

# INTEGRATING MOLECULAR, EPIDEMIOLOGIC, AND CLINICAL DATA TO INVESTIGATE A LARGE TUBERCULOSIS OUTBREAK IN CALIFORNIA, 2012–2016

Clint McDaniel<sup>1</sup>, Kathryn Winglee<sup>1</sup>, Martin Cilnis<sup>2</sup>, Sarah Talarico<sup>1</sup>, Yohani Ramos<sup>3</sup>, Veronica Pehl<sup>3</sup>, Alvaro Garza<sup>4</sup>, Lauren Cowan<sup>1</sup>, Julie Vaishampayan<sup>5</sup>, James Posey<sup>1</sup>, Tambi Shaw<sup>2</sup>, Benjamin Silk<sup>1</sup>

<sup>1</sup>Centers for Disease Control and Prevention, Atlanta, GA, USA  
<sup>2</sup>California Department of Public Health, Richmond, CA, USA

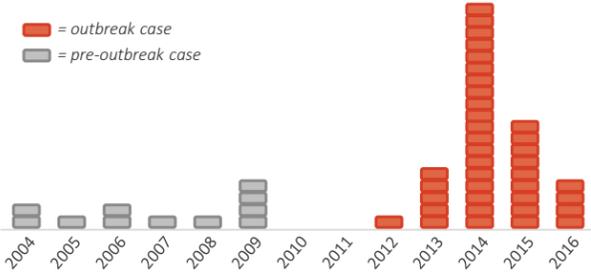
<sup>3</sup>San Joaquin County Public Health Services, Stockton, CA, USA  
<sup>4</sup>San Joaquin County Public Health Services, Stockton, CA, USA (retired)

<sup>5</sup>Stanislaus County Health Services Agency (formerly San Joaquin County Public Health Services), Modesto, CA, USA

## BACKGROUND

- Tuberculosis (TB) outbreak investigations can be complex and resource-intensive. Molecular, epidemiologic, and clinical data can help identify strategies to interrupt transmission, but best approaches for data integration and visualization are not widely established.
- In early 2014, local staff identified three TB cases with a rare genotype. Further review identified additional cases with the genotype; an outbreak investigation was initiated.
- As a case study, we present an integrated analysis from this large TB outbreak investigation in California during 2012–2016 (Figure 1).

Figure 1: Epidemiologic Curve of a Large Outbreak in California, 2004–2016



## METHODS

- We analyzed epidemiologic, whole-genome sequencing (WGS), and clinical data for genotype-matched or epidemiologically linked cases diagnosed in California during the 2012–2016 outbreak (n=38 cases) and the 2004–2011 pre-outbreak (n=11 cases) periods.
- Based on social connections and shared locations, we characterized epidemiologic linkages as definite, probable, or possible (Figure 2).
- We analyzed available WGS data to identify informative single nucleotide polymorphisms (SNPs):
  - SNP differences between isolates are shown as lines between isolates on a neighbor-joining tree phylogenetic tree (Figure 3).
  - Isolates within five SNPs of another isolate were considered genetically closely related.
- We developed a plausible transmission diagram (Figure 4) through an integrated analysis using the following data:
  - SNP difference between isolates
  - Indicators of infectiousness and infectious periods
  - Case relationships and potential exposure sites
  - Strength of epidemiologic links

## RESULTS

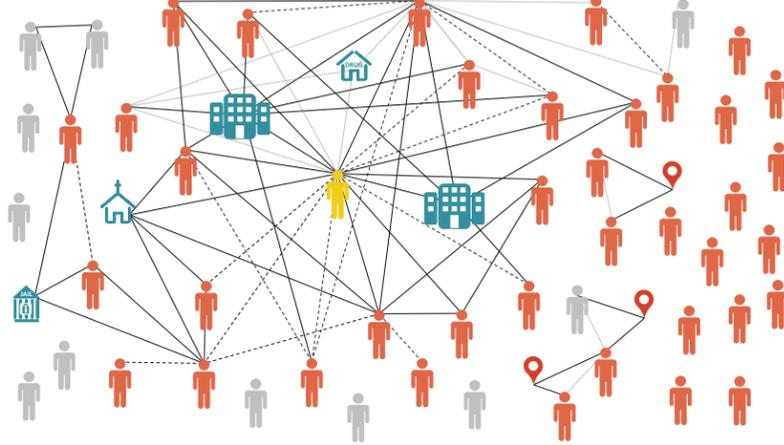


Figure 2: Epidemiologic linkages

Key findings:

- 25 of 38 (66%) cases had an epidemiologic link to ≥1 cases (range: 1–14 epidemiologic links per case; median=2).
- The patient shown in yellow with many social connections had a highly infectious TB and a long infectious period, which was key to understanding transmission in this outbreak.

