CLIA-compliant Validation of Whole Genome Sequencing (WGS) for Clinical Microbiological Applications:

Experience of one State Public Health Reference Laboratory

Varvara Kozyreva, PhD
Genotyping Unit Chief
Microbial Diseases Laboratory Program
California Department of Public Health
✉ varvara.kozyreva@cdph.ca.gov

ASM NGS 2018
September 24th 2018

Disclosure

I have no conflict of interest relevant to this presentation.

Clinical Laboratory Improvement Amendments (CLIA)- federal regulatory standards ensuring quality of clinical laboratory testing performed on samples from humans for the diagnostic purposes.

CLIA is implemented and enforced by the Centers for Medicare & Medicaid Services (CMS).

WGS Workload in the Microbial Diseases Laboratory of California Department of Public Health

CDPH California Department of PublicHealth

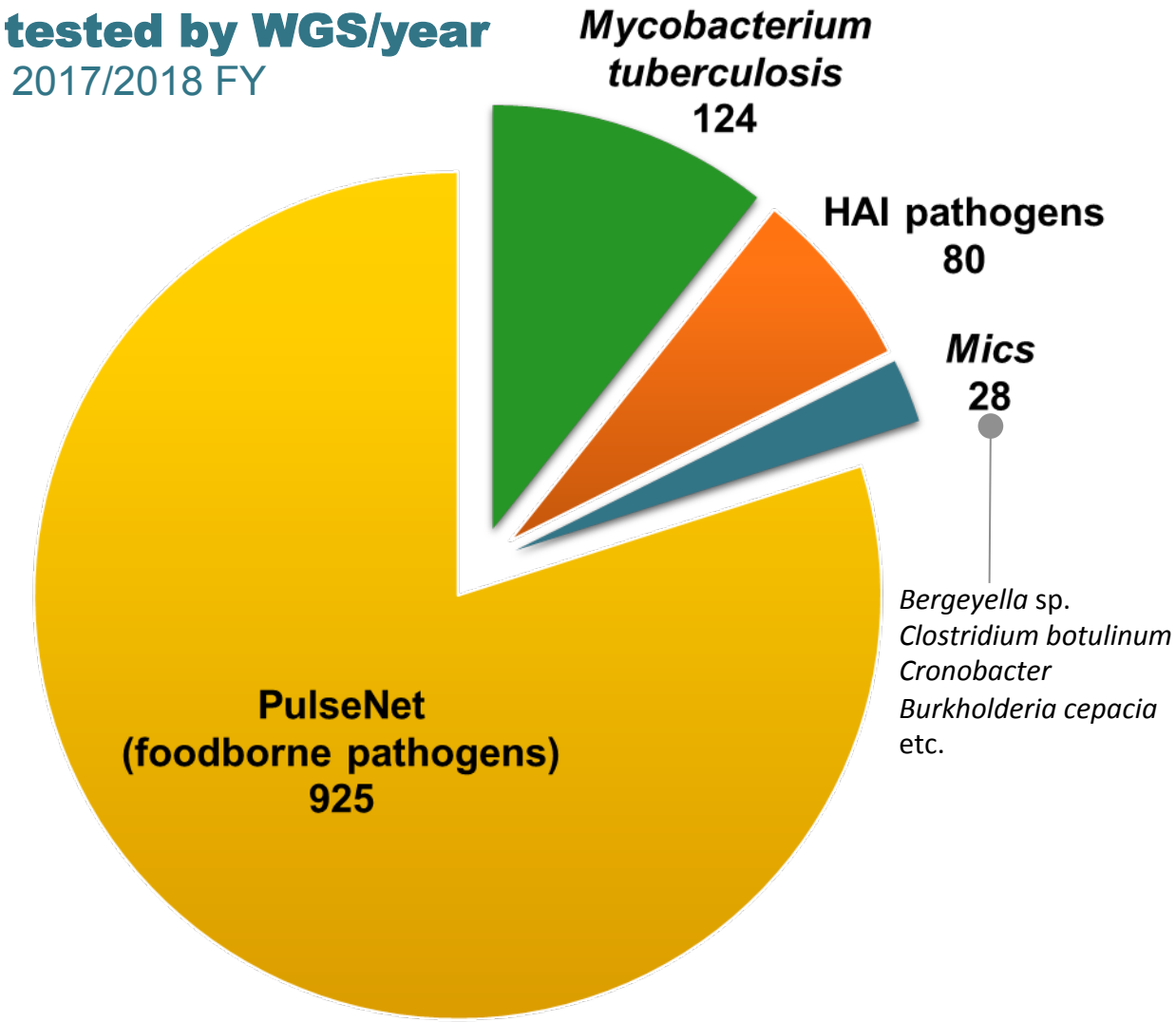
MDL Microbial Diseases Laboratory

Pathogen Experts Keeping California Safe

CMS CLIA Certified

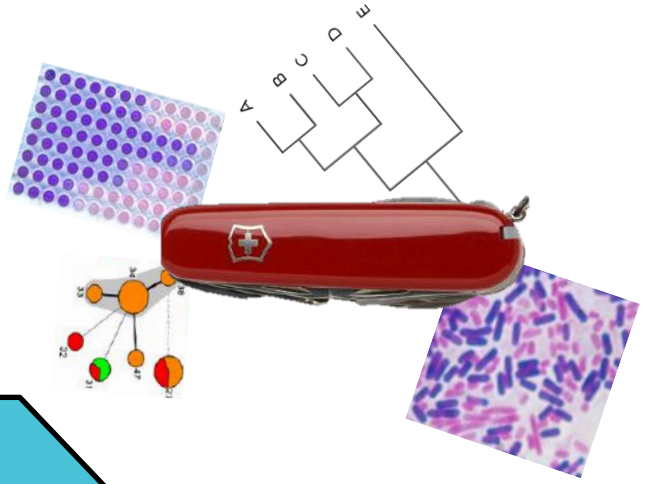
Isolates tested by WGS/year

2017/2018 FY



Data provided by Dr. Rituparna Mukhopadhyay

Simple public health lab WGS toolkit



In-house sequencing:
MiSeq sequencer
Nextera XT library prep
500- or -600-cycle seq kits

