Integrating Next Generation Technology into Healthcare-Associated Infection Outbreak Investigations

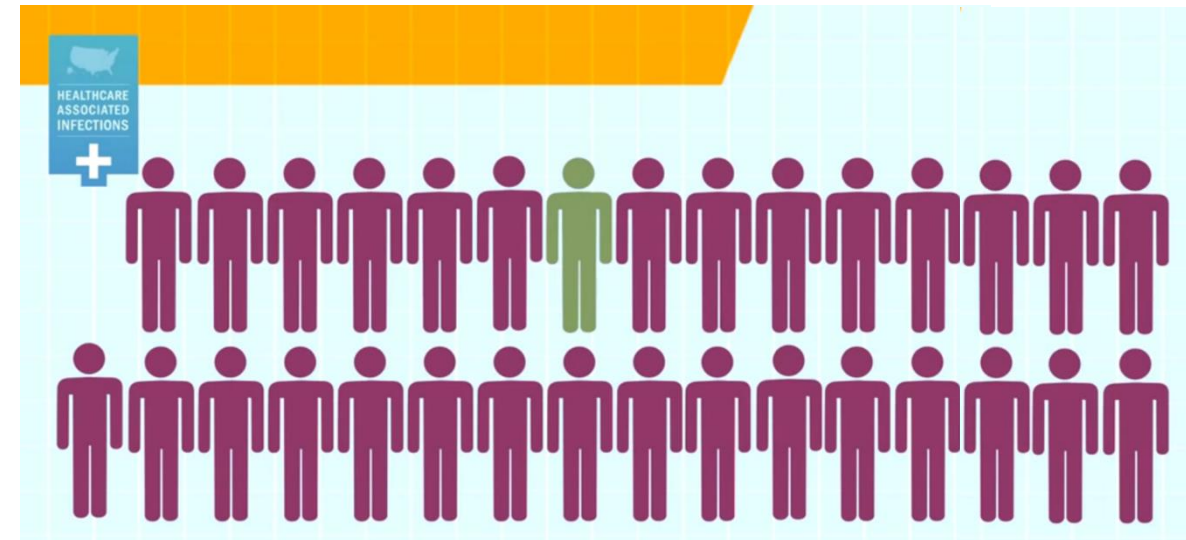**Division of Healthcare Quality Promotion**

Rich Stanton

Gillian McAllister

Division of Healthcare Quality Promotion

- DHQP focuses on healthcare associated infections (HAIs)

- Clinical and Environmental Microbiology Branch (CEMB)

- Metagenomics and Molecular Biology (MMB)

  - Gillian McAllister – Sequencing Activity Lead

  - Rich Stanton – Molecular Biology Activity Lead
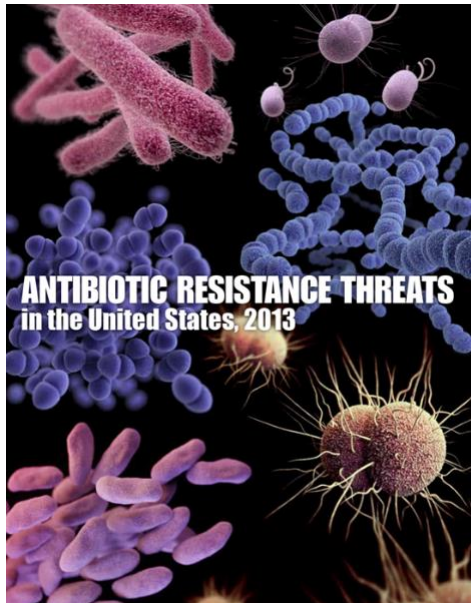
On any given day, about 1 in 31 hospital patients has at least one healthcare-associated infection (HAI).

