

CLIA Compliant Validation of WGS in a Public Health Laboratory

Cassandra Campion

Assay Development Supervisor

Indiana State Department of Health
Laboratories

May 23, 2019



Indiana State
Department of Health

Outline

- Introduction
- ISDH validation:
 - Study strategy
 - Accuracy
 - Precision



Indiana State
Department of Health

Introduction

- NGS was introduced to public health in 2013 when PulseNet began using NGS to detect *Listeria* outbreaks.
- Since then, most state public health labs in the U.S. have gained access to NGS.
- Now many CDC programs incorporate NGS, usually with their own analysis pipelines (eg., GHOST, *Legionella*).
- How can a public health/clinical lab bring in NGS on their own?



Indiana State
Department of Health

Challenges to Adopting NGS

- Bioinformatics:
 - Infrastructure — Many labs do not have network or computing power to perform full range of bioinformatics analyses.
 - Bioinformaticians — Many labs do not have financial ability to hire a full-time bioinformatician.



Indiana State
Department of Health

Challenges to Adopting NGS (cont.)

- Validation cost — Relatively high cost does not allow many clinical laboratories to perform large validation sets or multi-laboratory studies.
- Proficiency testing for clinical and public health laboratories is still in development.



Indiana State
Department of Health

NGS Implementation Guidelines

- Next-generation sequencing:
Standardization of Clinical Testing (Nex-StoCT) workgroup (CDC):
 - Developed principles and guidelines for validation, QC, PT, and reference materials.
 - Workgroup focused on heritable human disorders, but may offer insight to NGS implementation in clinical and public health laboratories.



Indiana State
Department of Health

NGS Implementation Guidelines

- CAP Molecular Pathology Checklist:
 - Includes wet and dry bench portions
- CLSI MM09 (Nucleic Acid Sequencing Methods in Diagnostic Laboratory Medicine)



Indiana State
Department of Health

The ISDH CLIA-Compliant Validation

- ISDH goal:
 - Validate the MiSeq and CLC Bio Genomic Workbench using the Nextera FLEX kit for outbreak analysis
- ISDH validation modeled after a paper written by the California Department of Public Health (Kozyreva *et al.*):
 - Full California validation available in supplementary documents



Indiana State
Department of Health

Kozyreva, V. K., Truong, C. L., Greninger, A. L., Crandall, J., Mukhopadhyay, R., & Chaturvedi, V. (2017). Validation and Implementation of Clinical Laboratory Improvements Act-Compliant Whole-Genome Sequencing in the Public Health Microbiology Laboratory. *Journal of clinical microbiology*, 55(8), 2502-2520.

ISDH Validation



Indiana State
Department of Health

How Did Our Validation Differ from California?

California

- 31 isolates
- In-house developed pipeline
- Additional analysis available

Indiana

- 10 isolates
- Commercial software
- Limited analysis

ISDH Validation Study Strategy

- 10 total isolates, including:
 - 5 *Enterobacteriaceae* (CP-CREs)
 - 1 VPD (*H. influenzae*)
 - 1 Gram positive (*Staph aureus*)
 - 1 Actinomycete (*Nocardia* sp.)
 - 1 *Mycobacterium* (non-MTBC)
 - 1 Enteric (*Salmonella*)
- Isolates were chosen based on frequency of receipt and likelihood of need.



Indiana State
Department of Health

ISDH Validation Study Strategy

- Run technical triplicates on the same cartridge for each isolate (all 10 isolates):
 - Single extraction per isolate
 - Independent library preps
 - Single sequencing run
- Operator variance (5 isolates):
 - Used the same extract as the technical triplicates
 - Library prep performed by second analyst
 - Single sequencing run



Indiana State
Department of Health

What Did We Measure?

