

TITLE: Whole Genome Sequencing of *Mycobacterium tuberculosis* for the Investigation of a Tuberculosis Outbreak Involving Prison and Community Cases in Florida, USA

AUTHORS AND AFFILIATIONS:

S raphin MN^{1,2}, May RJ^{1,2}, Khan M^{1,2}, Nolan DJ², Morris JG², Salemi M², Lauzardo M^{1,2}

¹Division of Infectious Diseases and Global Medicine, University of Florida

²Emerging Pathogen Institute, University of Florida, Gainesville, FL

Contact Email for Presenter: nseraphin@ufl.edu

CATEGORY: B

RANK OF PRESENTER: 2—Students and Postdoctoral Fellows

ABSTRACT

Introduction: An extensive tuberculosis (TB) outbreak involving U.S-born persons in the prison system and foreign-born persons in the community occurred in Florida, U.S. over a 10 year period (2004-2015). Genotyping by spoligotyping and 24-locus MIRU-VNTR showed the outbreak was clonal but contact investigation could not identify a source. We used whole-genome sequencing (WGS) and phylogenetic analyses to delineate the outbreak.

Hypothesis: In this investigation, we tested the hypothesis that the outbreak bacterial population was more diverse than observed when using traditional genotyping methods.

Methods: We sequenced the complete genomes of 21 of the 74 *Mycobacterium tuberculosis* isolates, which constituted a representative spatial and temporal sample of the outbreak cases. Contigs were ordered and aligned using the laboratory strain CDC1551 as a reference genome and SNPs called de Novo. We compared the mean genetic distances between and within community and prison strains; foreign and US-born cases. A phylogenetic relationship between the strains was investigated using distance-based and maximum likelihood methods. We applied a Bayesian coalescence-based phylogenetic framework implemented in Beast v2.4.1 to infer the timescale of the outbreak and estimate the date of the most recent common ancestor, using a strong prior for the clock rate. A posterior distribution of trees and model parameters were generated by running 5×10^7 Markov Chain Monte Carlo (MCMC) generations, sampling every 5,000 steps.

Results: The genomic data revealed two *M. tuberculosis* lineages with identical spoligotyping and 24-locus MIRU-VNTR profile, suggesting two concomitant TB outbreaks. The lineages descended from a common source of foreign origin. Almost thirty percent of the outbreak cases were HIV co-infected, 37% were incarcerated at the time of diagnosis, 16% had a history of drug use and 4% were homeless in the year prior to diagnosis. These clinical and social factors likely contributed to the outbreak. The Bayesian phylogenetic analyses suggest that the source strain circulated in the community for some time before the clustered cases in the prison system triggered the outbreak.

Conclusions: These data highlight the increasing need for more discriminatory methods such as WGS in the fight against TB in low incidence settings.