



# TB Genotyping and Outbreak Investigations

Lisa Paulos, Epidemiologist

September 17, 2020



# Definitions

---

- **TB genotyping:** a laboratory-based approach used to analyze the genetic material (e.g., DNA) of *Mycobacterium tuberculosis*.
- **GENType:** Each unique combination of the spoligotype and 24-locus MIRU-VNTR genotyping method. Current method of genotyping used for surveillance.
- **Whole Genome Sequencing (WGS):** wgSNP (single nucleotide polymorphism) analysis uses whole-genome sequencing data to identify SNPs that are useful for examining the genetic relationship among isolates. A SNP is a change at a single position (A,T,C, or G) in the DNA sequence.
  - All isolates are tested using this method since 2018.
  - Will be the method used for surveillance beginning 2021.

# TB Outbreak Definitions

---

- **CDC definition:** More TB cases than expected within a geographic area or population during a particular time period AND evidence of recent transmission.
- **Maryland definition:** 3 cases that match by GENType within a geographic area over 3 years.
- **Other considerations:**
  - Epidemiologic links among cases
    - Common location
    - Contacts to infectious case
  - Multi-drug resistant TB
  - Cases among children

# Purposes of TB Genotyping

---

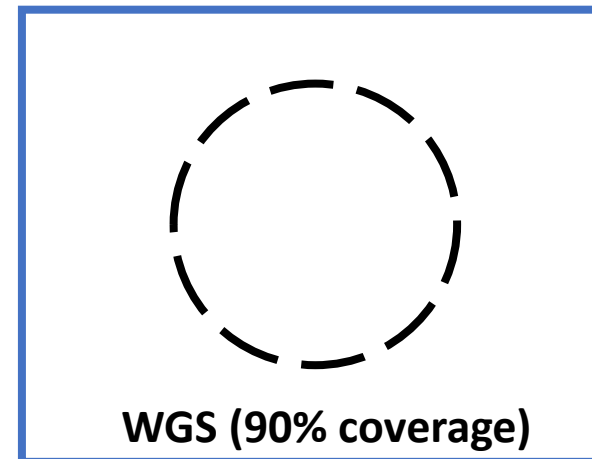
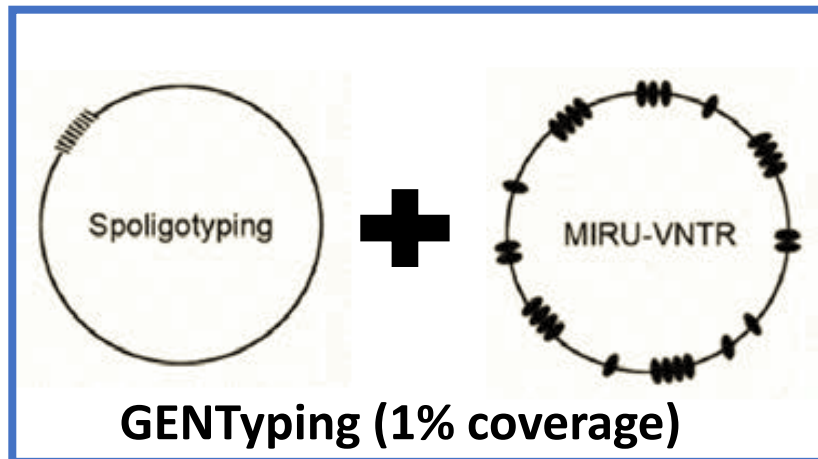
- Help identify persons with TB disease involved in the same chain of recent transmission.
- Distinguish between persons whose TB disease is the result of TB infection that was acquired in the past, as compared to recently or newly acquired infection.
- Tool that can add value to contact investigation activities.
- Identify or confirm laboratory cross contamination (false positive) episodes.
- Identify drug resistance at the molecular level.



