Visualization and Retrospective Analysis of Whole Genome Sequencing Data from Carbapenemase-Producing Organisms in the Northeast U.S.

Catharine Prussing[1], Kailee Cummings[1], Lauren Rose[1], Shannon Morris[1], Cyndel Vollmer[1], Wolfgang Haas[1], Kara Mitchell[1], Elizabeth Nazarian[1], Kimberlee Musser[1]

[1]Wadsworth Center, NYSDOH, Albany, NY

Objectives

1. To retrospectively analyze whole genome sequencing (WGS) data from carbapenemase-producing bacterial isolates sequenced at the Wadsworth Center, which is the Northeast Regional Lab for the Antimicrobial Resistance (AR) Lab Network.
2. To assess relatedness among isolates of the same multi-locus sequence type (MLST) that were not necessarily part of the same epidemiological investigation.

Methods

- Beginning in 2017, bacterial isolates were selected for WGS according to priorities established by the AR Lab Network[1]
- Genomic DNA from clinical isolates and isolates from colonization screenings (rectal and tracheal swabs) was sequenced using the Illumina MiSeq or NextSeq platform
- All sequenced isolates were characterized using an in-house developed, clinically validated AR pipeline[2], which identifies a sequence type (ST) using *in silico* MLST[3] and identifies AR genes by querying against 3 databases[4]
- Trends in STs and AR (specifically carbapenemase) gene variants were visualized over time, geography, and organism
  - Groups of >20 isolates with the same ST were pulled for a retrospective relatedness analysis
  - Currently, relatedness analyses for AR isolates are only conducted if requested by epidemiologists, usually for infection control
- ST groups were analyzed using an in-house developed pipeline, LegioCluster[5], which assesses relatedness of isolates based on mutation events (MEs, the sum of single nucleotide polymorphisms (SNPs) and short insertions/deletions)
  - An internal reference genome was selected for each ST group
- Analyses for two ST groups were selected for visualization

References

1. CDC Antibiotic Resistance Laboratory Network: Whole Genome Sequencing of Healthcare-Associated Infection/Antibiotic Resistant Pathogens in State and Local Public Health Laboratories. June 2020.
2. Prussing *et al.* 2020. *Front Microbiol* 11:2007.
3. https://github.com/tseemann/mlst
4. https://github.com/tseemann/abricate
5. Haas *et al.* 2020. *J Clin Microbiol* 59:e00967-20.
6. Treangen *et al.* 2014. *Genome Biology* 15:524.

Results

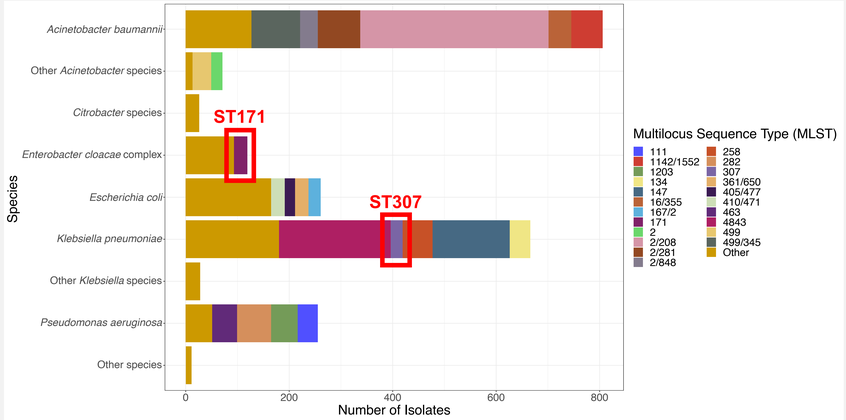

Figure 1. MLST by bacterial species. As of July 19, 2023, 2,276 isolates had been sequenced, representing 33 bacterial species, 277 MLST sequence types (STs), and 40 carbapenemase gene variants. All ST groups with <20 isolates are listed as "Other". Relatedness analysis results for STs highlighted in red boxes are shown below.

