Integration of Whole Genome Sequencing into the Molecular Surveillance of *Mycobacterium tuberculosis*

Angela M. Starks, PhD
Chief, Laboratory Branch
Division of TB Elimination
May 21, 2015

Molecular Surveillance of *M. tuberculosis*

- **National TB Genotyping Service (2004-present)**
  - Spoligotyping and 24 locus MIRU-VNTR typing
  - ~10,000 isolates/year
  - Provided through CDC contract awarded to Michigan Dept. of Community Health
- **TB Genotyping Information Management System (TB GIMS)**
  - Web based database to manage genotyping results
  - Integrates surveillance data with genotyping results
  - Provides tools for interpretation of genotyping results (maps, epi curves, profiles)

Primary Goal of Molecular Surveillance is to Detect Recent Transmission of *M. tuberculosis*

If…

genotyping methods provide sufficient discrimination,

genotyping markers have a molecular clock similar/faster than the rate of transmission,

and the population is mobile

then…

the only reasonable explanation for two cases being infected with genotype clustered isolates is recent transmission

The Challenges of Molecular Surveillance

- **Excellent discriminatory power (28,000 distinct genotypes) but sometimes insufficient**
- **Molecular clock of markers not at same pace as transmission**
  - Clustered cases where transmission was remote in time or geography
- **Stepwise evolution of VNTR patterns may lead to repeated convergence/divergence of strains with similar genotypes**
  - Cluster isolates of divergent strains over long periods of time
- **Can be difficult to interpret genotype results**
  - One large group of related cases?
  - Small group(s) of related cases?
  - Clustering of entirely unrelated cases?

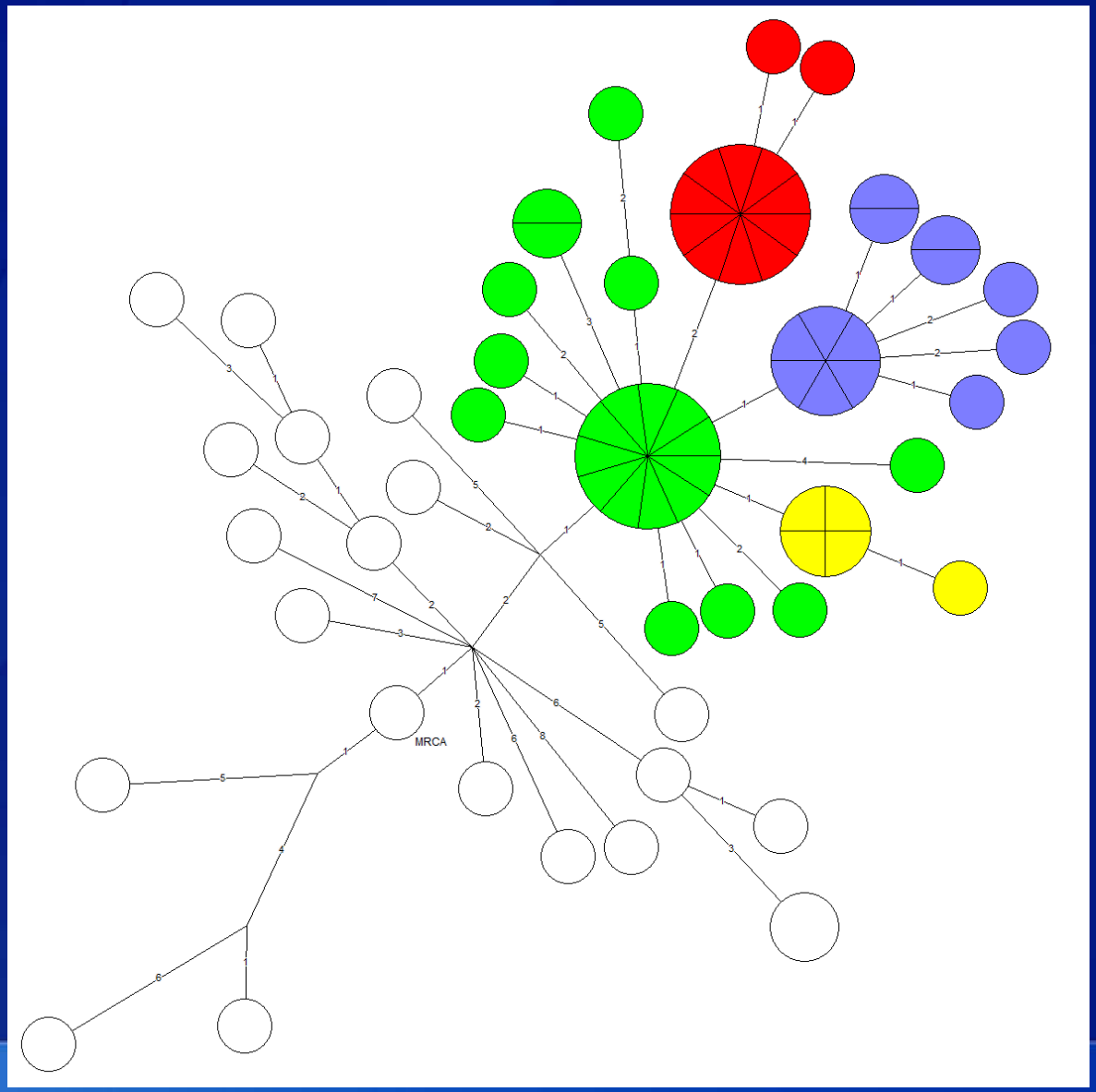
Whole Genome Sequencing (WGS)