# Genotyping Methods

---

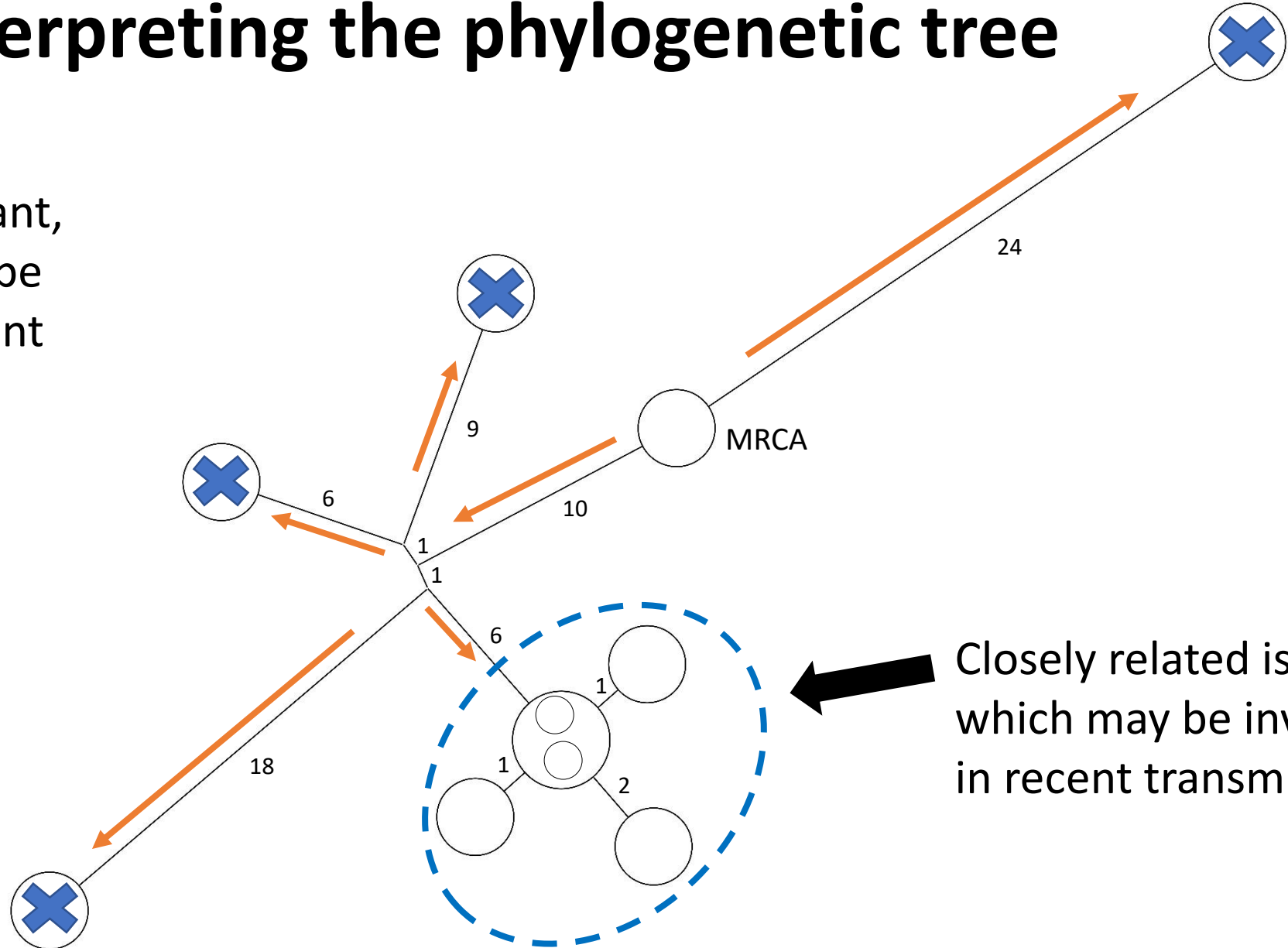
- GENTyping covers about 1% of the TB genome.
- WGS expands coverage of the genome to ~90%—Captures much more of the genetic changes that occur



# Guide for interpreting the phylogenetic tree

✘ = genetically distant, and unlikely to be involved in recent transmission

→ = direction of genetic change



Closely related isolates, which may be involved in recent transmission

# Recent TB Transmission in Maryland

---

Recent Transmission Clusters*, Maryland Data, 2017-2020					
Year	2 cases	3 cases	4-9 cases	10+ cases	Multistate
2017	3	2	1	1	0
2018	2	0	0	1	2
2019	4	0	1	1	3
2020	1	2	0	1	1