ANTIBIOTIC RESISTANCE THREATS
in the United States, 2013

Estimated minimum number of illnesses and deaths caused annually by antibiotic resistance*:

At least

2,049,442 illnesses

23,000 deaths

*bacteria and fungus included in this report

HEALTHCARE ASSOCIATED INFECTIONS

Healthcare Pathogens Incredibly Diverse

- **Gram-positive bacteria**
  - *Staphylococcus*
  - *Enterococcus*
  - Nontuberculous mycobacteria
- **Spore-formers**
  - *Clostridioides difficile*
  - *Clostridium sordellii*

- **Gram-negative bacteria**
  - Enterobacteriaceae (>50 genera)
    - Carbapenem-resistant Enterobacteriaceae (CRE)*
      - *E. coli, Klebsiella, Enterobacter, Citrobacter, Serratia, Morganella, Providencia)*
    - *Burkholderia*
  - *Acinetobacter*
    - Carbapenem-resistant *Acinetobacter baumannii* (CRAB)*
  - *Pseudomonas aeruginosa*
    - Carbapenem-resistant *Pseudomonas aeruginosa* (CRPa)*

Often discover novel species, or strains of previously identified species

➢**Untenable to sequence every HAI or MDRO pathogen**

Diversity Complicates WGS for Outbreak Investigations

- Complex biology:
  - Multiple routes of transmission
  - Ubiquitous in healthcare environment (sinks, devices, products, asymptomatic carriers, etc.)
  - Heterogenous infections
  - Vertical (clonal expansion) and horizontal (plasmids) transmission of AR genes
- Structural factors:
  - Persistent outbreaks (weeks – years)
  - Single facility, regional, national, international
- WGS limitations:
  - Sampling limitations (single sample/patient, no asymptomatic isolates, limited environmental samples)
  - No standards for comparison of relatedness for all species
  - Different growth rates in different environments

Diversity Reflected Outbreak Investigations

- 41 different species from 20 genera sequenced as part of 102 HAI outbreak investigations over last six years

- 21 outbreak investigations involved multiple species

- Sequenced anywhere from 2-70 (median: 8) isolates per outbreak investigation

WGS of HAI Outbreaks Counts

Year	Count	# of Species	Total Isolates
2014	2	2	40
2015	9	14	197
2016	17	10	146
2017	20	15	334
2018	23	18	299
2019	31	21	351
Total	102	46	1367

WGS of HAI Outbreaks by Pathogen

Species	Count
Pseudomonas aeruginosa	22
Klebsiella pneumoniae	21
Acinetobacter baumannii	16
Escherichia coli	16
Mycobacterium chimaera	10
Staphylococcus aureus	10
Enterobacter cloacae	5

Multiple Approaches to Determine Relatedness

- Initial processing through QuAISAR-H pipeline

- Multilocus Sequence Typing (MLST)

- Core Genome Multilocus Sequence Typing (cgMLST)

- High Quality Single Nucleotide Polymorphisms (hqSNPs)

- AR Genes and Plasmid Replicons Overlaps

Standard WGS HAI Processing: QuAISAR-H Pipeline

- **Q**uality control, **A**ssembly, species **I**dentification, **S**equence typing, **A**nnotation, and **R**esistance mechanisms for **H**ealthcare-associated pathogens

- Automates routine evaluation

- Provides an easy, all-in-one tool for initial processing and analyses of raw sequence data

