculosis complex demonstrates that Mycobacterium africanum, Mycobacterium bovis, Mycobacterium caprae, Mycobacterium microti and Mycobacterium pinnipedii are later heterotypic synonyms of Mycobacterium tuberculosis. Int J Syst Evol Microbiol. 2018;68(1):324-32. Available from: https://doi.org/10.1099/ijsem.0.002507

5. Brode SK, Dwilow R, Kunimoto D, Menzies D, Khan FA. Canadian tuberculosis standards. 8th ed. Can J Respir Crit Care Sleep Med. 2022;6(sup1):109-28. Chapter 8, Drug-resistant tuberculosis. Available from: https://doi.org/10.1080/24745332.2022.2039499

Appendix: Data Dictionary

Table A1: Laboratory-Confirmed Culture-Positive Tuberculosis Cases Field Names and Descriptions

Field Name	Description
Client_ID	iPHIS identifier denoting a unique person
Episode_ID	iPHIS identifier denoting a unique case/episode of illness (a person may have >1 episode)
WGS_ID	PHO laboratory identifier for sequenced specimen
Rec_Date	Date specimen received at PHO laboratory
YOB	Client year of birth as provided on PHO test requisition
Sex	Sex of patient as provided on PHO test requisition
Birth_Country	Combined iPHIS Client Immigration Birth Country or Born in Canada status
City	Client city at time of illness
DHU	Diagnosing Public Health Unit identified in iPHIS
RHU	Responsible Public Health Unit identified in iPHIS
Onset	iPHIS Onset Date of symptoms
Diagnosis	Based on the iPHIS field Disease Description this field will display “Resp” for respiratory, "Resp-Pulm" for pulmonary, "Non-Resp" for non-respiratory, or any combination of the three for patients with multiple disease sites indicated
Source	Anatomical site and/or material from which specimen was collected
Complex_ID	Mycobacterium tuberculosis Complex variant. For historical isolates that were not sequenced, Complex_ID was derived from TB-Insight using spoligotype data.
Lineage	Lineage designation of the sequenced specimen. For historical isolates that were not sequenced, lineage designation was derived from TB-Insight using spoligotype data.

Field Name	Description
AFB_Smear	AFB smear results are displayed as Neg (No acid-fast bacilli seen), few (3-9 per smear), 1+ (1-9 per 10 fields), 2+ (1-9 per field), 3+ (10-90 per field), or 4+ (>90 per field)
SM24_No	PHOL-assigned cluster number indicating identical spoligotype (S) and 24-locus MIRU-VNTR (M24) genotypes. SM24 Number is only available for historical isolates. If an isolate underwent genotyping but had no identical matches, the value is displayed as "-". If an isolate did not undergo genotyping or had missing genotyping data, the value is displayed as "NA".
Drug_Res	Drug resistance results for the isolate will be displayed as S (susceptible), Mono-R (Mono-resistant), Poly-R (Poly-resistant), MDR (Multi-drug resistant), Pre-XDR (Pre-extensively-drug-resistant), or XDR (Extensively-drug-resistant). For more details on the drug resistance category definitions, please see technical notes

Table A2: Genetic Relatedness by Whole Genome Sequencing to Selected Case Field Names and Descriptions

Field Name	Description
WGS_ID	PHO laboratory identifier for the selected case
Match_Type	Categorizes genetic relatedness based on the number of single nucleotide polymorphism (SNP) differences between two specimens (Match - 0 to 3 SNPs, Close - 4 to 10 SNPs, and Related - 11 to 15 SNPs). A difference of ±2 SNPs is within acceptable variation for SNP distances determined by WGS.
SNP	The number of SNP differences between the selected case and matched case
Match_Case	WGS ID of the genetically related case
Client_ID	iPHIS identifier denoting a unique person
Episode_ID	iPHIS identifier denoting a unique case/episode of illness (a person may have >1 episode)
Rec_Date	Date specimen received at PHOL
YOB	Client year of birth as provided on PHO test requisition
Sex	Sex of patient as provided on PHO test requisition
Birth_Country	Combined iPHIS Client Immigration Birth Country or Born in Canada status

Field Name	Description
Arrival	Date that person arrived in Canada (if applicable)
Onset	iPHIS Onset Date of symptoms
Postal_Code	Client postal code at time of illness
DHU	Diagnosing Public Health Unit identified in iPHIS
RHU	Responsible Public Health Unit identified in iPHIS
SM24_No	PHOL-assigned cluster number indicating identical spoligotype (S) and 24-locus MIRU-VNTR (M24) genotypes. SM24 Number is only available for historical isolates. If an isolate underwent genotyping but had no identical matches, the value is displayed as “-”. If an isolate did not undergo genotyping or had missing genotyping data, the value is displayed as “NA”.
Drug_Res	Drug resistance results for the isolate will be displayed as S (susceptible), Mono-R (Mono-resistant), Poly-R (Poly-resistant), MDR (Multi-drug resistant), Pre-XDR (Pre-extensively-drug-resistant), or XDR (Extensively-drug-resistant). For more details on the drug resistance category definitions, please see technical notes

Citation

Ontario Agency for Health Protection and Promotion (Public Health Ontario). Technical notes: Ontario universal typing of tuberculosis program by whole genome sequencing. Toronto, ON: King's Printer for Ontario; 2026.

How to Cite this Tool

Generic Citation

Ontario Agency for Health Protection and Promotion (Public Health Ontario). The ontario universal typing of tuberculosis program by whole genome sequencing >> [indicator title in sentence case] [Internet]. Toronto, ON: King's Printer for Ontario; cYYYY [modified YYYY Mon DD; cited YYYY Mon DD]. Available from: URL

Example Citation

Ontario Agency for Health Protection and Promotion (Public Health Ontario). The ontario universal typing of tuberculosis program by whole genome sequencing >> exact title of chart in sentence case [Internet]. Toronto, ON: King's Printer for Ontario; c2024 [modified 2024 Oct 25; cited 2024 Oct 31]. Available from: URL

Disclaimer

Public Health Ontario

Public Health Ontario is an agency of the Government of Ontario dedicated to protecting and promoting the health of all Ontarians and reducing inequities in health. Public Health Ontario links public health practitioners, front-line health workers and researchers to the best scientific intelligence and knowledge from around the world.

For more information about PHO, visit publichealthontario.ca.

Ontario