\*does not include genotype clusters evaluated and refuted by WGS

# Case Scenarios

---

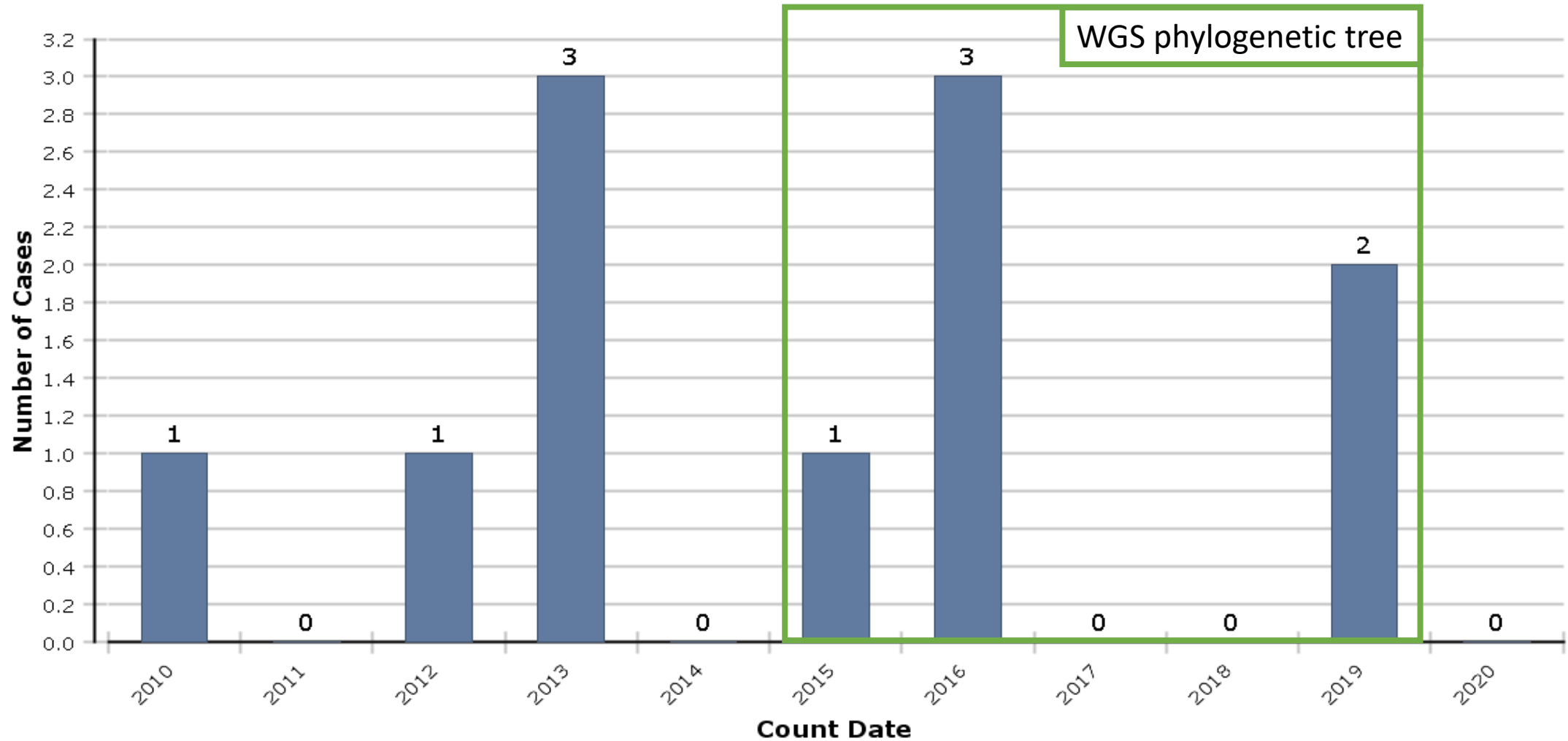
G08737