- **High resolution comparison of genomes**
  - Conventional genotyping examines <1% of the genome
  - WGS examines >90% of the genome

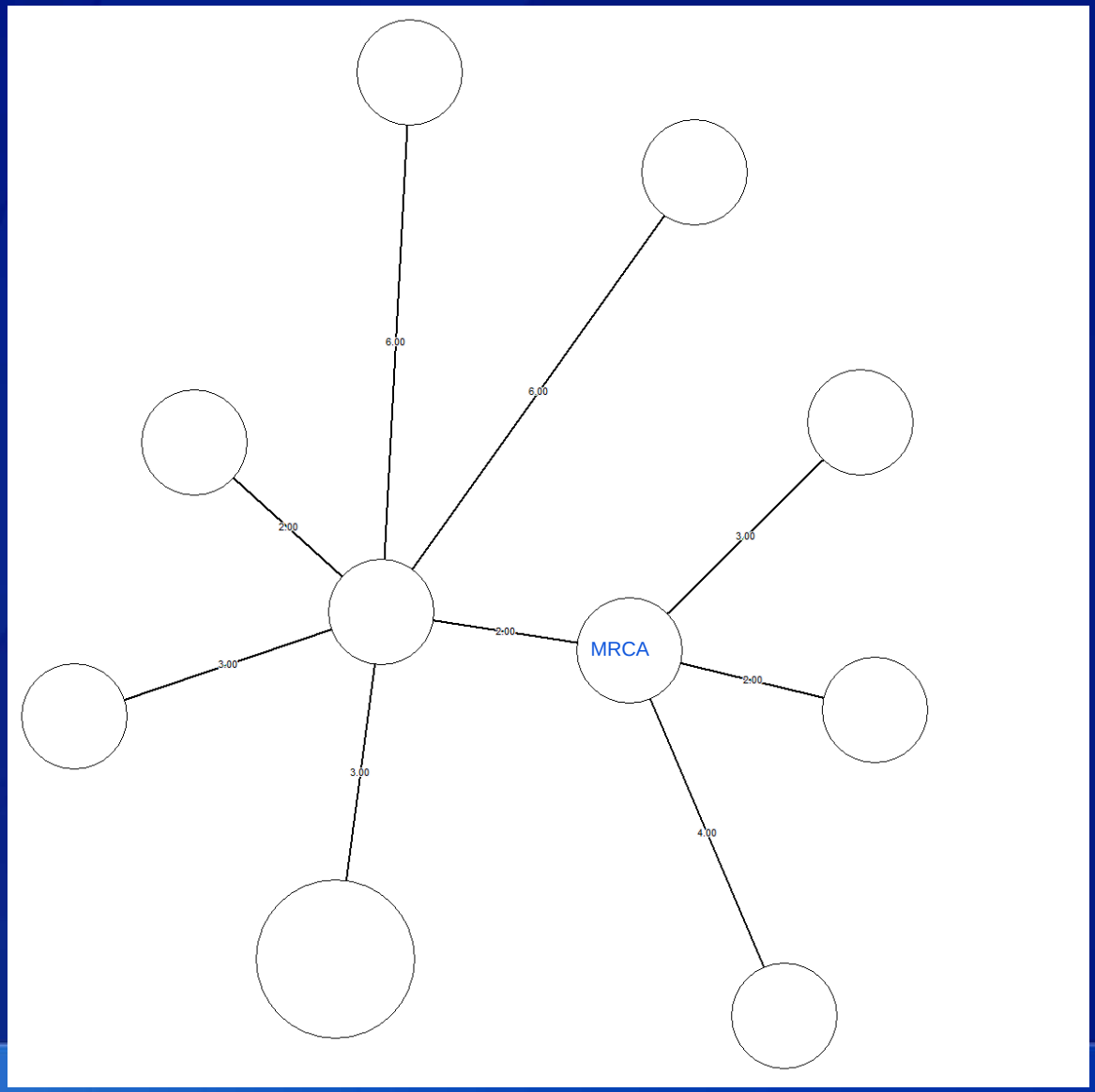
Comparison of Genomes

- **“High quality differences between genomes”**
  - Compare the reference genome with all of the assembled reads from the sample and identify all single nucleotide polymorphisms (SNPs) that are present in at least 2 mapped reads and that are present in at least 30% of the reads mapped to that base (SNP report)
  - Compare the SNP reports from all of the samples of interest and identify those SNP positions that are unique to a sample or subset of samples (informative SNP panel)
    - Assembly at repetitive genes (PE/PPE gene family), homologus/pseudogenes, and near insertion/deletions is error prone
  - Build phylogenetic maps based on the informative SNP panel