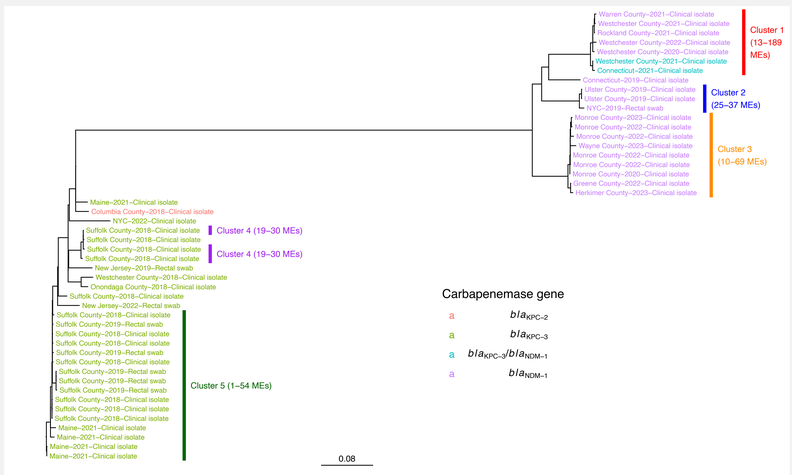

Figure 2. Relatedness analysis of ST171 *Enterobacter cloacae* complex isolates. Phylogenetic tree comparing genome sequences generated by Parsnp[6]. Node labels indicate the isolate location (New York State county or other Northeast Region state), year of collection, and isolate source (clinical isolate or rectal swab). Node label text colors indicate the carbapenemase gene variant(s) carried by the isolate. All clusters in which there are isolates separated from each other by <50 MEs are indicated with vertical bars. Cluster 5 consists of isolates from two separate investigations from two different states that were not previously known to be related to each other.

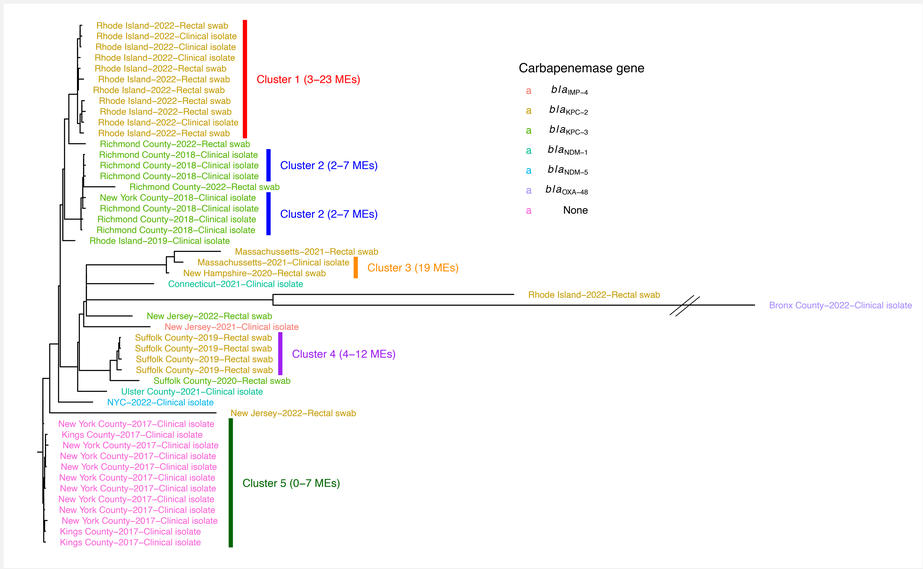

Figure 3. Relatedness analysis of ST307 *Klebsiella pneumoniae* isolates.

Phylogenetic tree comparing genome sequences generated by Parsnp[6]. Node labels indicate the isolate location (New York State county or other Northeast Region state), year of collection, and isolate source (clinical isolate or rectal swab). Node label text colors indicate the carbapenemase gene variant carried by the isolate. All clusters in which there are isolates separated from each other by <50 MEs are indicated with vertical bars. Cluster 3 consists of two isolates from two different states that were not previously known to be related to each other.

Conclusions

- WGS analysis of AR isolates has utility both for characterizing AR genes present in individual isolates at the gene variant level, and for monitoring trends in surveillance data and assessing relatedness to inform infection control and outbreak investigations.
- In our retrospective analysis of AR WGS data from the Northeast Region of the AR Lab Network, we identified 22 groups of isolates with the same ST that are being analyzed further to investigate relatedness among isolates not previously known to be epidemiologically linked.
- Analysis of ST171 *E. cloacae* complex isolates and ST307 *K. pneumoniae* isolates identified two clusters containing isolates from different states that were not previously known to be closely related.
- Ongoing prospective analysis of relatedness among AR isolates may identify previously unknown links between isolates that could help in understanding pathogen transmission.

Acknowledgements

- We thank the Wadsworth Center Advanced Genomic Technologies Center for conducting sequencing, Emily Snavely (former APHL-CDC AR fellow) for initial development of the AR pipeline, and epidemiology and public health laboratory colleagues from the Northeast Region for their collaboration.
- This work was supported by the New York State Department of Health, Cooperative Agreement Number NU50CK000423 funded by the Centers for Disease Control and Prevention (CDC), and Cooperative Agreement U60OE000103 funded by CDC through the Association of Public Health Laboratories.