G10249

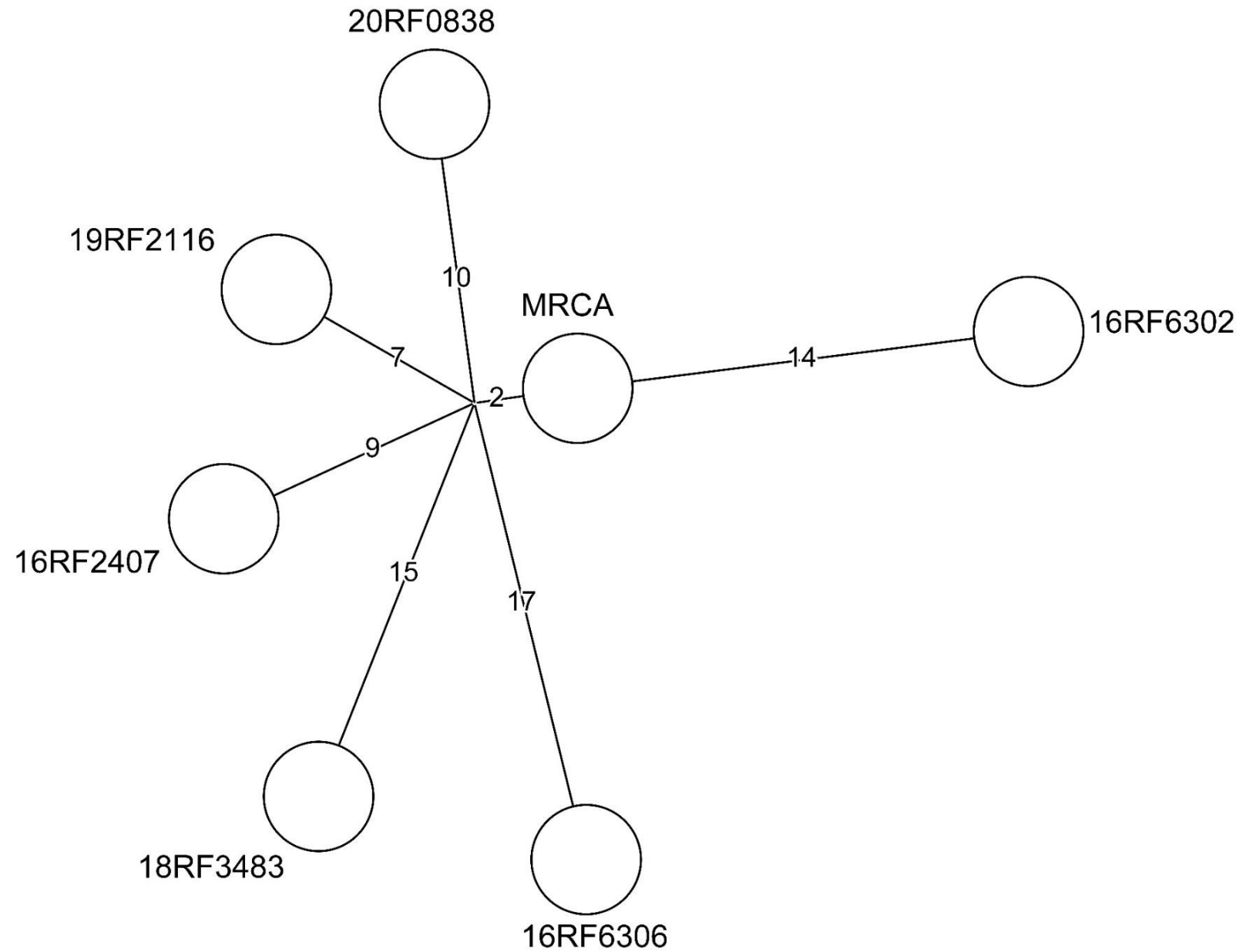
G17444-Multistate



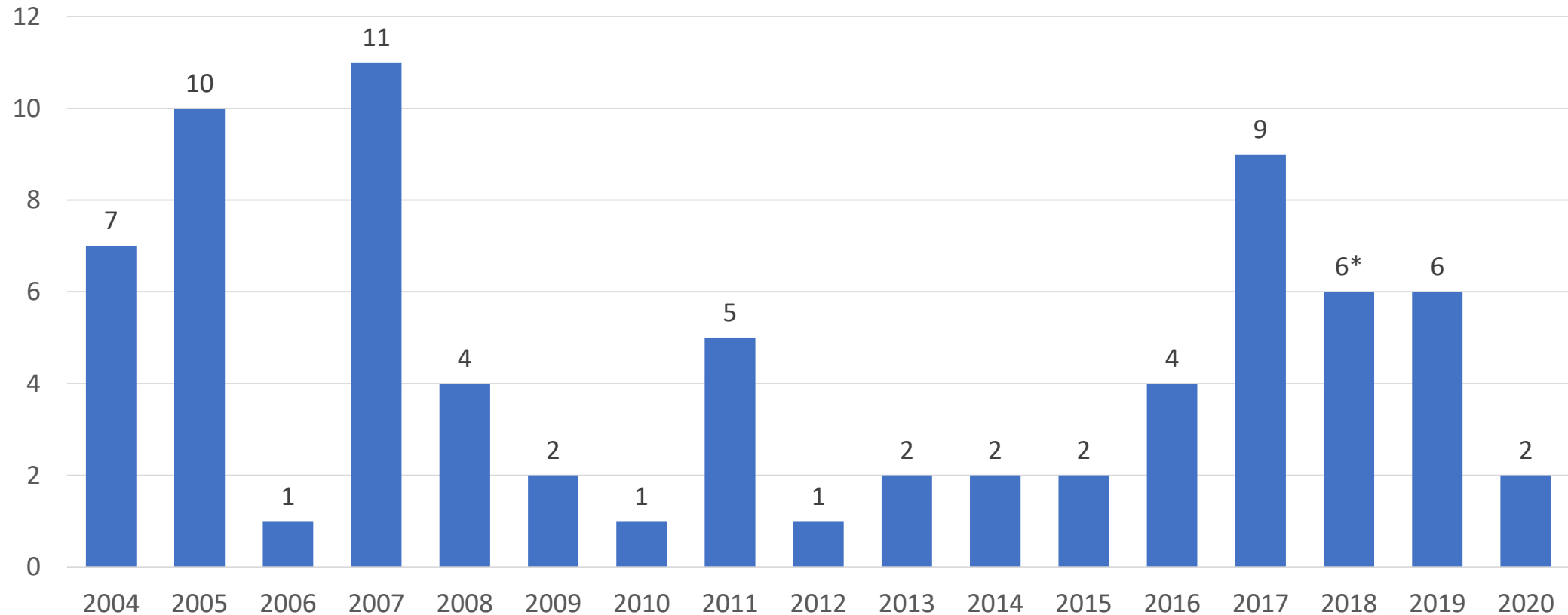
# G08737 WGS refutes