BIOINFORMATIC ANALYSIS PIPELINE

Raw reads import
Trimming

***De novo* assembly**

- **16S rRNA ID**
- **ABR genes detection**
- ***In silico* MLST**

Mapping to reference

- **hqSNP calling**
- **Phylogenetic tree generation**

Report to
state epidemiologists
& outside clients

… but dangerous if used incorrectly

Image sources:
Basmaci et al, JCM 2014
De Agostini Picture Library / Getty Images
Agrawal et al, PlosOne 2015
https://www.khanacademy.org/
https://www.swissinfo.ch/
Illumina, Inc.
http://www.dorkly.com/

WGS → • a laboratory-developed test (LDT) • not approved/cleared by FDA → **WGS validation in microbiological PHL settings**

Challenges:

- **Complexity** of technology with extensive computational analysis requirements
- Insufficient guidelines for validation of WGS for **microbiological applications**
  - Publications addressing NGS standardization and validation are mainly focused on cancer diagnostics, inherited human diseases, and pathology

- **CDC Nex-StoCT Workgroups I & II:** two publications on standardization of clinical NGS- *validation, QC, ref materials, PT* (Gargis et al. Nat Biotechnol 2012 & 2015)

- **College of American Pathologists (CAP):** *validation, QC, QA, ref materials* (CAP NGS Inspection Checklist, 2014; Aziz et al. Arch Pathol Lab Med, 2015; Jennings et al. J Mol Diag, 2017; Roy et al. J Mol Diag, 2018)

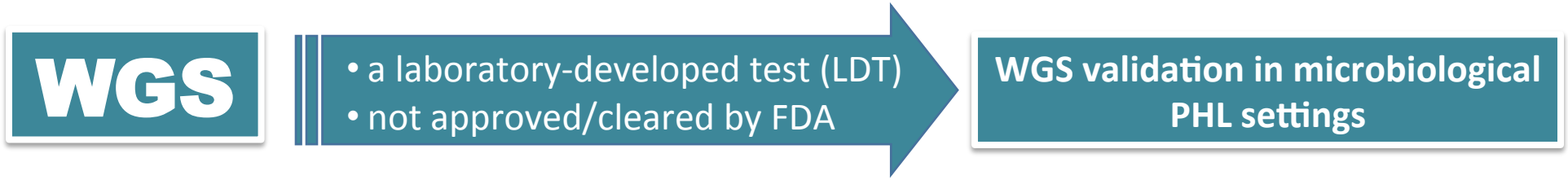
- **Clinical and Laboratory Standards Institute (CLSI):** MM09 - Nucleic Acid Sequencing Methods in Diagnostic Laboratory Medicine (2014) – *validation, QC, QA, PT, competency, ref materials*

- **American College for Medical Genetics and Genomics (ACMG):** Clinical laboratory standards for NGS- *Validation, QC, QA, ref material, PT.* (Rehm et al. Genet Med. 2013)

- Validation of NGS in **Microbial Forensics** (Budowle et al. Investigative Genetics, 2014)

- **Food and Drug Administration (FDA):** Draft Guidance for Infectious Disease NGS-Based Diagnostic Devices: ID, antimicrobial resistance, virulence (2016)- *for Industry and FDA staff*

- Validation of **Metagenomic NGS** for pathogen detection (Schlaberg et al. Arch Pathol Lab Med, 2017)



Challenges:

- **Complexity** of technology with extensive computational analysis requirements
- Insufficient **guidelines** for validation of WGS for **microbiological applications**
  - Publications addressing NGS standardization and validation are mainly focused on cancer diagnostics, inherited human diseases, and pathology
- **No examples** of the validation of WGS-based tests used in PHL[1]
- Poorly established **performance parameters**
- Lack of concept for CLIA **validation of epidemiological tests**
- Lack of established **reference materials**
  - ✧ Well-characterized
  - ✧ Relevant to public health
  - ✧ Diverse

Possible Reference Materials sources:

- **CDC PulseNet**- foodborne (PulseNet) organisms only (Isolates)
- **NIST**- National Institute of Standards and Technology – *S. enterica, S. aureus, P. aeruginosa, Clostridium sporogenes*[2] (DNA, not isolates)[2]
- **FDA-ARGOS**- Database for Reference Grade Microbial Sequences (*in silico* validation)[3]

1 Examples of Validation of Metagenomic NGS for pathogen detection (Schlaberg et al. Arch Pathol Lab Med, 2017)
2 https://www.nist.gov/programs-projects/microbial-genomic-measurements
3 https://www.ncbi.nlm.nih.gov/bioproject/231221

Main objectives of WGS validation in microbiological PHL:

- Establish the **performance specifications** of WGS used for public health applications
- Create a **validation set of microorganisms** which can be used for future validations of WGS platforms
- Determine the **optimal conditions** that will generate reliable, reproducible, and accurate results for the intended application
- Develop a **set of Quality Assurance (QA) & Quality Control (QC) measures** to ensure high quality and consistency of routine testing
- Establish **reporting language**

VALIDATION SET

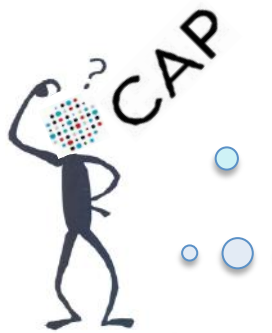
10- *Enterobacteriaceae*
5- Gram-positive cocci isolates
5- Gram-negative non-fermenting bacterial isolates
9- *Mycobacterium tuberculosis*
5- representatives of miscellaneous species

Validated a test panel of bacterial isolates for WGS

- **Source:**
  - ATCC strains
  - Strains sequenced by the CDC
- **34 bacterial isolates**
- **19 species**
- **Genome sizes** 1.8 to 6.7 Mb
- Wide range of **GC content** 32.1-66.1%
- Can be used as **reference materials for future validation** of new sequencing platforms and tests, QC procedures, proficiency testing, and the independent assessment of test performance

Acinetobacter baumannii
Aeromonas hydrophilia
Bacteroides fragilis
Corynebacterium jeikeium
Enterobacter cloacae
Enterococcus faecalis
Escherichia coli
Haemophilus influenzae
Legionella pneumophila
Moraxella catarrhalis
Mycobacterium tuberculosis
Neisseria gonorrhoeae
Pseudomonas aeruginosa
Salmonella enterica ser Adelaide
Salmonella enterica ser Enteritidis
Salmonella enterica ser Infantis
Salmonella enterica ser Typhimurium
Salmonella enterica ser Worthington
Staphylococcus aureus
Staphylococcus epidermidis
Staphylococcus saprophyticus
Stenotrophomonas maltophilia
Streptococcus pneumoniae