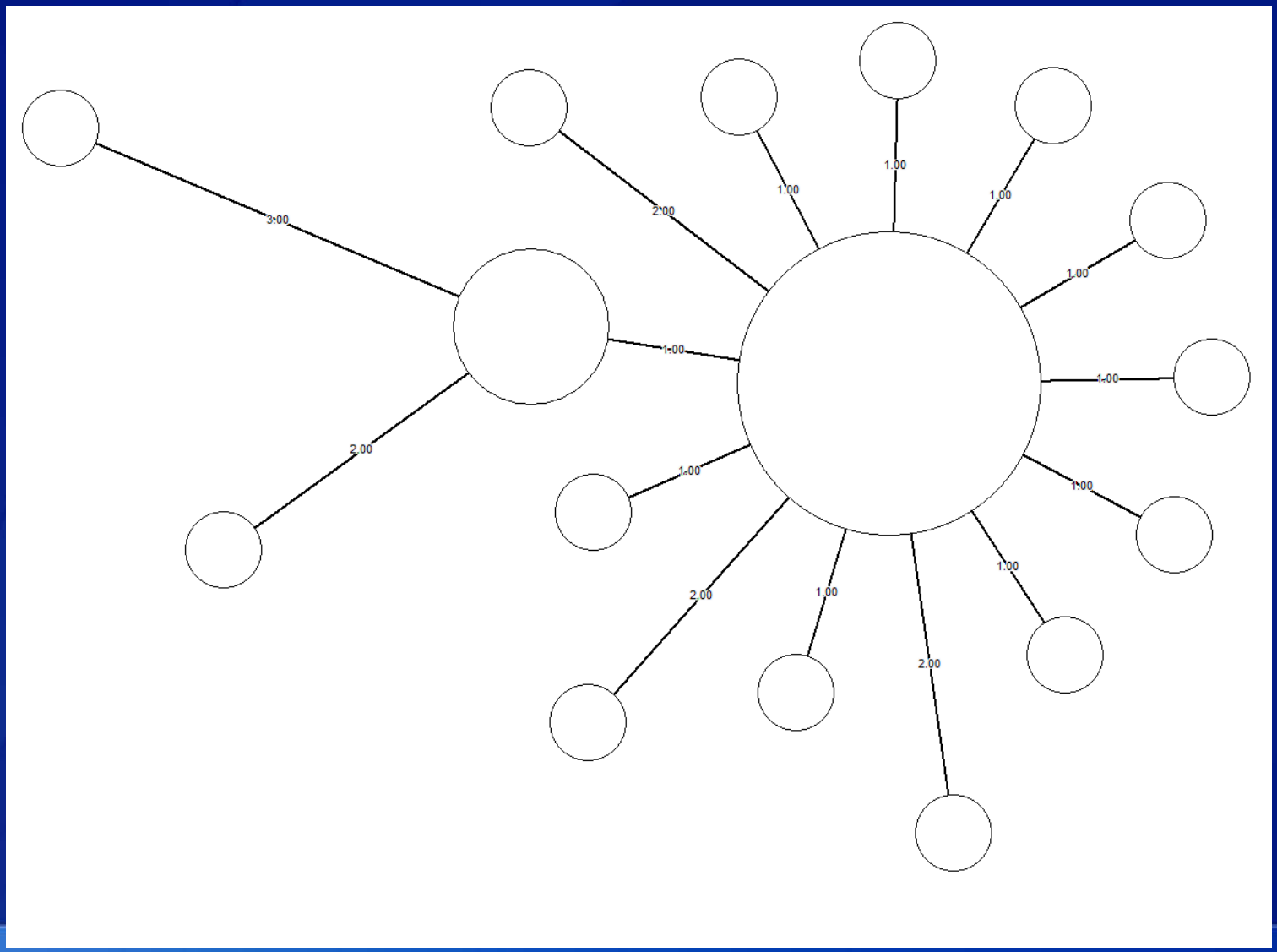
WGS may reveal diversity between isolates in a genotype cluster