- Platform accuracy
- Application accuracy
- Bioinformatics pipeline accuracy
- Platform repeatability
- Platform reproducibility
- Application repeatability
- Application reproducibility



Indiana State
Department of Health

Accuracy



Indiana State
Department of Health

Platform Accuracy

- Defined as the accuracy of base calls against a reference sequence.
- Technical triplicates for each isolate were sequenced, generated reads mapped to a reference sequence, and hqSNPs identified.
- To rule out differences in strains, SNPs present in all replicates of an isolate were considered “true” and did not count as sequencing error.
- Due to computing limitations, raw sequence reads were analyzed by Nick Florek at Wisconsin State Laboratory of Hygiene.



Indiana State
Department of Health

Application Accuracy

- Assessed four applications:
 - 16S identification
 - Resistance typing
 - *in silico* MLST
 - hqSNP genotyping
- For comparison against gold standard, only one replicate was used per strain (16S and resistance typing).



Indiana State
Department of Health

16S Identification

All 10 sequenced isolates used for analysis with 100% agreement.

Isolate Number	Gold Standard Identification	16S Identification
ARLN 0347	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>
ARLN 0076	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>
ARLN 0349	<i>E. coli</i>	<i>E. coli</i>
ATCC BAA 2146	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>
ARLN 0056	<i>Acinetobacter baumannii</i>	<i>Acinetobacter baumannii</i>
ATCC 10211	<i>Haemophilus influenzae</i>	<i>Haemophilus influenzae</i>
ATCC 29213	<i>Staphylococcus aureus</i>	<i>Staphylococcus aureus</i>
18Myc0103	<i>Nocardia farcinica</i>	<i>Nocardia farcinica</i>
ATCC 25291	<i>Mycobacterium avium</i>	<i>Mycobacterium avium</i>
C17090866	<i>Salmonella enterica</i>	<i>Salmonella enterica</i>

Resistance Typing

Five ARLN/ATCC isolates used for analysis with 100% agreement.

ARLN/ATCC Number	ARLN/ATCC Resistance Mechanisms	Detected Resistance Mechanisms
ARLN 0347	KPC-3	KPC-3
ARLN 0076	VIM-1	VIM-1
ARLN 0349	mcr-1, CTXM-14, CTXM-55, TEM-1B	mcr-1, CTXM-14, CTXM-55, TEM-1B
ATCC BAA 2146	NDM-1	NDM-1
ARLN 0056	OXA-23, OXA-66	OXA-23, OXA-66

in silico MLST

- ISDH does not conduct MLST testing, and therefore did not know the MLST schemes of the validation isolates.
- To determine the accuracy of the *in silico* MLST tool, sequences used in the Kozyreva *et al.* paper were pulled from NCBI.
- MLST tool was considered in agreement if results obtained from CLC Bio matched published scheme.
- 15 sequences used for analysis with 100% agreement.



Indiana State
Department of Health

in silico MLST

California Isolate Number	California MLST Scheme	ISDH MLST Scheme
C1 (EDL-933)	ST-11	ST-11
C2 (ATCC 7966)	ST-1	ST-1
C3 (ATCC 8739)	ST-3021	ST-3021
C4 (ATCC 13047)	ST-1	ST-1
C5 (ATCC 25923)	ST-243	ST-243
C6 (ATCC 14028)	ST-19	ST-19
C46 (ATCC 29212)	ST-30	ST-30
C47 (ATCC 12228)	ST-8	ST-8
C51 (ATCC 13637)	ST-14	ST-14
C55 (ATCC 25922)	ST-73	ST-73
CDPH C73	ST-11	ST-11
CDPH C74	ST-32	ST-32
CDPH C75	ST-440	ST-440
CDPH C76	ST-592	ST-592
CDPH C72	ST-655	ST-655

hqSNP Genotyping

- Genotyping analysis performed in CLC Bio to determine if software was able to correctly cluster isolates of known relatedness.
- 10 previously sequenced KPC positive ISDH isolates were analyzed with 17 KPC positive sequences provided by CDC as part of known outbreak.