- The **platform validation is not sufficient** to demonstrate that sequencing assay is able to detect the [disease-associated] sequence variation.
- A validation tailored **specifically to the application (**the specific implementation of the **bioinformatics pipeline)** should be undertaken. [1]



- NGS tests are better approached through **method-based validation** than analyte-specific validation, if it is **impractical/impossible to obtain positive samples** of all targeted mutations.
- The **ability of a laboratory to validate**, perform, and interpret certain **types of genomic sequencing analysis** is not wholly dependent on the specific variants being tested, but rather on **ability of the lab to conduct that type of sequencing analysis in its totality.** [2]

It is **not** possible to validate all theoretically possible variants that can occur, therefore it is necessary to use a **combination** of a '**methods-based**' and '**analyte-specific**' validation approach for determining a test's analytic performance. [3]

wet & dry

beginning-to-end

1. CLSI. MM09-A2, 2nd ed., 2014 CLSI, Wayne, PA
2. Schrijver I et al., J Mol Diagn. 2014 May;16(3):283-7.
3. Aziz N et al., Arch Pathol Lab Med. 2015 Apr;139(4):481-93.

10- *Enterobacteriaceae*
5- Gram-positive cocci isolates
5- Gram-negative non-fermenting bacterial isolates
9- *Mycobacterium tuberculosis*
5- representatives of miscellaneous species

Validation Set
34 bacterial isolates

WHOLE GENOME SEQUENCING VALIDATION IN PUBLIC HEALTH MICROBIOLOGY LAB SETTINGS

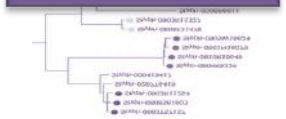
Accuracy

- MiSeq Illumina platform accuracy

- Test accuracy

  - Genotyping
    - MLST
      - 16S rRNA gene ID
        - Antibiotic resistance genes detection

- Bioinformatics pipeline accuracy



Reproducibility/ Repeatability

triplicates

- Genotyping: per replicate per base pair
  - MLST
    - 16S rRNA gene ID

Analytical sensitivity and specificity

- False-positive & False-negative rates
  - Genotyping
    - MLST
- Limit of Detection & Interference/ Cross-reactivity
  - Coverage limit of SNP detection
    - Specificity of SNP detection in the presence of interfering sequences

Base calls vs. WGS assays

10- *Enterobacteriaceae*
5- Gram-positive cocci isolates
5- Gram-negative non-fermenting bacterial isolates
9- *Mycobacterium tuberculosis*
5- representatives of miscellaneous species

Validation Set
34 bacterial isolates

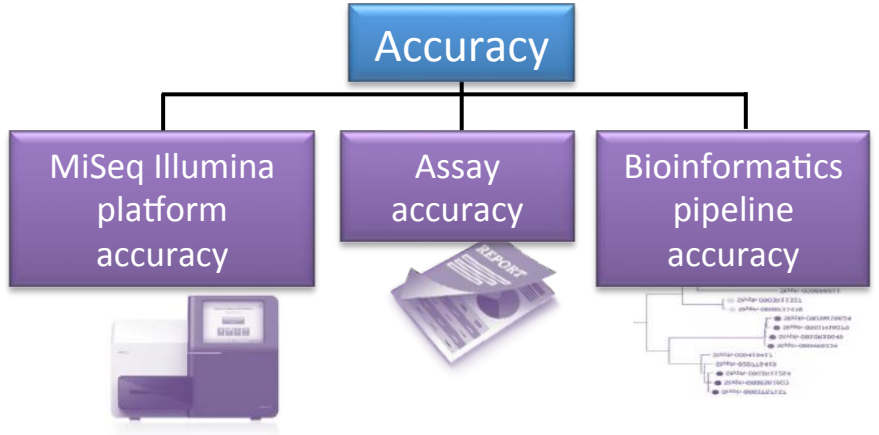
WHOLE GENOME SEQUENCING VALIDATION IN PUBLIC HEALTH MICROBIOLOGY LAB SETTINGS

Accuracy
- MiSeq Illumina platform accuracy
- Test accuracy
- Bioinformatics pipeline accuracy

Reproducibility/ Repeatability

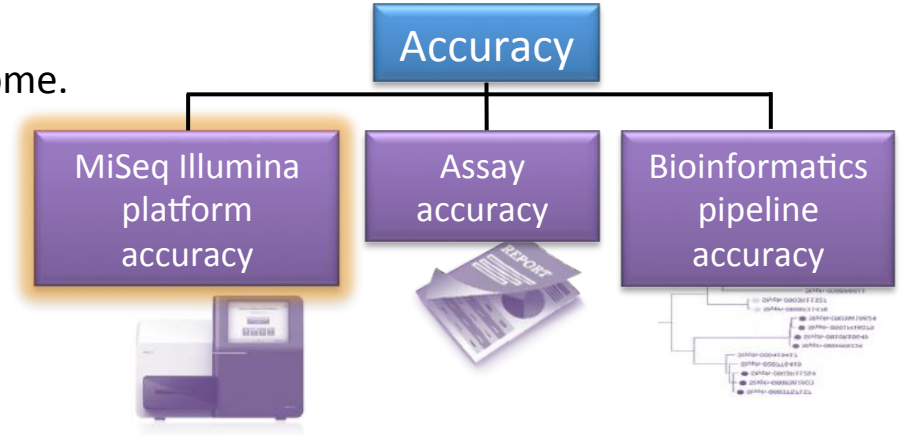
Analytical sensitivity and specificity
- False-positive & False-negative rates
- Limit of Detection & Interference/ Cross-reactivity

	Accuracy: MiSeq Illumina platform accuracy	Accuracy: Test accuracy	Accuracy: Bioinformatics pipeline accuracy	Reproducibility/ Repeatability	Analytical sensitivity and specificity: False-positive & False-negative rates	Analytical sensitivity and specificity: Limit of Detection & Interference/ Cross-reactivity
Correct base calls (SNPs)	✔			✔		✔
Genotyping (topology)		✔	✔		✔	
MLST		✔		✔	✔	
16S rRNA gene ID		✔		✔		
Antibiotic resistance genes detection		✔	✔			


Accuracy
MiSeq Illumina platform accuracy
Assay accuracy
Bioinformatics pipeline accuracy
REPORT

