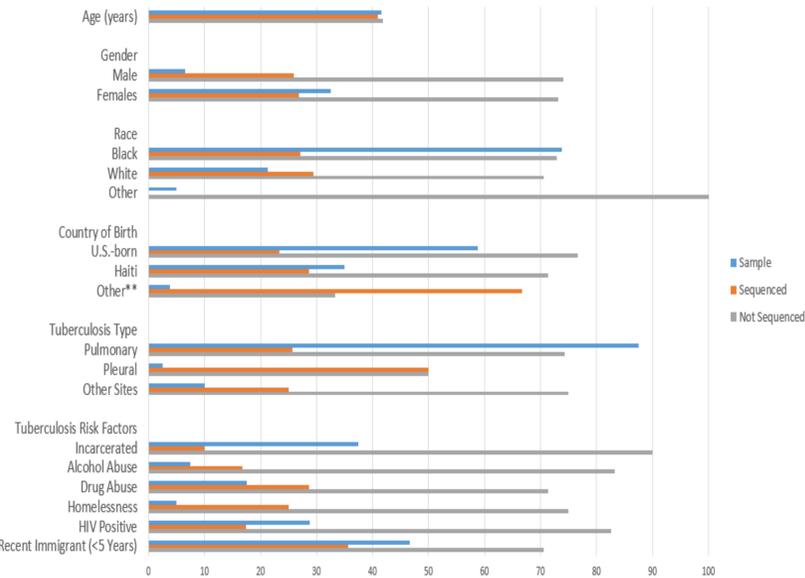


Marie Nancy Séraphin, PhD, MPH <sup>1,2</sup>; R. Justin May <sup>1,2</sup>; Md. Siddiqur Rahman Khan <sup>1,2</sup>; David J. Nolan <sup>2,3</sup>; J. Glenn Morris, MD, MPH&TM <sup>2</sup>; Marco Salemi, PhD <sup>2,3</sup>; Michael Lauzardo, MD, MSc.<sup>1,2</sup>

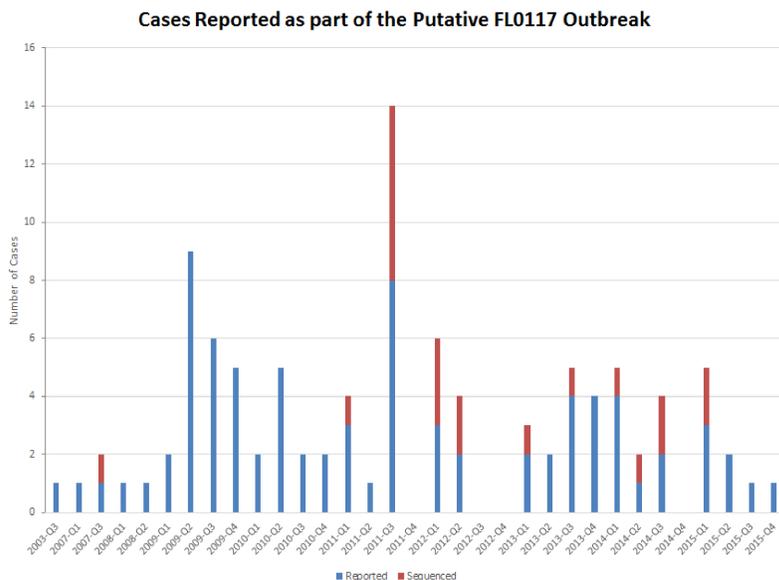
<sup>1</sup>Department of Medicine, Division of Infectious Diseases and Global Medicine, University of Florida  
<sup>2</sup>Emerging Pathogens Institute, University of Florida  
<sup>3</sup>Department of Pathology, Immunology and Laboratory Medicine, University of Florida

## INTRODUCTION

In March 2009, an inmate with pulmonary tuberculosis (TB) was reported to the Florida Department of Health, which triggered an outbreak investigation. Twenty-seven more cases, all U.S.-born, were identified among the prison population between March 2009 and October 2013. Genotyping by Spoligotype and 24-locus MIRU-VNTR on culture confirmed cases showed they all shared identical genotype profile. A review of the Florida TB genotyping management system identified 44 more cases with the same genotype profile in the community, the majority Haitian-born. Contact tracing could not identify a source for the outbreak. We used whole genome sequencing (WGS), phylogenetic analyses and the available contact investigation data to delineate the outbreak. We tested the **hypothesis** that the outbreak bacterial population is more diverse than observed using traditional genotyping methods compared to WGS.