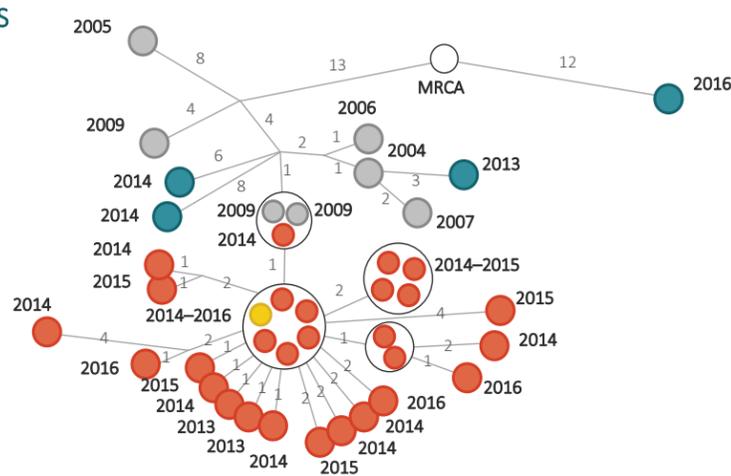


Figure 3: Whole-genome sequencing results

Key findings:

- 29 of 33 (88%) outbreak cases with WGS data had an isolate that was closely related to ≥1 other isolates (range: 1–27 isolates closely related to the given isolate; median=21).
- 21 of 22 (95%) epidemiologically linked cases, and 8 of 11 (73%) cases without an identified epidemiologic linkage, were closely related to at least one other isolate.
- Recent transmission was ruled out for four cases during the outbreak period (shown in teal).

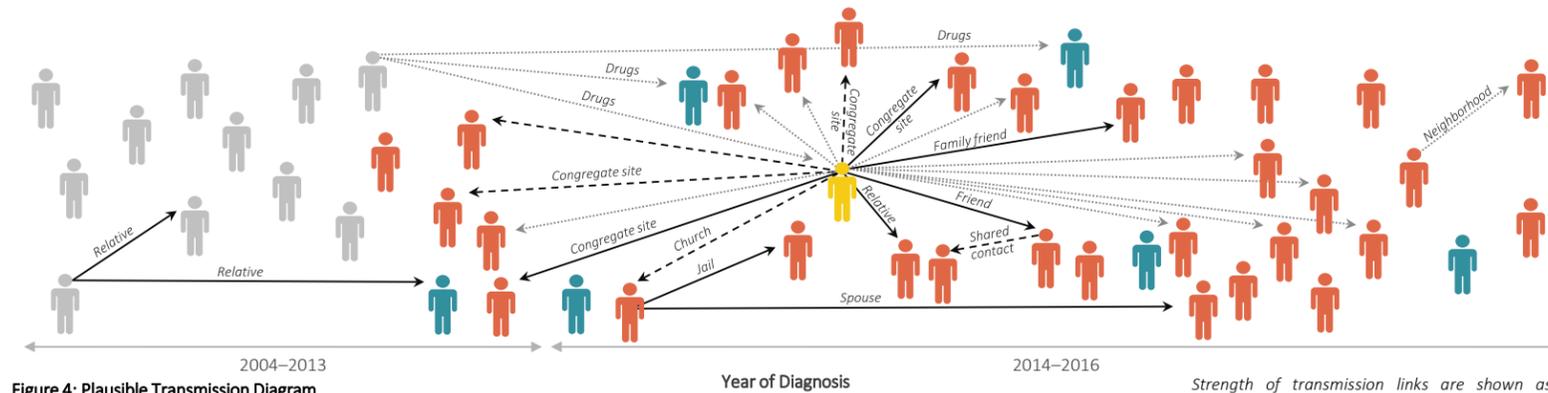


Figure 4: Plausible Transmission Diagram

Key findings:

- Integrated analysis of epidemiologic, molecular, and clinical data for each case identified a plausible source for 26 of 38 (55%) cases.
- Recent transmission was ruled out for 6 of 38 (16%) cases in the outbreak period (shown in teal).
- The proposed transmission diagram suggested seven likely transmission sites; one central case was identified as the most likely source for 14 secondary cases.

## LIMITATIONS

- Non-genotyped and clinically diagnosed cases without known epidemiologic links to an outbreak case could not be included, but may be involved in recent transmission.
- Imprecise or inaccurate symptom onset and epidemiologic data can affect the plausible transmission pathways identified.

## CONCLUSIONS

- TB outbreak investigations are complex, resource-intensive, and can require analysis of multiple types of data.
- The integrated analysis revealed that a few cases were responsible for a majority of transmission events, including one patient who had highly infectious TB and a long infectious period. This was key to understanding transmission in this outbreak.
- Staff could prioritize cases and associated contacts by identifying specific locations and time periods when transmission likely occurred.
- Data visualization aided in communicating findings to key stakeholders and directing resources to the public health response.

## RECOMMENDATIONS

- To enhance national capacity for data integration and visualization, a systematic evaluation of resources and program requirements is needed.

## CONTACT INFORMATION

Clint McDaniel  
 Email: cmcdaniel1@cdc.gov  
 Phone: (404) 639-6156

This work was made possible through support from the Advanced Molecular Detection (AMD) program at CDC.

National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention  
 Division of Tuberculosis Elimination