Accuracy of the platform -

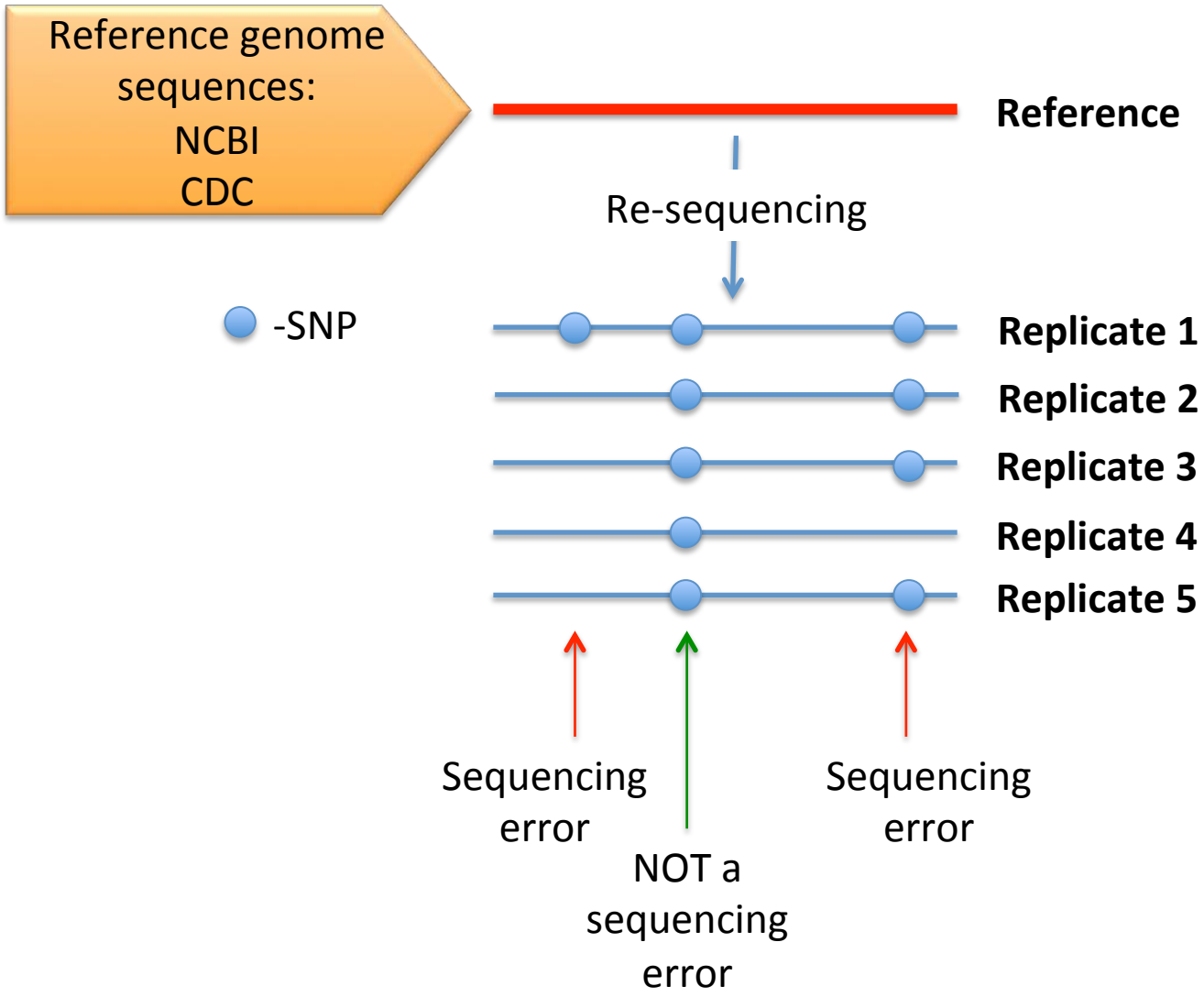
accuracy of the identification of individual base pairs (“base calling”) in the bacterial genome.



Determined by:

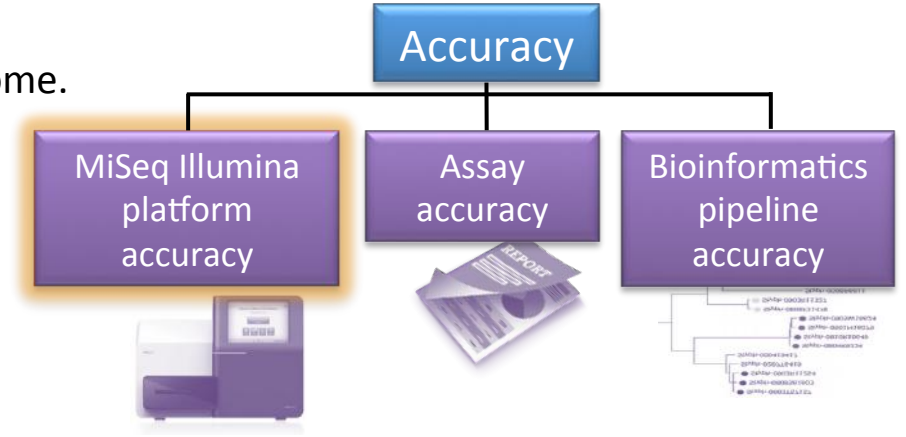
the proximity of agreement between base calling made by MiSeq sequencer (measured value) and NCBI/CDC reference complete genome sequence (the true value)

ID	Species	Total # of SNP difference with the reference
C1	*Escherichia coli*	**5**
C2	*Aeromonas hydrophilia*	**1**
C3	*Escherichia coli*	**22**
C4	*Enterobacter cloacae*	**10**
C5	*Staphylococcus aureus*	**0**
C6	*Salmonella* ser Typhimurium	**12**
C46	*Enterococcus faecalis*	**3**
C47	*Staphylococcus epidermidis*	**184**
C48	*Staphylococcus saprophyticus*	**27**
C51	*Stenotrophomonas maltophilia*	**39**
C52	*Legionella pneumophila*	**2**
C55	*Escherichia coli*	**14**
C72	*Escherichia coli* O121:H19	**0**
C73	*Salmonella ser* Enteritidis	**0**
C74	*Salmonella ser* Infantis	**3**
C75	*Salmonella ser* Adelaide	**0**
C76	*Salmonella ser* Worthington	**1**
C105	*Corynebacterium jeikeium*	**4**


Reference genome sequences:
NCBI
CDC
Reference
Re-sequencing
-SNP
Replicate 1
Replicate 2
Replicate 3
Replicate 4
Replicate 5
Sequencing error
NOT a sequencing error
Sequencing error