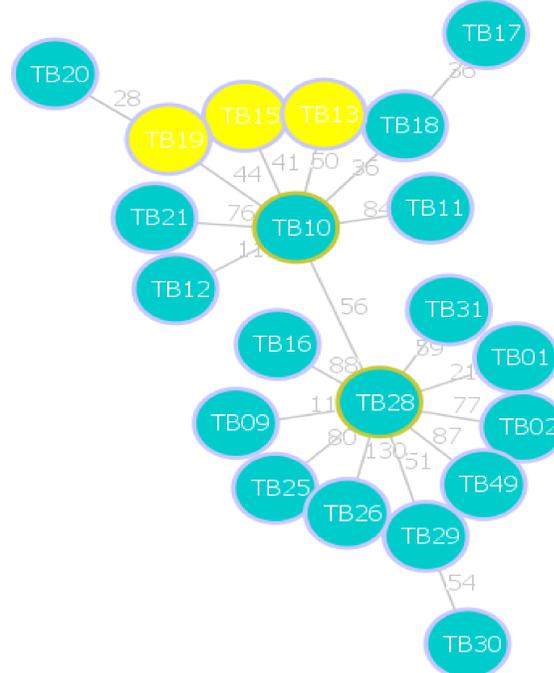
**Figure 1. Characteristics of the Cases Involved in the Outbreak by Whole Genome Sequencing Status.** \*\*Includes one (1) case born in the Dominican Republic, one (1) born in Vietnam and one (1) born in Grenada; \*\* Includes injection and non-injection drug use.



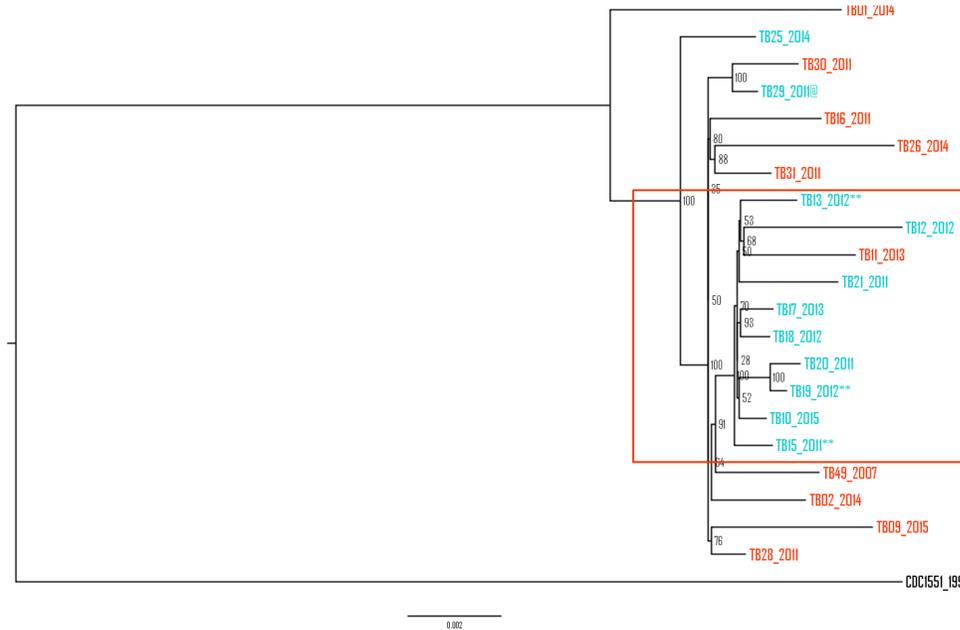
**Figure 2. Epidemic Curve of the Spoligotyping and 24-locus MIRU-VNTR Defined *M. tuberculosis* Outbreak.** Red bars indicate sequenced isolates as a proportion of all outbreak isolates reported in that quarter.

## METHODS

We sequenced 21 of the 74 cases involved in the outbreak, which constituted a representative spatial and temporal sample of the cases. Short paired-end reads were trimmed and independently *de novo* assembled into contigs, then the contigs were ordered, aligned and variable sites (SNPs) called using the reference strain CDC1551. We evaluated the evolutionary history between the sequences by comparing the SNP difference within and between prison and community cases, foreign-born and U.S.-born cases. We investigated the phylogenetic relationship between the strains using distance-based and maximum likelihood methods. We estimated the time to most recent common ancestor (TMRCA) and inferred a timescale for the outbreak using Bayesian coalescence theory implemented in BEAST v2.4.1



**Figure 3. Minimum Spanning Tree (MST) of FL0117 Cases.** Nodes represent each of the sequenced cases (n=21). Cyan identifies cases diagnosed in the community and Yellow identifies the three cases diagnosed while incarcerated. Yellow outline identifies central nodes. The numbers on the branches represent single-nucleotide variant (SNV) between each pairs of isolates.