recent transmission



# G08737 Phylogenetic Tree



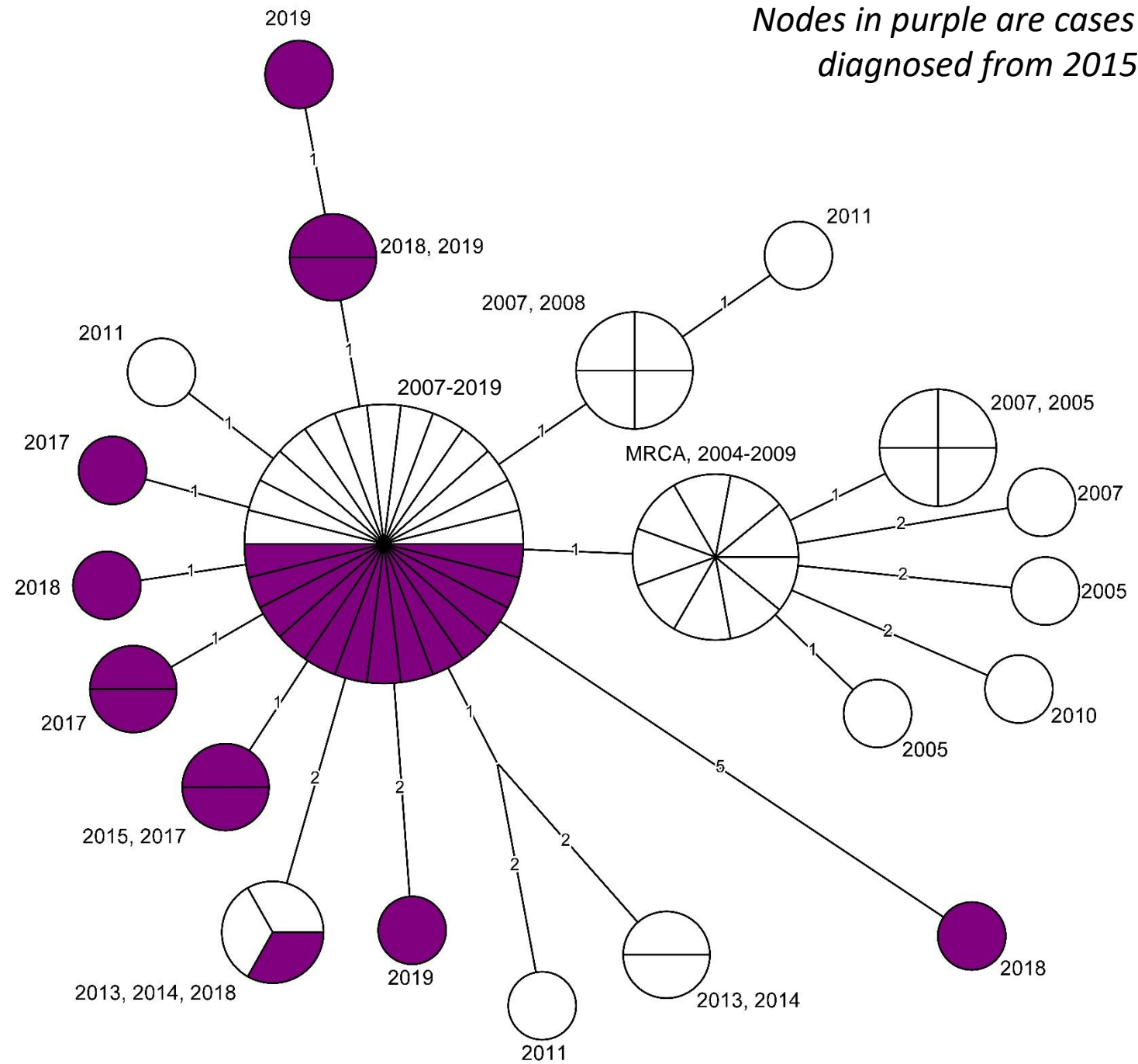
# G10249 in Maryland: Epidemiologic Curve