Accuracy of the platform -

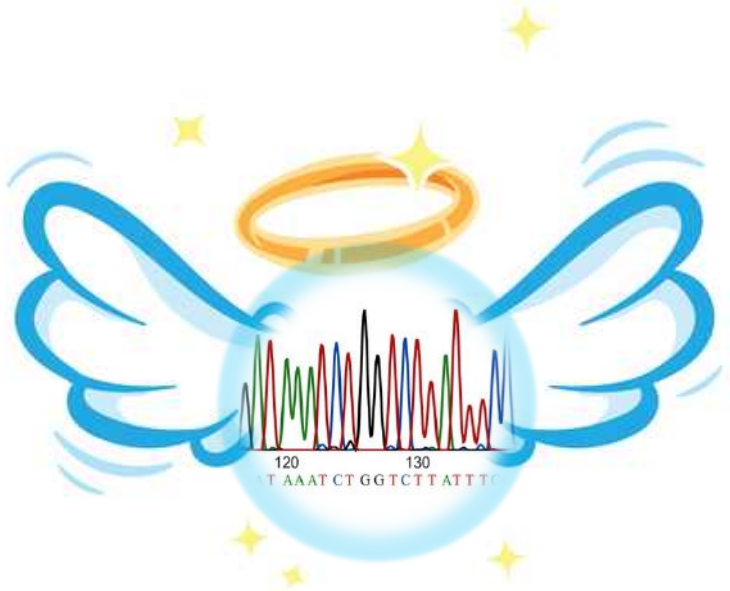
accuracy of the identification of individual base pairs (“base calling”) in the bacterial genome.



ID	Species	Total # of SNP difference with the reference	# of sequencing errors (SNP is supported only by 4 or less validation replicates)
C1	*Escherichia coli*	5	**0**
C2	*Aeromonas hydrophilia*	1	**0**
C3	*Escherichia coli*	22	**0**
C4	*Enterobacter cloacae*	10	**0**
C5	*Staphylococcus aureus*	0	**0**
C6	*Salmonella* ser Typhimurium	12	**0**
C46	*Enterococcus faecalis*	3	**0**
C47	*Staphylococcus epidermidis*	184	**2**
C48	*Staphylococcus saprophyticus*	27	**0**
C51	*Stenotrophomonas maltophilia*	39	**0**
C52	*Legionella pneumophila*	2	**0**
C55	*Escherichia coli*	14	**1**
C72	*Escherichia coli* O121:H19	0	**0**
C73	*Salmonella ser* Enteritidis	0	**0**
C74	*Salmonella ser* Infantis	3	**0**
C75	*Salmonella ser* Adelaide	0	**0**
C76	*Salmonella ser* Worthington	1	**0**
C105	*Corynebacterium jeikeium*	4	**0**

Searching for the reference…

Is Sanger sequencing so infallible?



Clinical Chemistry 62:4
647–654 (2016)

Molecular Diagnostics and Genetics

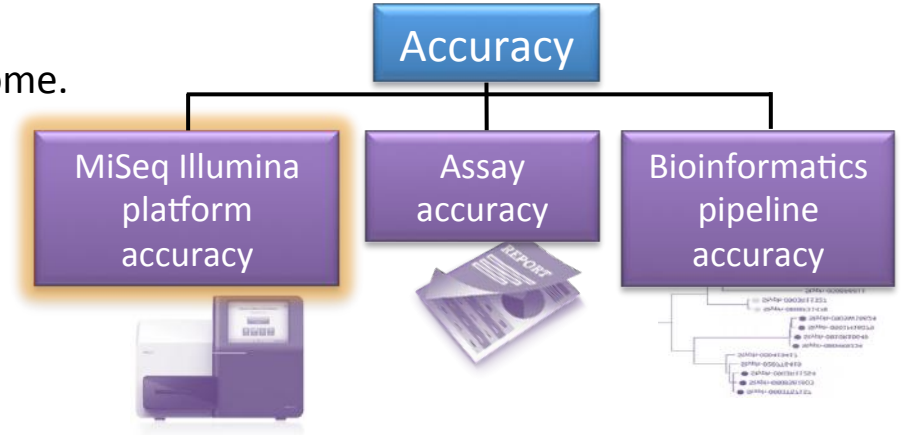
Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants

Tyler F. Beck,[1] James C. Mullikin on behalf of the NISC Comparative Sequencing Program,[1,2] and Leslie G. Biesecker[1*]

Beck TF et al., Clin Chem. 2016 Apr;62(4):647-54.
Image: https://www.123rf.com, cidegen.com

Accuracy of the platform -

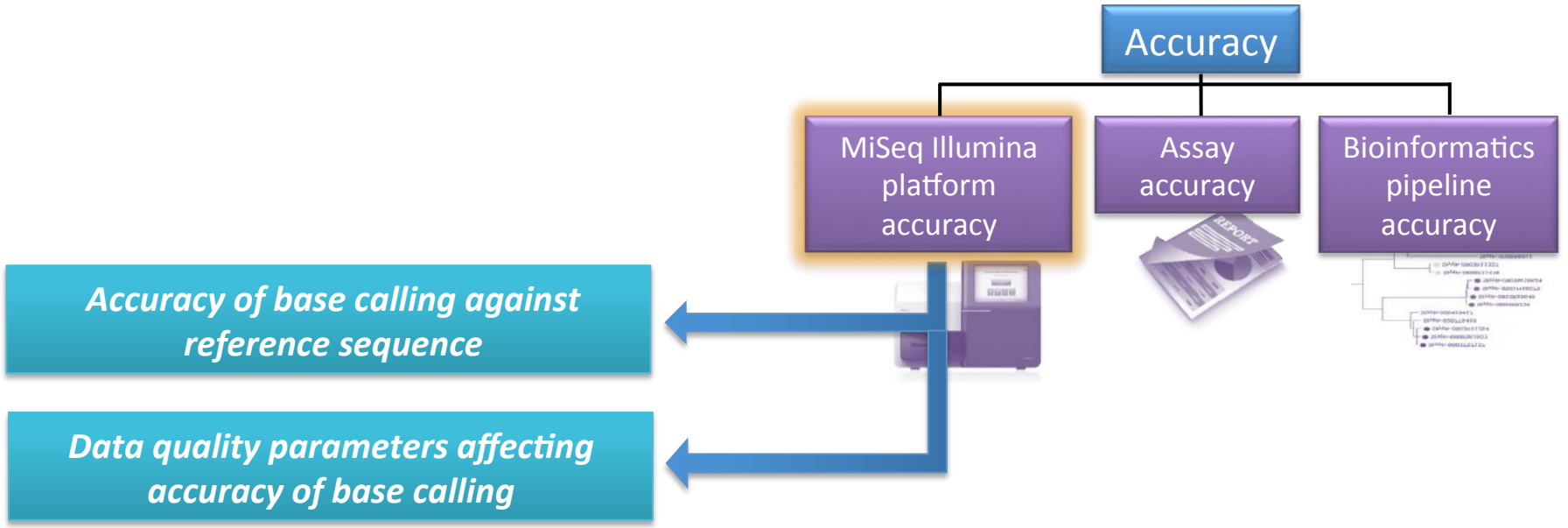
accuracy of the identification of individual base pairs (“base calling”) in the bacterial genome.



