Comparison as a Means to Optimization: A Ring Trial of Whole Genome Sequencing for Outbreak Detection in U.S. Public Health Laboratories.

Sharvari Narendra[1], Katie Barry[1], Shireen Kotay[1], Thomas J. X. Li[1], Salvador Castañeda-Barba[1], Emily Snavely[1], Amy J Mathers[1]
Division of Infectious Diseases and International Health, Departments of Medicine and Pathology, School of Medicine, University of Virginia

Background: The potential of pathogen genomics in public health is rapidly expanding. Public Health Laboratories (PHLs) are evolving to build infrastructure and establish standardized methods. To understand the current landscape of public health response to nosocomial outbreaks, we oversaw a Ring Trial to assess the whole genome sequencing (WGS) approaches across six PHLs under the Centers for Disease Control and Prevention (CDC)- Pathogen Genomics Centers of Excellence Network (PGCoE) with UVA as an academic partner. Ring trials are inter-laboratory exercises to compare methods and results using the same sample-set.

Methods: A contrived case study outlining a hospital outbreak was designed using 10 previously characterized, fully sequenced (both long and short read with closed genomes) carbapenemase-producing *Klebsiella pneumoniae* and one *Klebsiella quasipneumoniae* clinical isolates. The raw FASTQ files generated from an Illumina (San Diego, CA) platform and a survey were shared with CDC, Virginia Division of Consolidated Laboratory Services, Massachusetts Department of Health, Georgia Department of Health, Minnesota Department of Health and Washington Department of Health.

Results: While speciation differed, the overall bioinformatics and wet-lab approaches and results were similar across PHLs. The WGS speciation results showed that 57% of the PHLs correctly identified the *K. quasipneumoniae* isolate. We investigated the bioinformatics tools and databases used and found that Kraken (v1.1.1) and 16S methods were unreliable in *K. quasipneumoniae* speciation. While the Kraken database used (v2017019_4GB) contains the isolate, Kraken1 assigns low number of reads to it. The isolate was identified on all instances using Kraken2 however, thus indicating differences in Kraken algorithms between versions. Older databases that do not contain *K. quasipneumoniae*, like Rfam (RF00177) and REFSEQ (v20150430) were used in some cases. Consensus reporting of species despite using a combination of tools could also account for misidentification.

Conclusion: The first iteration of the Ring Trial revealed gaps in WGS analysis across PHLs with respect to speciation. Differences in software versions and databases could account for some variability in the results, thus emphasizing the importance of using updated bioinformatics resources in WGS analysis. Ongoing and future iterations of the Ring Trials will expand on these findings, aiming to refine and inform best practices and enhance outbreak detection capabilities across PHLs.