\*1 2018 case diagnosed in California

# G01249 Phylogenetic Tree as of 1/23/2020

*Nodes in purple are cases that were  
diagnosed from 2015-2019*



# G10249 Demographic Information Over Time

---

## 2005-2008\*, 26 cases

- 85% Male
- 73% Homeless
- 27% Excess alcohol use
- 50% Substance use
- 42% HIV positive
- 15% Died

## 2009-2014, 13 cases

- 69% Male
- 23% Homeless
- 31% Excess alcohol use
- 15% Substance use
- 15% HIV positive
- 7% Died

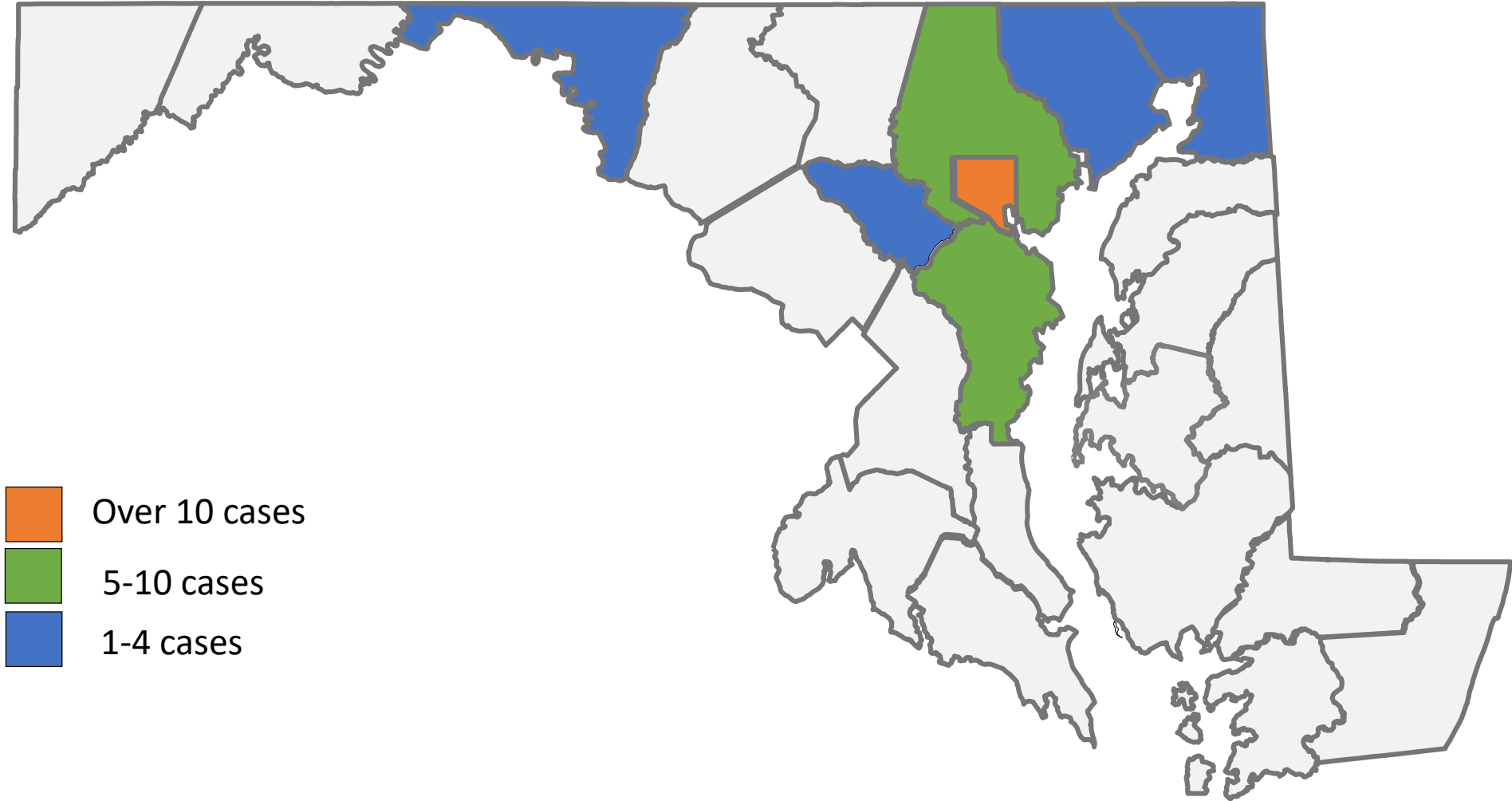
## 2015-2019, 27 cases

- 52% Male
- 15% Homeless
- 19% Excess alcohol use
- 33% Substance use
- 15% HIV positive
- 11% Died