- Available through AMD Portal (amdportal-sams.cdc.gov)
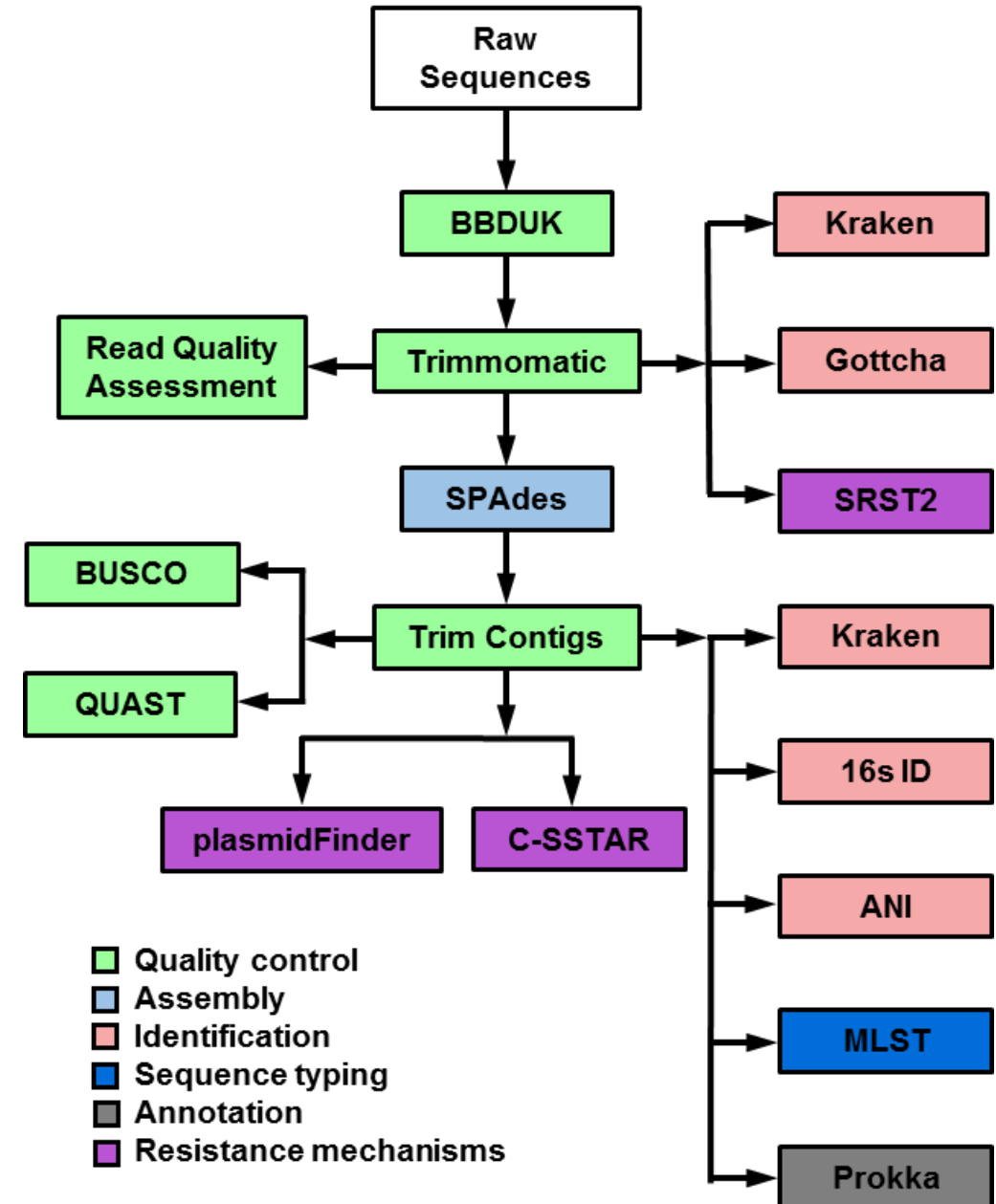
MLST/cgMLST

- Multilocus Sequence Typing MLST (pubmlst.org)
  - Compares of alleles of 7 housekeeping genes and names sequence type (ST)
  - Command line tool (github.com/tseemann/mlst)
- Core Genome MLST (cgMLST)
  - Compares of alleles of 1000s of genes common to a specific species
  - Implemented through BioNumerics
- Both can be used to rapidly identify clusters for hqSNP analysis