Indiana State
Department of Health

100% agreement between known clusters

CDC isolates (known outbreak)

- SRR2915816 trimmed
- SRR2915817 trimmed
- SRR2915818 trimmed
- SRR2915819 trimmed
- SRR2915820 trimmed
- SRR2915821 trimmed
- SRR2915822 trimmed
- SRR2915823 trimmed
- SRR2915824 trimmed
- SRR2915825 trimmed
- SRR2915826 trimmed
- SRR2915827 trimmed
- SRR2915828 trimmed
- SRR2915829 trimmed
- SRR2915830 trimmed
- SRR2915831 trimmed
- SRR2915832 trimmed

- C16032836_S2_L001_R1_001
- C16032862_S3_L001_R1_001
- C16032887_S7_L001_R1_001
- C16033634_S8_L001_R1_001
- C16033795_S6_L001_R1_001

ISDH isolates
(confirmed outbreak)

- C16032864_S5_L001_R1_001
- C16032835_S1_L001_R1_001
- C16033801_S10_L001_R1_001
- C16033393_S9_L001_R1_001
- C16032841_S4_L001_R1_001

ISDH
isolates
(suspected outbreak)

0.030

Pipeline Accuracy

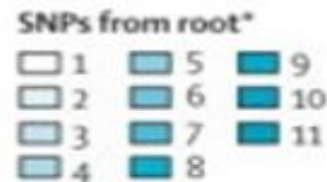
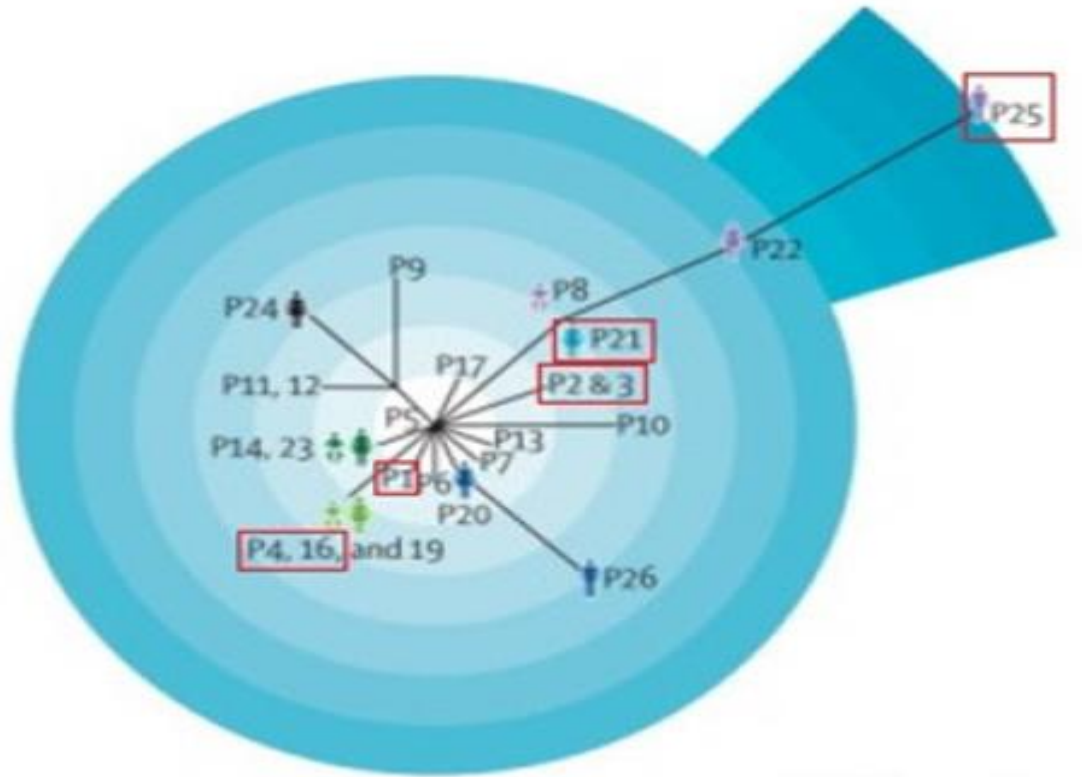
- Phylogenetic analysis performed on raw WGS reads of isolates from well-characterized MRSA outbreak (Harris, *et al.*)
- Seven strains from publication used in ISDH validation
- Validation results compared to published phylogenetic results