\*Data for 2004 cases not available

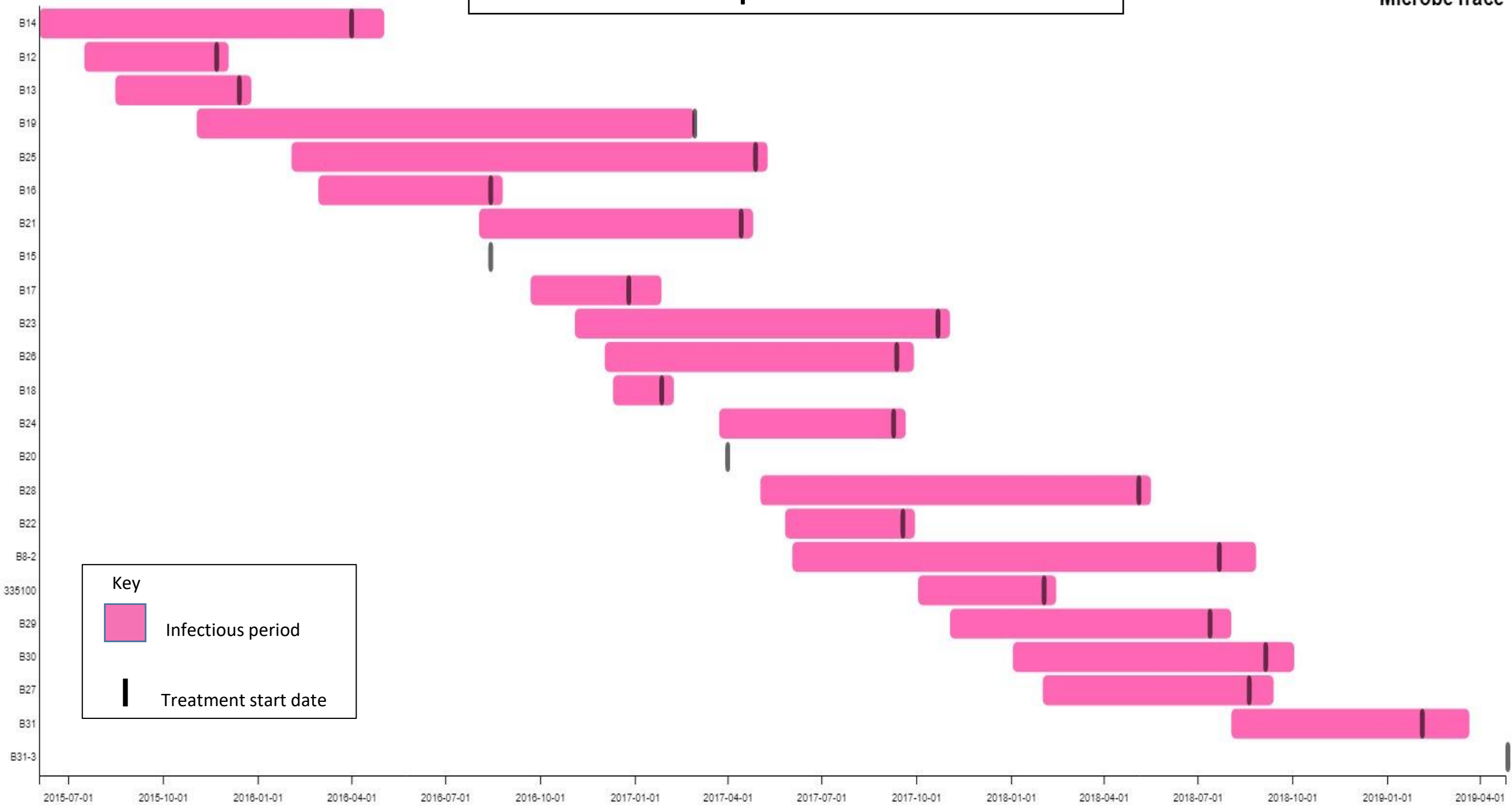
# G10249 Cases by County

---



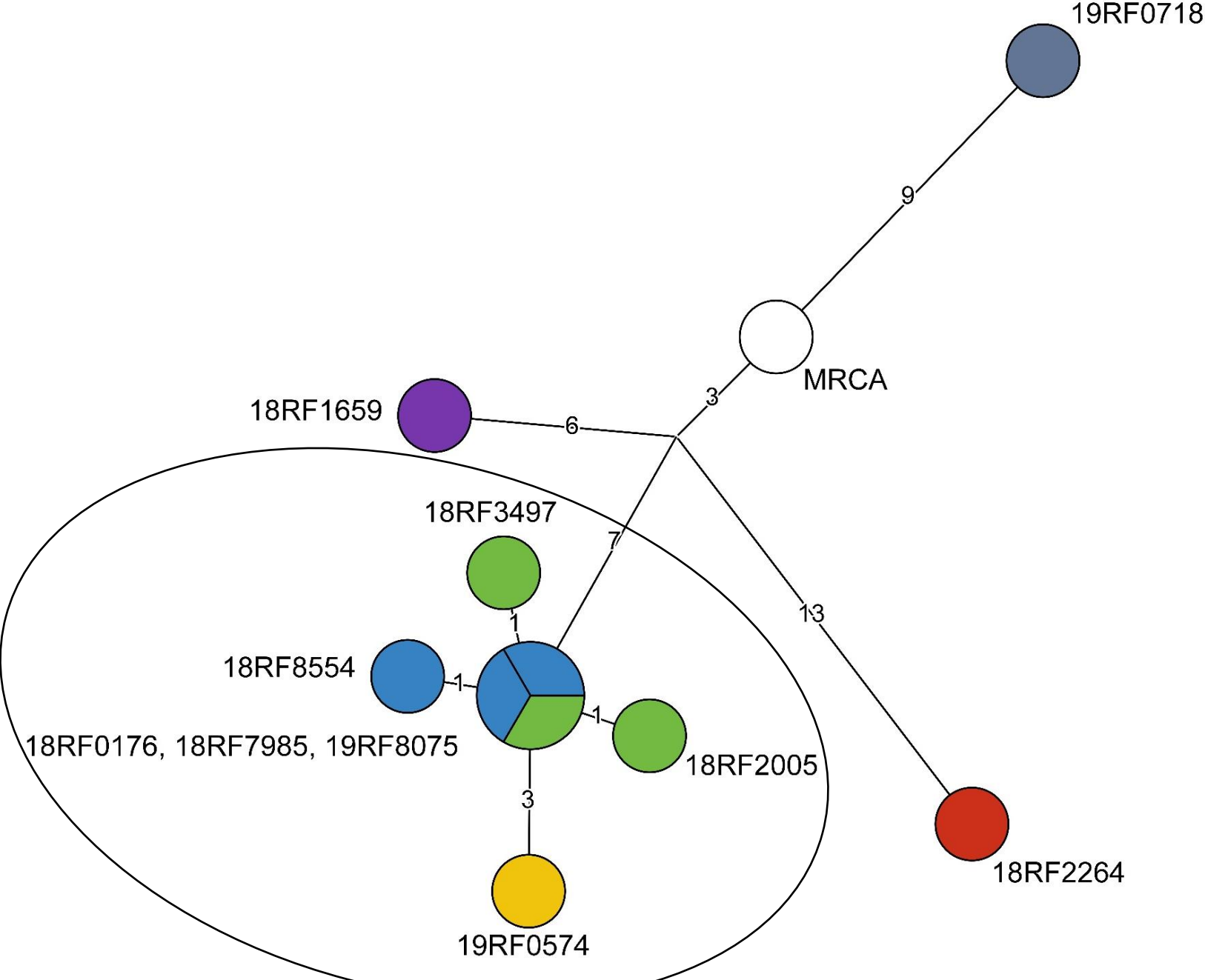
# G10249 Infectious period timeline 2015-2019

MicrobeTrace



# G17444 in the United States

- VA
- MD
- MA
- MT
- AZ
- TX





# G17444 Investigation Activities

---

- Interjurisdictional cluster of 6 cases with same Genotype
  - Maryland
    - 2 counties
  - Virginia
    - 2 counties
- Requested WGS analysis
- Shared interview tool
- Shared cluster investigation tool
- Call with CDC Partners to review cluster
- Reviewed with Massachusetts for possible epi-links
  - None identified

# Summary

---

- Whole genome sequencing provides additional information about GENType clusters in Maryland.
- Clusters across county and state lines can be identified.
- Transmission of TB in Maryland is occurring.
- Cluster and outbreak investigations are resource and time intensive.
- Robust contact investigations can help prevent TB outbreaks.

# Resources and Contact Information

---

Guide to the Application of Genotyping to Tuberculosis Prevention and Control

<https://www.cdc.gov/tb/programs/genotyping/default.htm>

TB Self study module 9: Tuberculosis Outbreak Detection and Response

Lisa Paulos, Maryland Center for TB Control and Prevention

410-767-6692, [lisa.paulos@maryland.gov](mailto:lisa.paulos@maryland.gov)

**Thank You!**



# Informational Slides

---

- The following slides are sent by CDC with the phylogenetic tree (WGS analysis) results.
- Please review for additional information about TB genotyping I was not able to cover in the presentation.
- Please contact me if you have additional questions—I love to talk about genotyping!