cgMLST tree of *P. aeruginosa* isolates from KPC-outbreak in single facility.
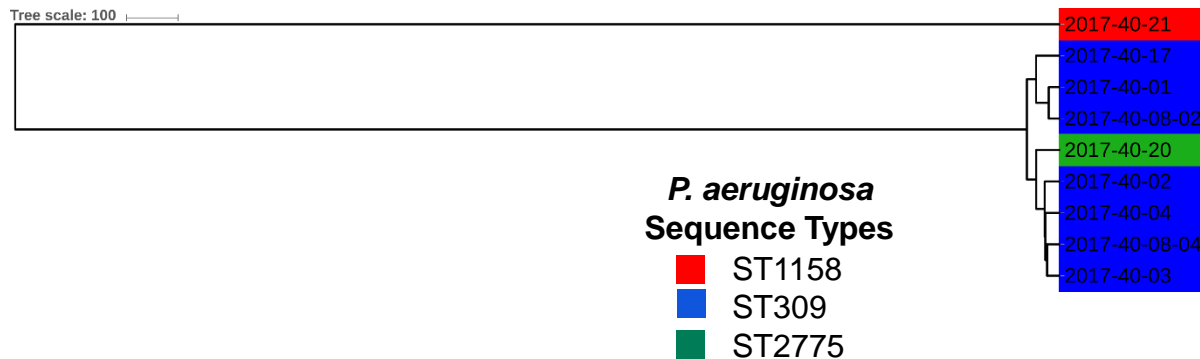
cgMLST tree of *K. pneumoniae* isolates from KPC-outbreak in single facility.
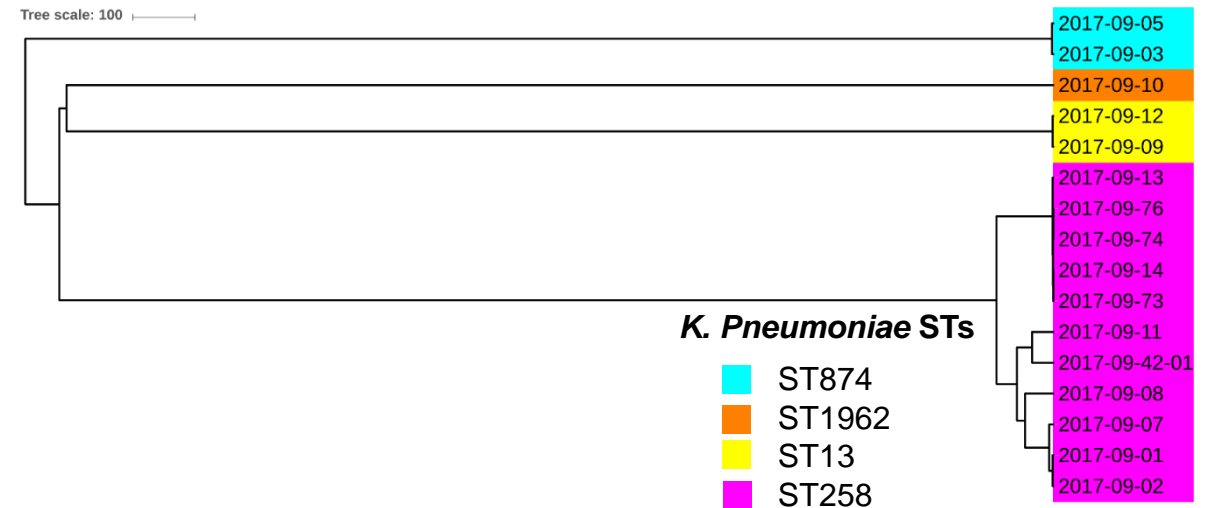
High Quality SNPs

- hqSNPS determined over core genome through SNVPhyl (github.com/phac-nml/snvphyl-galaxy)
- Includes core genome % calculation (varies with species but increases confidence in calls)
- Median outbreak genome by Mash distance (github.com/marbl/Mash) used as mapping reference

Tree scale: 0.1

0 SNPs

Color Key
- ST258 outbreak isolates
- AR Bank control isolates

Core Genome Size: 89.94%

SNP tree of *K. pneumoniae* ST258 isolates from KPC-outbreak in single facility.