ID	Species	Total # of SNP difference with the reference	# of sequencing errors (SNP is supported only by 4 or less validation replicates)
C1	*Escherichia coli*	5	**0**
C2	*Aeromonas hydrophilia*	1	**0**
C3	*Escherichia coli*	22	**0**
C74	*Salmonella ser* Infantis	3	**0**
C75	*Salmonella ser* Adelaide	0	**0**
C76	*Salmonella ser* Worthington	1	**0**
C105	*Corynebacterium jeikeium*	4	**0**

$$\%\ \text{agreement with reference*} = \frac{(\text{Covered genome length}) - (\text{Total \# of SNP differing from reference})}{\text{Covered genome length}} \times 100\%$$

*Each nucleotide call - an independent test.



Sequencing Run Q

- % of bases with quality score ≥ Q30 for the run (%Q30)
- PhiX error rate
- Cluster density
- Cluster passing filter

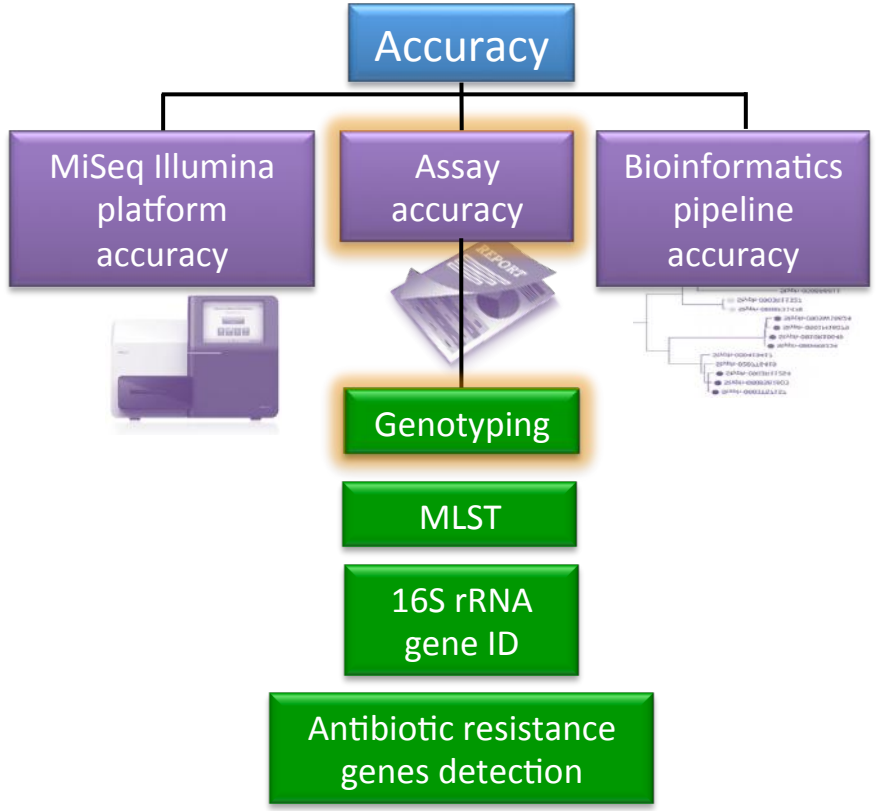
Raw Data Q

- Average depth of coverage
- The average read length with ≥ Q30
- Minimum read length of the fragments which have ≥75% of bases with Q30 score

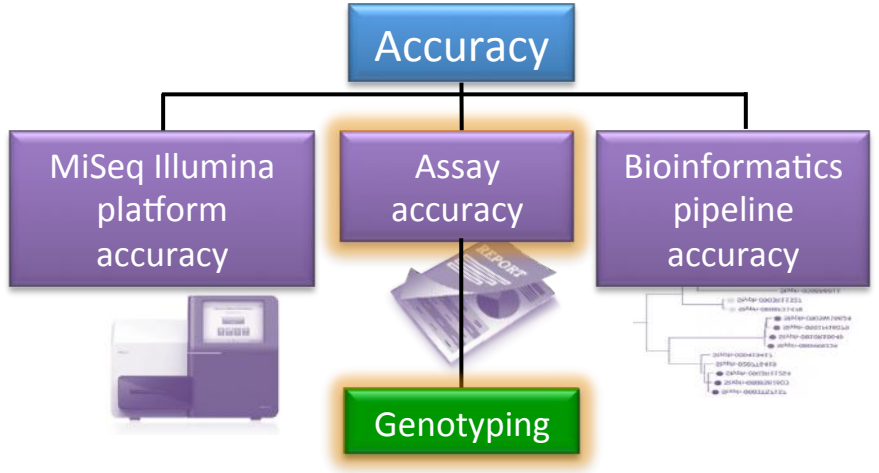
Secondary/Tertiary Data Analysis Q

- % genome covered after the mapping to reference
- Uniformity of coverage
- N50 for *de novo* assembled reads
- Number of assembled contigs