# Whole-genome single nucleotide polymorphism (wgSNP) analysis

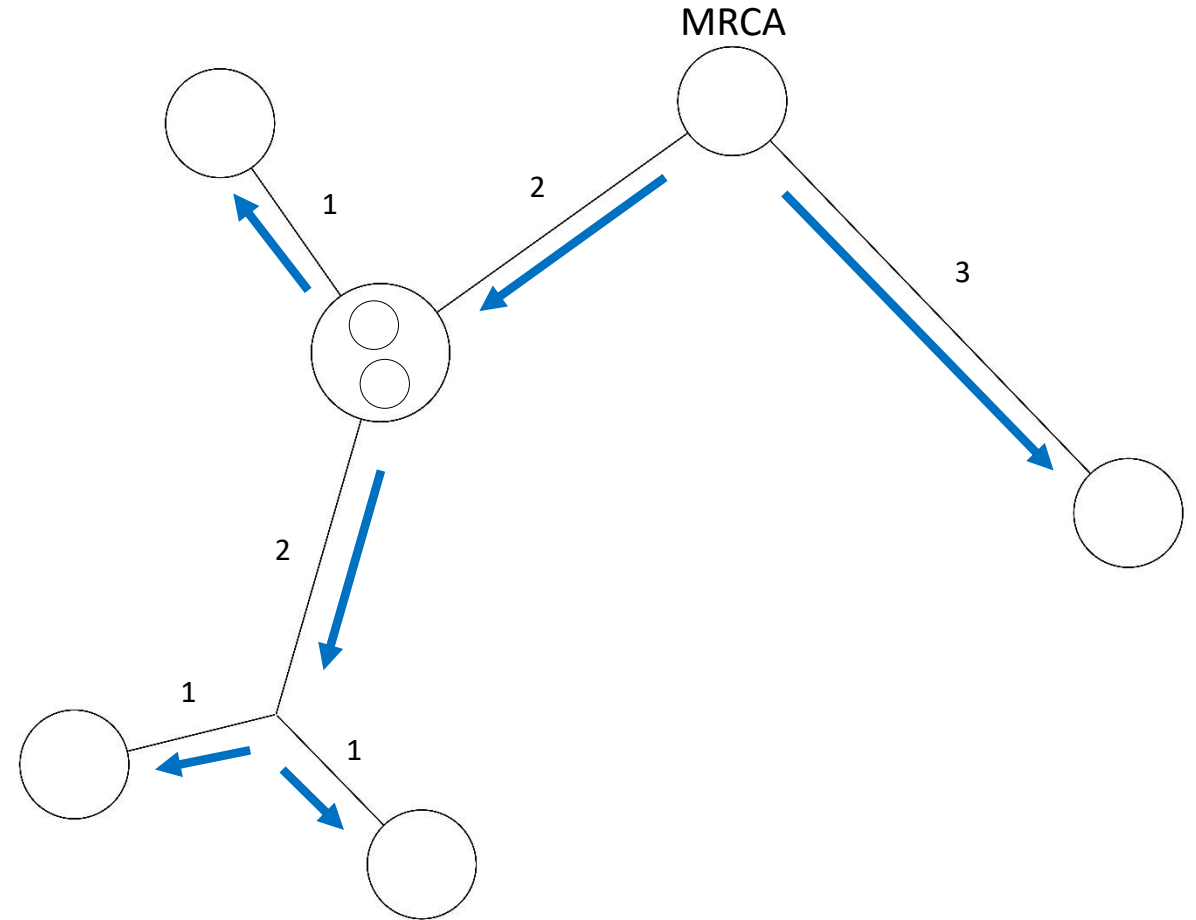
- wgSNP analysis was performed for GENType clustered isolates to further assess the genetic relatedness of the isolates. Isolates that are closely related indicate possible recent transmission. wgSNP analysis expands coverage of the genome to ~90% (i.e., compared to ~1% coverage with conventional genotyping).
- wgSNP analysis uses WGS (whole-genome sequencing) data to identify SNPs that are useful for examining the genetic relationship among isolates. A SNP is a change at a single position (A,T,C, or G) in the DNA sequence.
- SNPs that are identified in the wgSNP analysis are mapped on to a phylogenetic tree to diagram the genetic relationship among isolates.
- The phylogenetic tree can be used to target and inform epidemiologic investigation of these cases.

# Guide for interpreting the phylogenetic tree

- Isolates are shown as circles (called nodes) and are labeled with the isolate accession number
- Isolates with the same genome type (i.e., same sequence) are displayed together in one node
- Nodes are connected by lines proportional in length to the number of SNPs that differ between the isolates
- The lines are labeled with the number of SNPs

## MRCA = Most Recent Common Ancestor

- Hypothetical genome type (not an actual isolate) from which all isolates on the tree are descended
- Serves as a reference point for examining the direction of genetic change ( → )







# Guide for interpreting the phylogenetic tree

- SNP thresholds for categorizing *M. tuberculosis* isolates as genetically distant or closely related have not been formally established for CDC's wgSNP analysis yet
- Based on CDC's general experiences using wgSNP analysis for investigating recent transmission:
  - Isolates with 0 – 5 SNP differences are considered closely related
  - Isolates with 6 or more SNP differences are considered genetically distant
- SNP thresholds will vary depending on the methods used for the wgSNP analysis, and cannot be compared to thresholds used by other groups with different analysis methods
- These recommended SNP thresholds may change as CDC's wgSNP analysis methods are further developed or based on results of a formal validation analysis of SNP thresholds

# Important Considerations (1)

- The phylogenetic tree should be used to identify clusters of closely related isolates that may be the result of recent transmission. It should not be used to draw conclusions about direction of transmission among individual patients.
- In reviewing the phylogenetic tree, it is important to remember that not all closely related isolates will necessarily be depicted on the tree.
  - The inclusion criteria for selected isolates (e.g. time period, geographic boundaries, and GENTypes) as well as the exclusion of clinically diagnosed or undetected cases may influence interpretation of the analysis.

# Important Considerations (2)

- Recent transmission is easier to rule out than to confirm with WGS
- Even isolates that are closely related or identical by WGS can be due to reactivation
  - This is because mutations may not occur as frequently during latent infection and therefore SNPs may not accumulate
- The phylogenetic tree should be used in conjunction with clinical and epidemiologic information to assess recent transmission
- For additional information, an extended training titled “Tuberculosis WGS Training Module” is available under “New Resources” at the following location :

<https://www.cdc.gov/tb/programs/genotyping/default.htm>