0-11 SNPs

2017-09-76
2017-09-74
2017-09-73
2017-09-14
2017-09-13
2017-09-11
2017-09-42-01
ARB0364
2017-09-08
ARB0453
2017-09-02
2017-09-07
2017-09-01
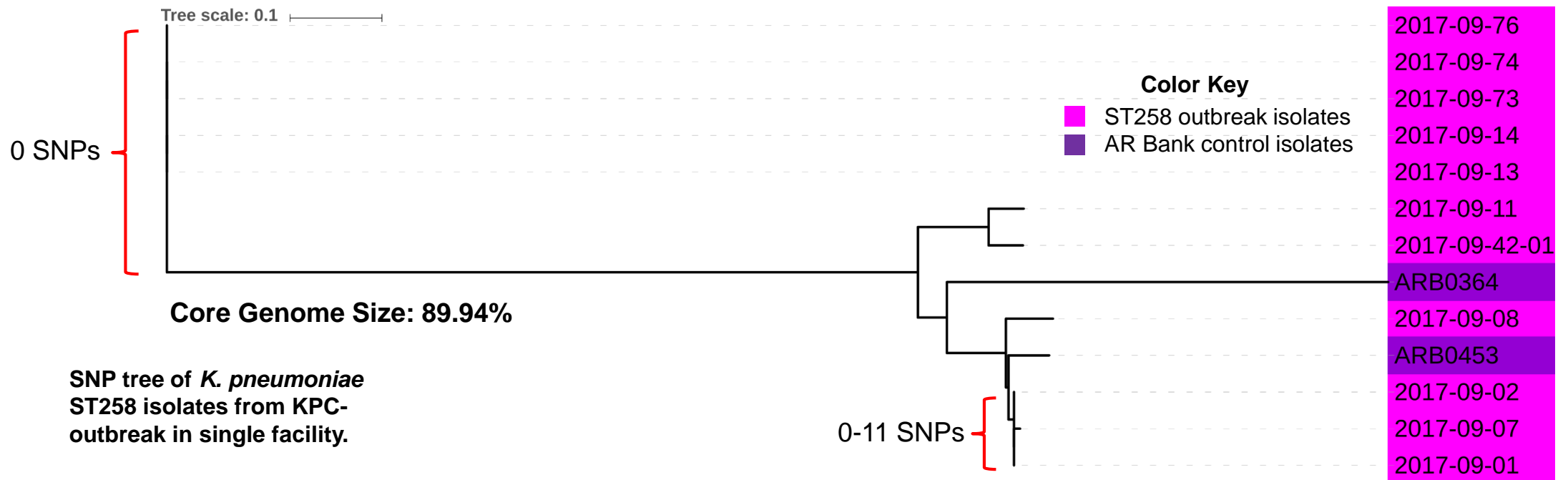
AR Gene/Plasmid Replicons Overlaps

- AR genes called using c-SSTAR (github.com/chrisgulvik/c-SSTAR) and srst2 (github.com/katholt/srst2)

- ResFinder (cge.cbs.dtu.dk/services/ResFinder/) and ARG-ANNOT (ifr48.timone.univ-mrs.fr/blast/arg-annot_nt.html) databases

- Common genes in unrelated isolates suggests horizontal transfer

cgMLST Trees from multiple species/STs from KPC-outbreak in single facility.