Indiana State
Department of Health

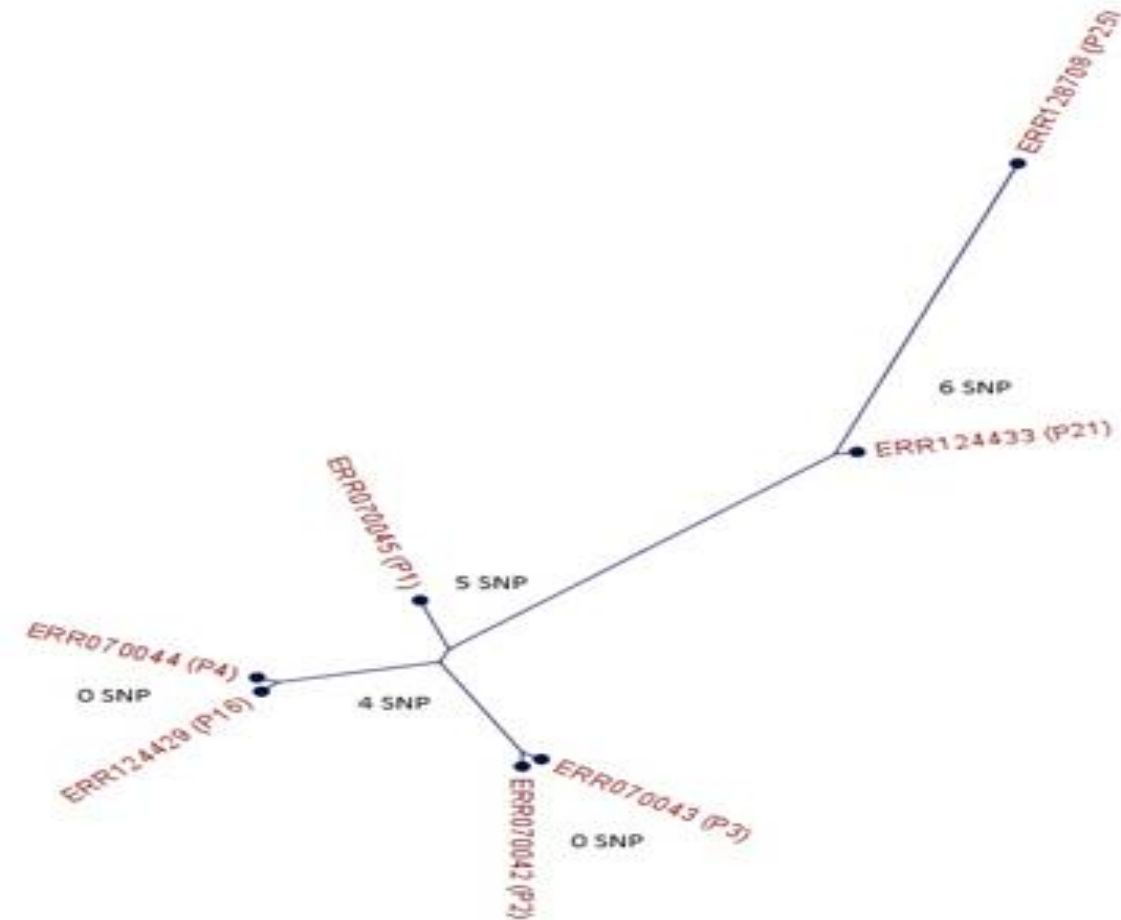
S.R. Harris, *et al.* 2013. Whole-genome sequencing for analysis of and outbreak of methicillin-resistant *Staphylococcus aureus*: a descriptive study.

Phylogenetic Tree Published in Harris, *et al.*:



Isolates used in the ISDH validation are indicated by the red boxes around the patient numbers.