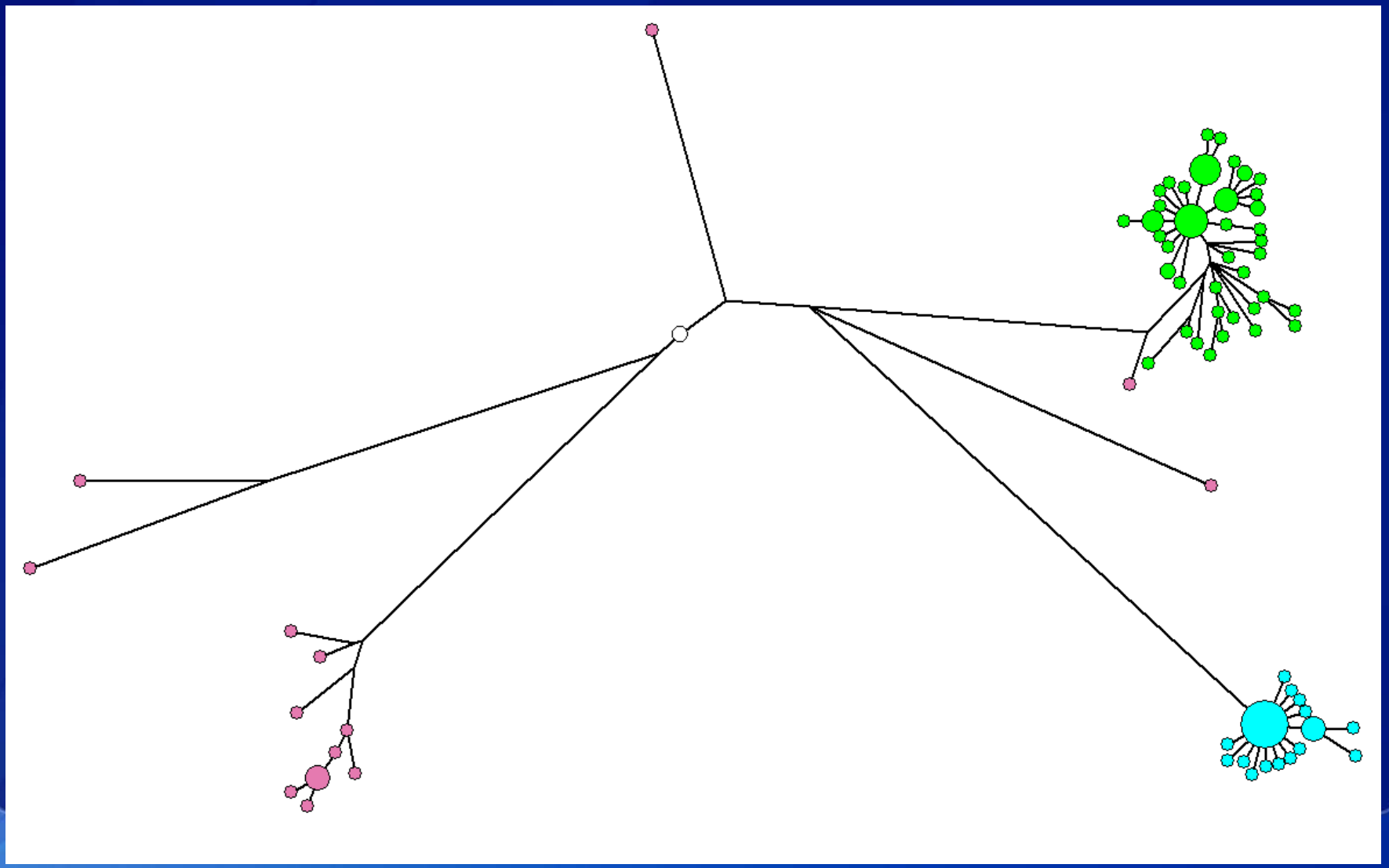
WGS may reveal significant diversity between isolates in a genotype cluster



WGS may reveal limited diversity between isolates in a genotype cluster



Genome from one cluster may be more similar to member of a different genotype cluster than to members of its own genotype cluster



Important Points

- ***M. tuberculosis* genome is clonal**
  - Current estimates of molecular clock is 0.5 nucleotides/year
  - Molecular clock might start and stop
- ***M. tuberculosis* population at transmission might not be the same at culture**
- **WGS provides high resolution data to identify cases that may be due to recent transmission but does not provide proof of transmission**
- **Current short read WGS platform does not allow extraction of MIRU-VNTR patterns**

Future Directions

- **Funding regional labs to increase sequencing capacity to 750–1,000 isolates/year**
- **Research to optimize DNA extraction**
- **Research to optimize assembly and comparison algorithm**
  - Whole genome Multi Locus Sequencing Typing (wgMLST)
  - Universal WGS in NYC (2012-2014) dataset (1,600 samples)

Acknowledgements

- **DTBE Laboratory Branch/Applied Research Team**
  - Jamie Posey
  - Paige Chopra
  - Lauren Cowan
  - Melisa Willby
- **Biotechnology Core Facility Branch/Genome Sequencing Laboratory**
  - Mike Frace
  - Mark Burroughs
  - Kristen Knipe
  - Mili Sheth
- **DTBE Surveillance, Epidemiology, and Outbreak Investigations Branch**
- **CDC OAMD**