Sequence Types
- *K. pneumoniae* ST874
- *K. pneumoniae* ST1962
- *K. pneumoniae* ST13
- *K. pneumoniae* ST258
- *E. coli* ST69
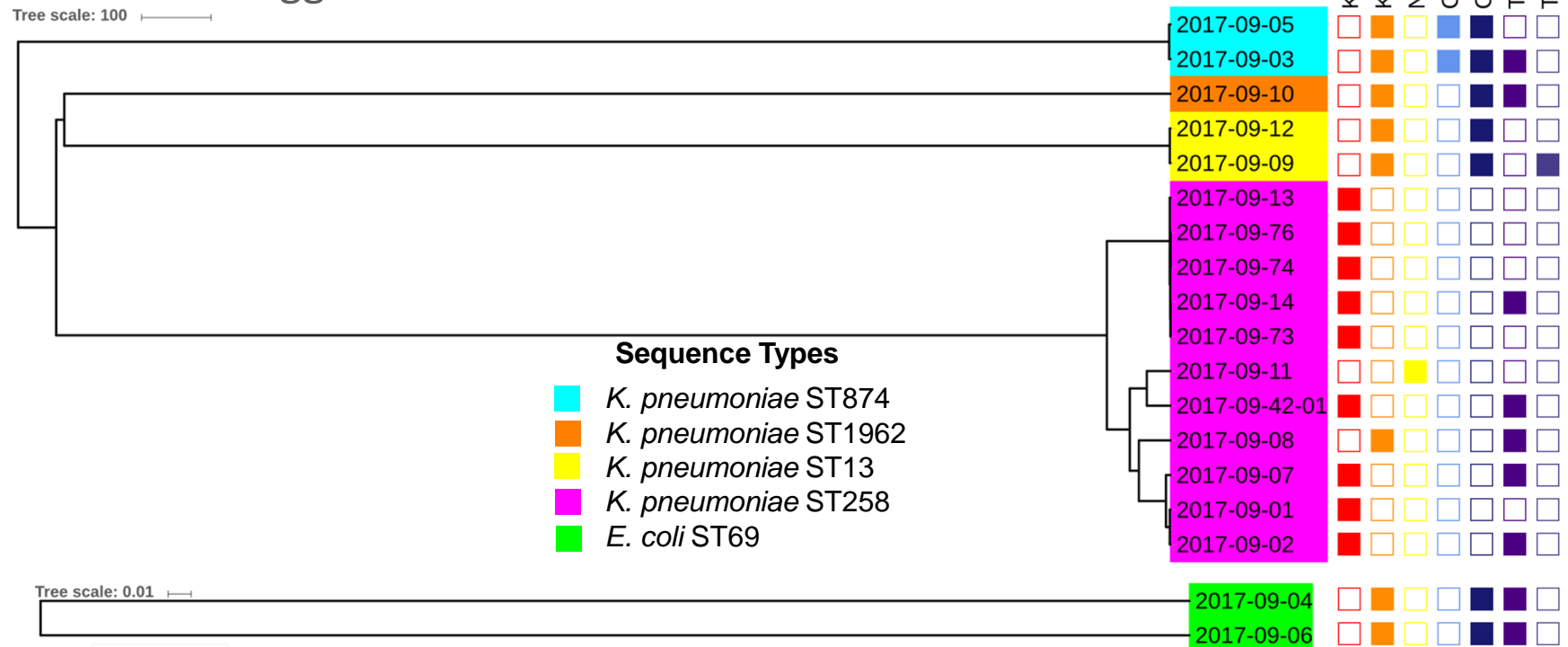
AR Gene/Plasmid Replicons Overlaps

- Plasmid replicons identified through PlasmidFinder (cge.cbs.dtu.dk/services/PlasmidFinder/)
- Long read sequencing can resolve plasmids, but overlapping genes and markers (especially on the same contig) can be used as indicators of horizontal transfer

Short read sequences mapped to plasmids from KPC-outbreak.

Sequence Types
- *K. pneumoniae* ST874
- *K. pneumoniae* ST1962
- *K. pneumoniae* ST13
- *K. pneumoniae* ST258
- *E. coli* ST69