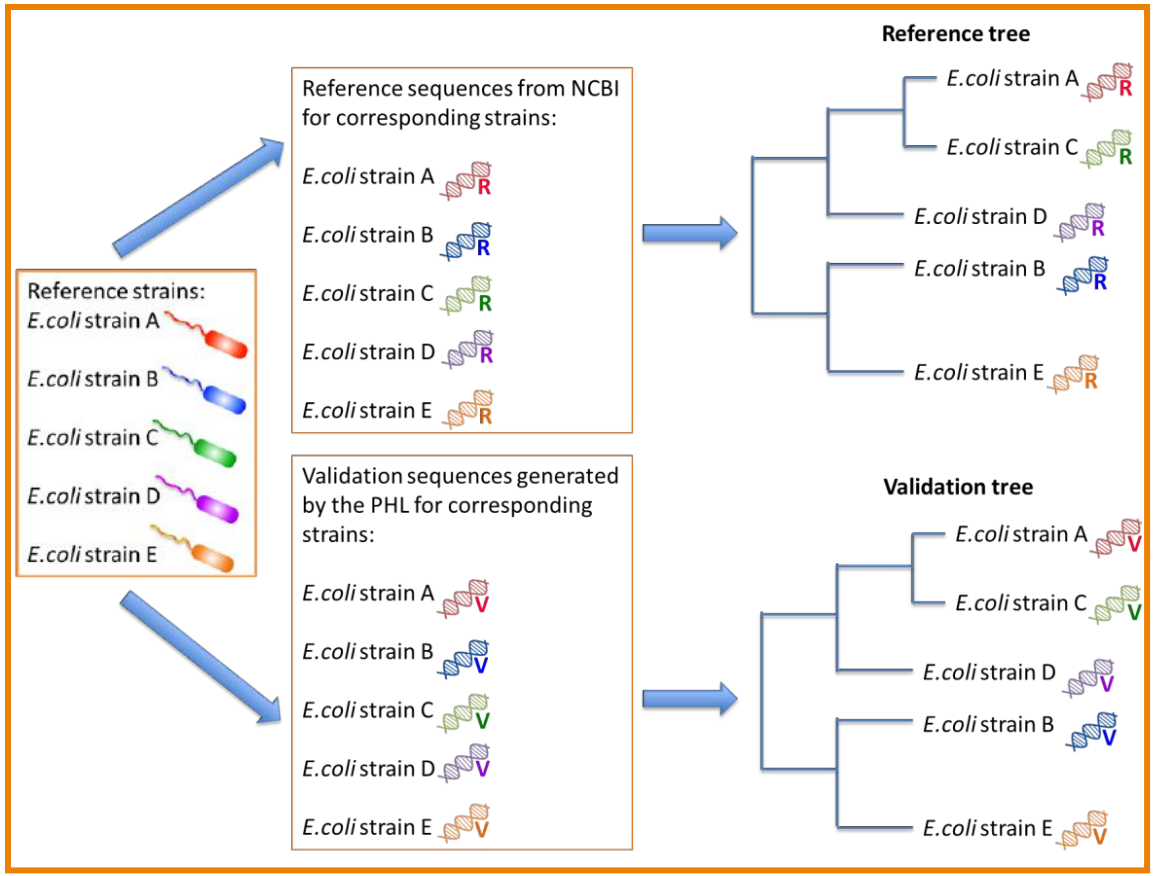
Assay accuracy - an agreement of the assay results for the validation sequences with the assay results for the reference sequences of the same strains.



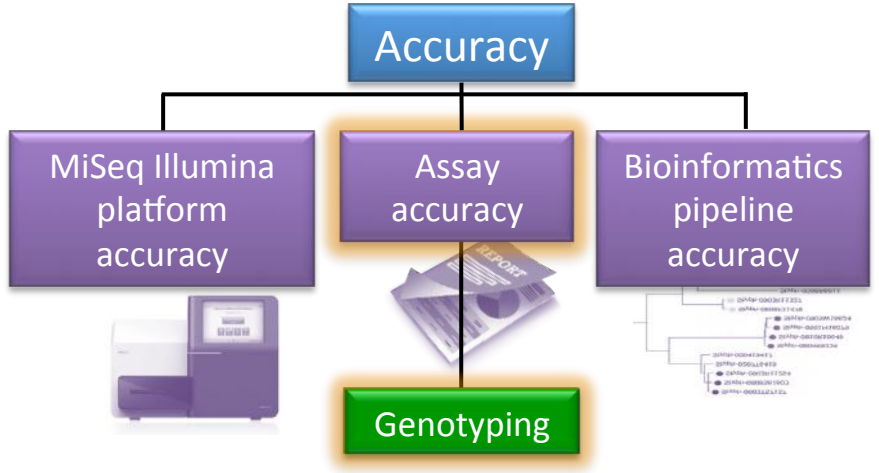
Assay accuracy - an agreement of the assay results for the validation sequences with the assay results for the reference sequences of the same strains.



Accuracy of genotyping assay: congruence of phylogenetic trees built using reference sequences and validation sequences



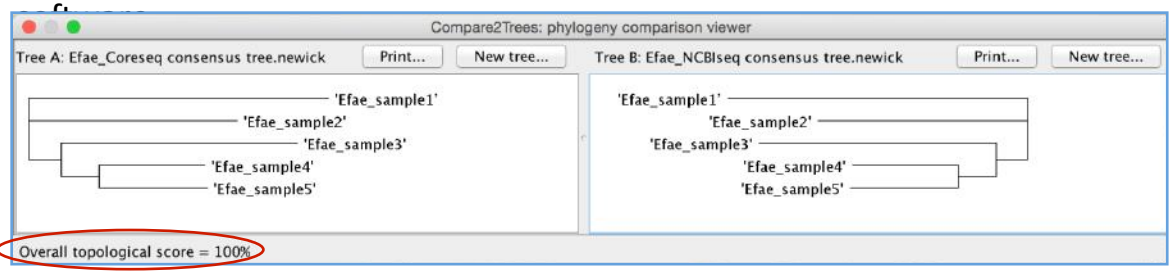
Assay accuracy - an agreement of the assay results for the validation sequences with the assay results for the reference sequences of the same strains.