Phylogenetic Tree Created in CLC Bio



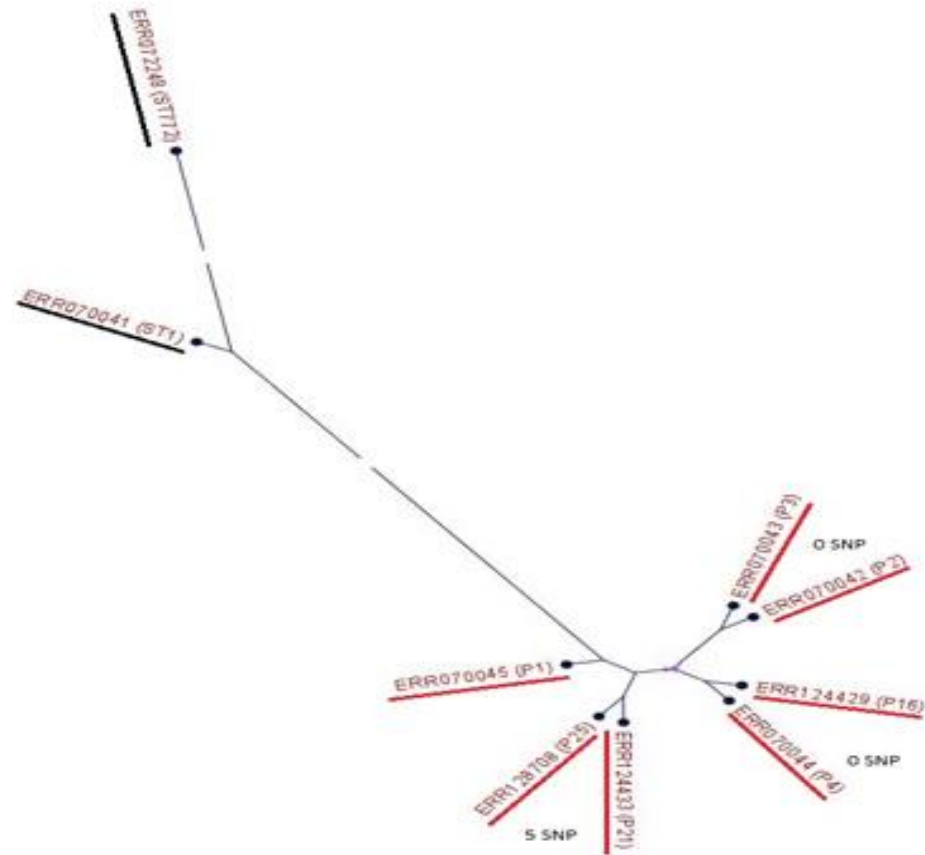
Clustering is reproducible with 100% agreement, although topology differs.

Note: Branches shorter than 0.0002 are shown as having length 0.0002

CLC Bio Tree with Unrelated Isolates

As suggested in Kozyreva, *et al.*, epidemiologically unrelated isolates were added to analysis.

Unrelated isolates did not cluster with outbreak isolates (comparison tree not published).



Note: Branches shorter than 0.0194 are shown as having length 0.0194

Analytical Sensitivity and Specificity

- Analytical sensitivity and specificity were calculated from the sequences used for hqSNP genotyping analysis.
- Overall sensitivity and specificity = 100%.



Indiana State
Department of Health

Analytical Sensitivity and Specificity

- True positive – Number of related sequences that cluster together (CDC sequences)
- True negative – Number of unrelated sequences that cluster together (ISDH sequences)
- False negative – Number of related sequences (CDC sequences) that cluster with unrelated sequences (ISDH sequences)
- False positive – Number of unrelated sequences (ISDH sequences) that cluster with related sequences (CDC sequences)



Indiana State
Department of Health

Precision



Indiana State
Department of Health

Platform Repeatability

- “Within-run precision”
- Sequenced the same isolate multiple times under the same conditions (*i.e.* same cartridge)
- Performed SNP analysis on all within-run replicates for each validation isolate
- Final average platform repeatability per base pair >99.9%