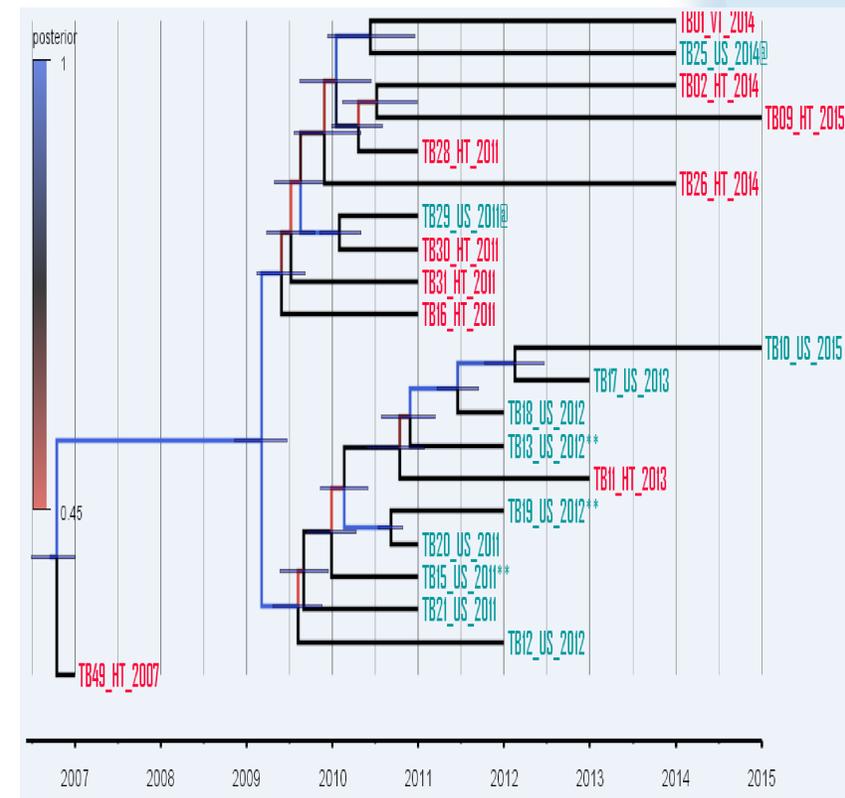
**Figure 4. Midpoint rooted Maximum Likelihood Phylogeny of the 21 FL0117 Isolates and ONE (1) Reference Isolate.** Tip labels colors indicate patient birth origin: Blue indicates U.S.-born cases, Red indicates Foreign-born cases and the reference strain is in Black; @ indicate pediatric case with Haitian guardian; \*\*denotes cases diagnosed while incarcerated; all others were diagnosed in the community. Numbers indicate bootstrap support for clusters. The analysis assumes the sequences are evolving according to the Kimura 3-parameter model.

## FINDINGS/CONCLUSIONS

- The genomic data showed two bacterial populations, suggesting two concomitant TB outbreaks
- The two populations descent from a common source of foreign origin
- It is likely that the source strain circulated in the community for some time before the clustered cases in the prison triggered the outbreak
- Almost 30% of the 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.
- These data highlight the increasing need for systematic genotyping using WGS in the fight against TB in low incidence settings.

## ACKNOWLEDGEMENTS

We gratefully acknowledge the contribution of the staff from the Florida Department of Health; The Florida Bureau of Public Health Laboratories (BPHL), and the County Health Departments staff that collected the data used in this study. This study was funded by the College of Medicine, University of Florida.



**Figure 5. Maximum Clade Credibility Phylogeny of the FL0117 Cluster illustrating the relationship between U.S.-born (Blue) and Foreign-born cases (Red) cases.** Results assume a strict molecular clock with a strong prior of  $3.03 \times 10^{-10}$ ; \*\* indicates cases diagnosed while incarcerated; all others were diagnosed in the community; @ indicated pediatric case born to Haitian parents. Blue horizontal bars indicate the level of uncertainty in the node age estimates. Branches colors represent the posterior distribution of trees. The analysis assumed the sequences are evolving according to the General Time Reversible (GTR) model.