KPC-2+ Plasmid
127,681 bp

KPC-3+ Plasmid
163,693 bp
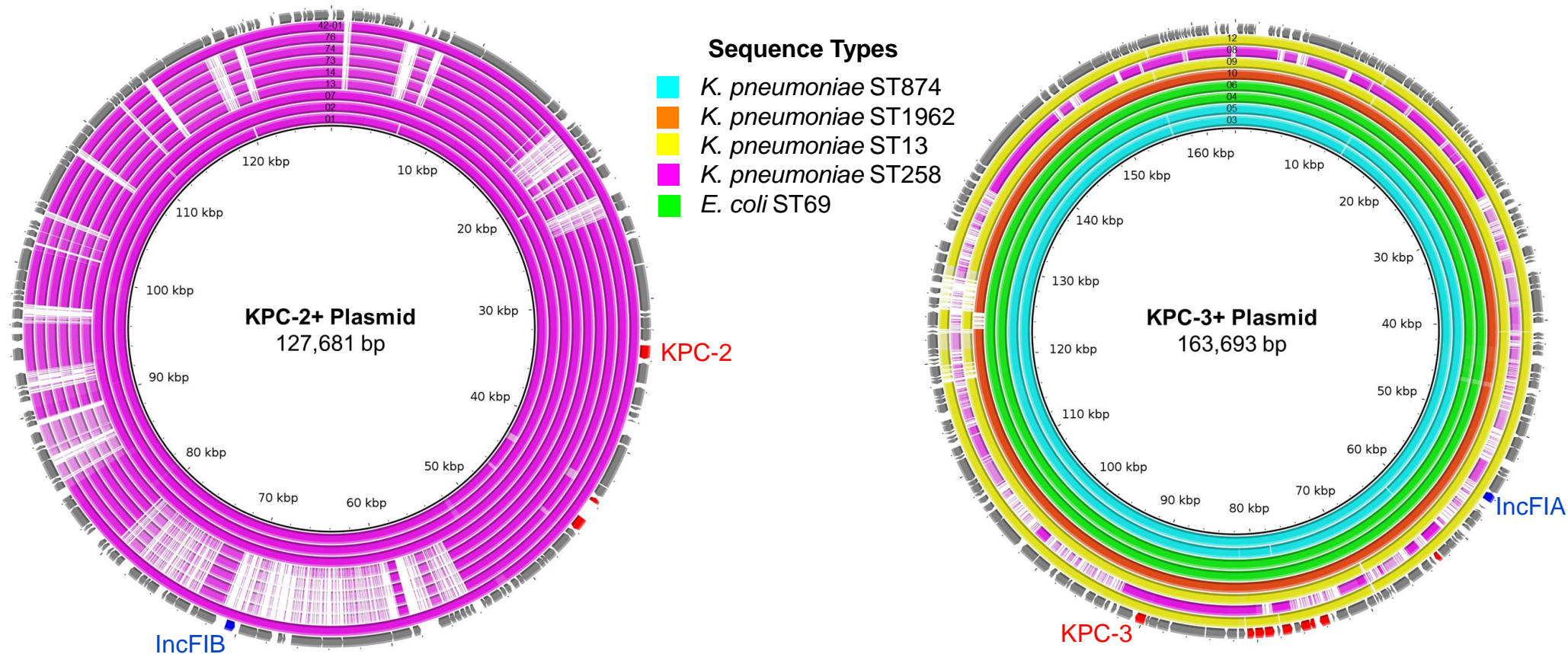
Best Practices/General Tips

- Because of the diversity of outbreaks, a variety of approaches strengthens conclusions
- Non-epidemiologically related isolates from the same ST can reveal clustering
- Relative SNP/gene distances rather than cutoffs of relatedness
- Always incorporate epidemiological data