Indiana State
Department of Health

Platform Repeatability

Isolate	Length of Covered Genome (bp)	Total # of SNP Difference for Within-run Replicates	Repeatability
ARLN 0056	3,645,732	0	100%
ARLN 0076	4,789,070	0	100%
ARLN 0349	4,506,303	2	99.99995562%
ARLN 0347	5,022,339	0	100%
C17030866	4,731,694	0	100%
BAA 2146	5,127,545	0	100%
ATCC 10211	1,796,677	0	100%
ATCC 29213	2,692,122	0	100%
ATCC 25291	4,658,151	2	99.9999571%
18Myc0103	5,363,168	5	99.99990677%
Average repeatability per bp			99.99998194%

Platform Reproducibility

- “Between-run precision”
- Sequenced the same isolate under different conditions (*i.e.*, different cartridges, different library prep, different analyst)
- Performed SNP analysis on all replicates for 5 validation isolates
- Final average platform reproducibility per base pair >99.9%



Indiana State
Department of Health

Platform Reproducibility

Isolate	Length of Covered Genome (bp)	Total # of SNP Difference for Between-run Replicates	Reproducibility
ARLN 0056	3,645,732	0	100%
ARLN 0076	4,789,070	0	100%
ARLN 0349	4,506,303	3	99.99993343%
ARLN 0347	5,022,339	0	100%
BAA 2146	5,127,545	0	100%
Average reproducibility per bp			99.99998668%

Application Repeatability and Reproducibility

- Resistance typing, 16S identification, and MLST tools were analyzed between replicates of the same run (repeatability) and between runs (reproducibility).
- All analysis tools showed 100% repeatability and reproducibility.



Indiana State
Department of Health

Repeatability-Resistance Typing

ARLN/ATCC Number	ARLN/ATCC Resistance Mechanisms	Detected Resistance Mechanisms
ARLN 0347-1	KPC-3	KPC-3
ARLN 0347-2	KPC-3	KPC-3
ARLN 0347-3	KPC-3	KPC-3
ARLN 0076-1	VIM-1	VIM-1
ARLN 0076-2	VIM-1	VIM-1
ARLN 0076-3	VIM-1	VIM-1
ARLN 0349-1	mcr-1, CTXM-14, CTXM-55, TEM-1B	mcr-1, CTXM-14, CTXM-55, TEM-1B
ARLN 0349-2	mcr-1, CTXM-14, CTXM-55, TEM-1B	mcr-1, CTXM-14, CTXM-55, TEM-1B
ARLN 0349-3	mcr-1, CTXM-14, CTXM-55, TEM-1B	mcr-1, CTXM-14, CTXM-55, TEM-1B
ATCC BAA 2146-1	NDM-1	NDM-1
ATCC BAA 2146-2	NDM-1	NDM-1
ATCC BAA 2146-3	NDM-1	NDM-1
ARLN 0056-1	OXA-23, OXA-66	OXA-23, OXA-66
ARLN 0056-2	OXA-23, OXA-66	OXA-23, OXA-66
ARLN 0056-3	OXA-23, OXA-66	OXA-23, OXA-66