cgMLST tree shows that putative outbreak isolates all cluster distinctly from control isolates from same ST.
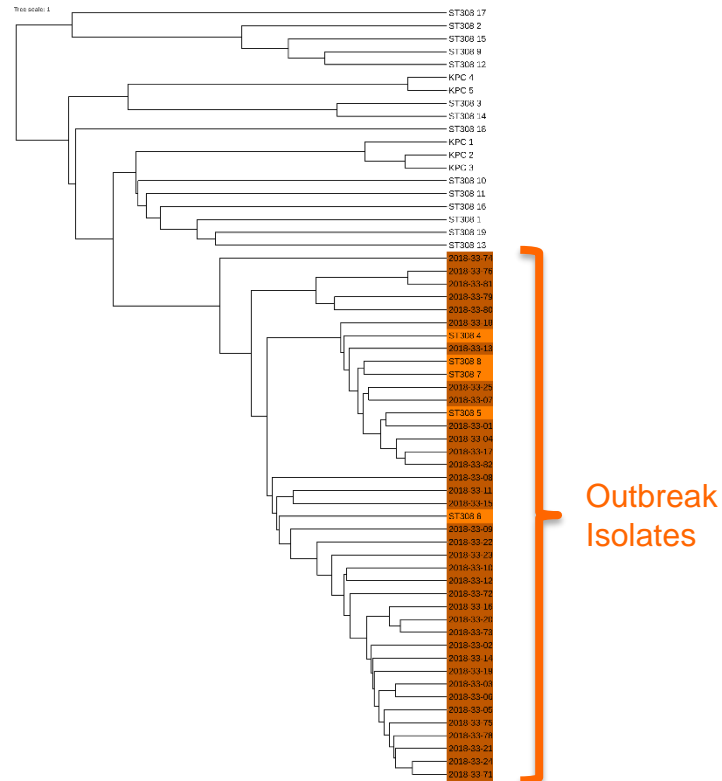
SNP tree from same outbreak shows wide range in relatedness (0-105 SNPs), but clusters that correlate with epi data.
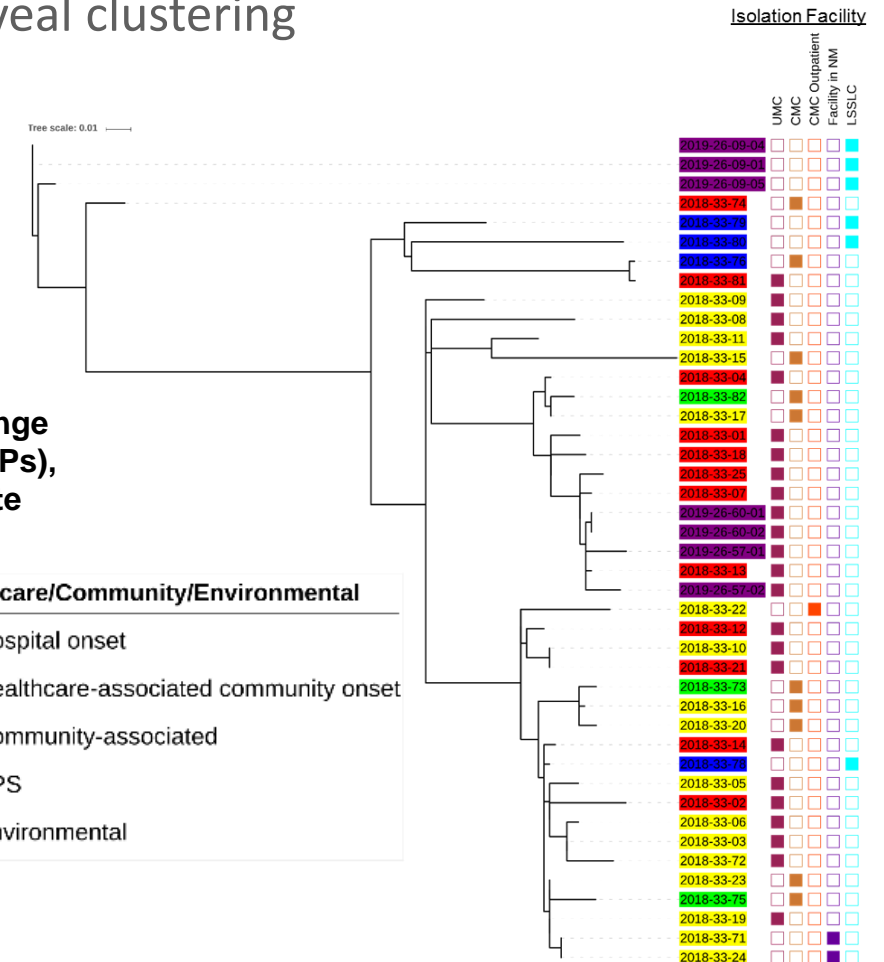
Best Practices/General Tips

- Standardized outbreak analysis reports include STs, AR Genes, Plasmid Replicons, SNP Matrices, and trees

- Provide epidemiologists an overall summary of the WGS analysis

Umbrella BioProject for Posting of WGS Data to NCBI

CDC HAI-Seq

Accession: PRJNA531911   ID: 531911

CDC's Division of Healthcare Quality Promotion-hosted Umbrella for any sequence data (genome, plasmid, amplicon, shotgun) generated by CDC and other public health laboratories that are related to healthcare-associated infections (HAIs). Can include any data format, including raw reads and/or assemblies. We have implemented a standard scheme for isolate/sample identification. We request for those sites interested in submitting data to this umbrella project, please contact: HAISeq@cdc.gov to be assigned your unique code. DHQP has guidance available for posting metadata, which can be provided upon request. In general, submitters should not post any metadata that compromises confidentiality. Questions? Please email HAISeq@cdc.gov.

Accession	PRJNA531911
Type	Umbrella project
Submission	Registration date: 10-Apr-2019 **Centers for Disease Control and Prevention. Division of Healthcare Quality Promotion**
Relevance	Medical

https://www.ncbi.nlm.nih.gov/bioproject/531911

Questions?

Metagenomics and Molecular Biology, Division of Healthcare Quality Promotion
HAIseq@cdc.gov