Repeatability-16S ID

Isolate Number	Gold Standard Identification:	16S Identification
ARLN 0347-1	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>
ARLN 0347-2	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>
ARLN 0347-3	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>
ARLN 0076-1	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>
ARLN 0076-2	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>
ARLN 0076-3	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>
ARLN 0349-1	<i>E. coli</i>	<i>E. coli</i>
ARLN 0349-2	<i>E. coli</i>	<i>E. coli</i>
ARLN 0349-3	<i>E. coli</i>	<i>E. coli</i>
ATCC BAA 2146-1	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>
ATCC BAA 2146-2	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>
ATCC BAA 2146-3	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>
ARLN 0056-1	<i>Acinetobacter baumannii</i>	<i>Acinetobacter baumannii</i>
ARLN 0056-2	<i>Acinetobacter baumannii</i>	<i>Acinetobacter baumannii</i>
ARLN 0056-3	<i>Acinetobacter baumannii</i>	<i>Acinetobacter baumannii</i>
ATCC 10211-1	<i>Haemophilus influenzae</i>	<i>Haemophilus influenzae</i>
ATCC 10211-2	<i>Haemophilus influenzae</i>	<i>Haemophilus influenzae</i>
ATCC 10211-3	<i>Haemophilus influenzae</i>	<i>Haemophilus influenzae</i>
ATCC 29213-1	<i>Staphylococcus aureus</i>	<i>Staphylococcus aureus</i>
ATCC 29213-2	<i>Staphylococcus aureus</i>	<i>Staphylococcus aureus</i>
ATCC 29213-3	<i>Staphylococcus aureus</i>	<i>Staphylococcus aureus</i>
18Myc0103-1	<i>Nocardia farcinica</i>	<i>Nocardia farcinica</i>
18Myc0103-2	<i>Nocardia farcinica</i>	<i>Nocardia farcinica</i>
18Myc0103-3	<i>Nocardia farcinica</i>	<i>Nocardia farcinica</i>
ATCC 25291-1	<i>Mycobacterium avium</i>	<i>Mycobacterium avium</i>
ATCC 25291-2	<i>Mycobacterium avium</i>	<i>Mycobacterium avium</i>
ATCC 25291-3	<i>Mycobacterium avium</i>	<i>Mycobacterium avium</i>
C17090866-1	<i>Salmonella enterica</i>	<i>Salmonella enterica</i>
C17090866-2	<i>Salmonella enterica</i>	<i>Salmonella enterica</i>
C17090866-3	<i>Salmonella enterica</i>	<i>Salmonella enterica</i>

Repeatability-MLST

Isolate Number	MLST Scheme
ARLN 0347-1	ST 11
ARLN 0347-2	ST 11
ARLN 0347-3	ST 11
ARLN 0076-1	ST 86
ARLN 0076-2	ST 86
ARLN 0076-3	ST 86
ARLN 0349-1	ST 354
ARLN 0349-2	ST 354
ARLN 0349-3	ST 354
ATCC BAA 2146-1	ST 11
ATCC BAA 2146-2	ST 11
ATCC BAA 2146-3	ST 11
ARLN 0056-1	ST 423
ARLN 0056-2	ST 423
ARLN 0056-3	ST 423
ATCC 10211-1	ST 6
ATCC 10211-2	ST 6
ATCC 10211-3	ST 6
ATCC 29213-1	ST 4618
ATCC 29213-2	ST 4618
ATCC 29213-3	ST 4618
C17090866-1	ST 19
C17090866-2	ST 19
C17090866-3	ST 19

Reproducibility-Resistance Typing

ARLN/ATCC Number	Operator 1 Detected Mechanisms	Operator 2 Detected Mechanisms
ARLN 0347	KPC-3	KPC-3
ARLN 0076	VIM-1	VIM-1
ARLN 0349	mcr-1, CTXM-14, CTXM-55, TEM-1B	mcr-1, CTXM-14, CTXM-55, TEM-1B
ATCC BAA 2146	NDM-1	NDM-1
ARLN 0056	OXA-23, OXA-66	OXA-23, OXA-66

Reproducibility-16S Identification

ARLN/ATCC Number	Operator 1 16S Identification	Operator 2 16S Identification
ARLN 0347	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>
ARLN 0076	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>
ARLN 0349	<i>E. coli</i>	<i>E. coli</i>
ATCC BAA 2146	<i>Klebsiella pneumoniae</i>	<i>Klebsiella pneumoniae</i>
ARLN 0056	<i>Acinetobacter baumannii</i>	<i>Acinetobacter baumannii</i>

Reproducibility-MLST

ARLN/ATCC Number	Operator 1 MLST	Operator 2 MLST
ARLN 0347	ST 11	ST 11
ARLN 0076	ST 86	ST 86
ARLN 0349	ST 354	ST 354
ATCC BAA 2146	ST 11	ST 11
ARLN 0056	ST 423	ST 423

Questions?



Indiana State
Department of Health

Acknowledgements

- Bioinformatics assistance: Kelsey Florek, Wisconsin State Laboratory of Hygiene
- KPC outbreak sequences: Tom de Man, CDC



Indiana State
Department of Health