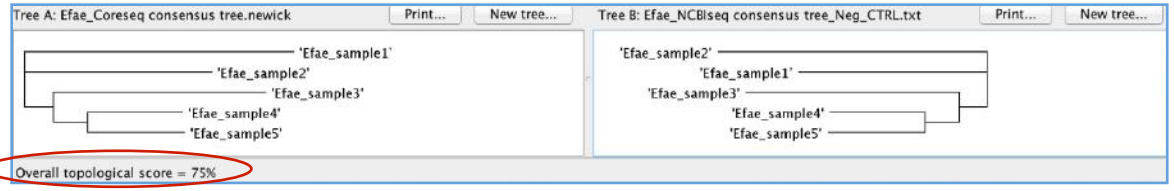
Accuracy of genotyping assay- congruence of phylogenetic trees built using reference sequences and validation sequences

Topological similarity between reference tree and validation tree

Percentage of topological similarity measured by Compare2Trees*

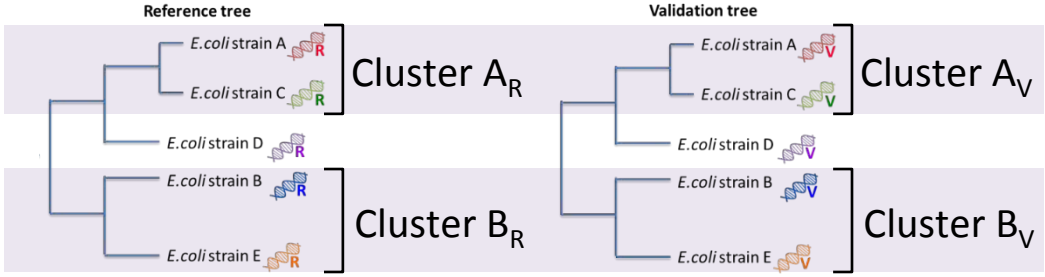


Overall topological score = **100%**



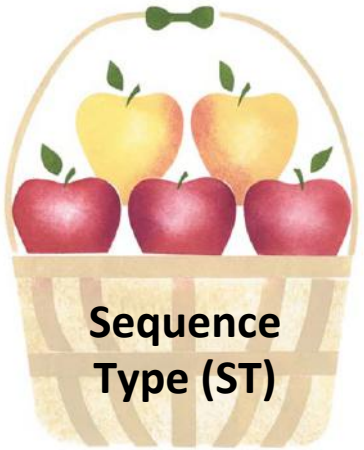
Overall topological score = **75%**

Comparison of clustering pattern of validation tree and reference tree

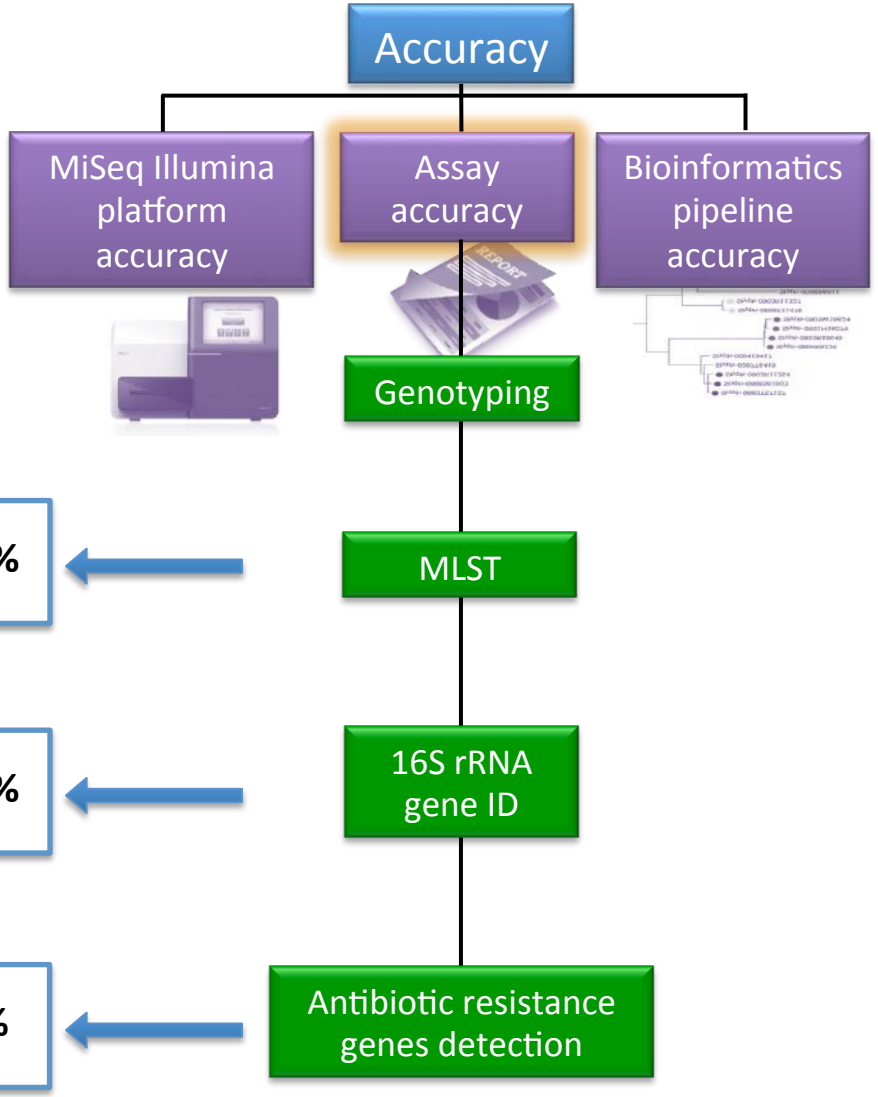


*Nye TM et al. Bioinformatics. 2006 Jan 1;22(1):117-9.

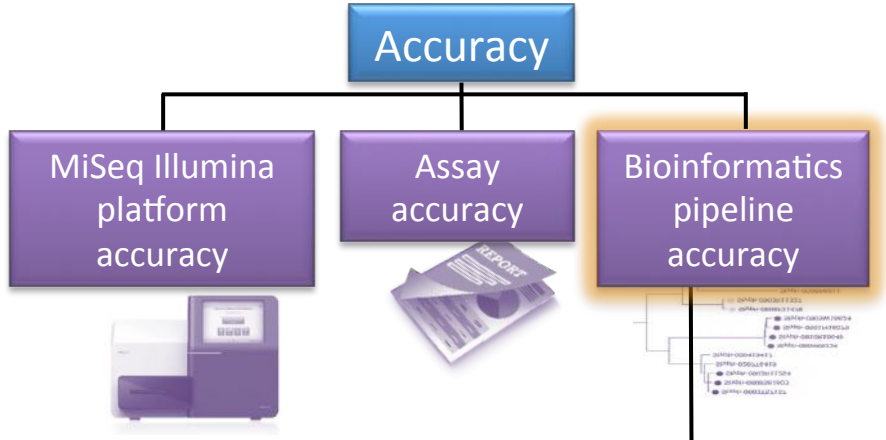
Assay accuracy - an agreement of the assay result for validation sequences generated by the lab with the assay result for reference sequences of the same strains.

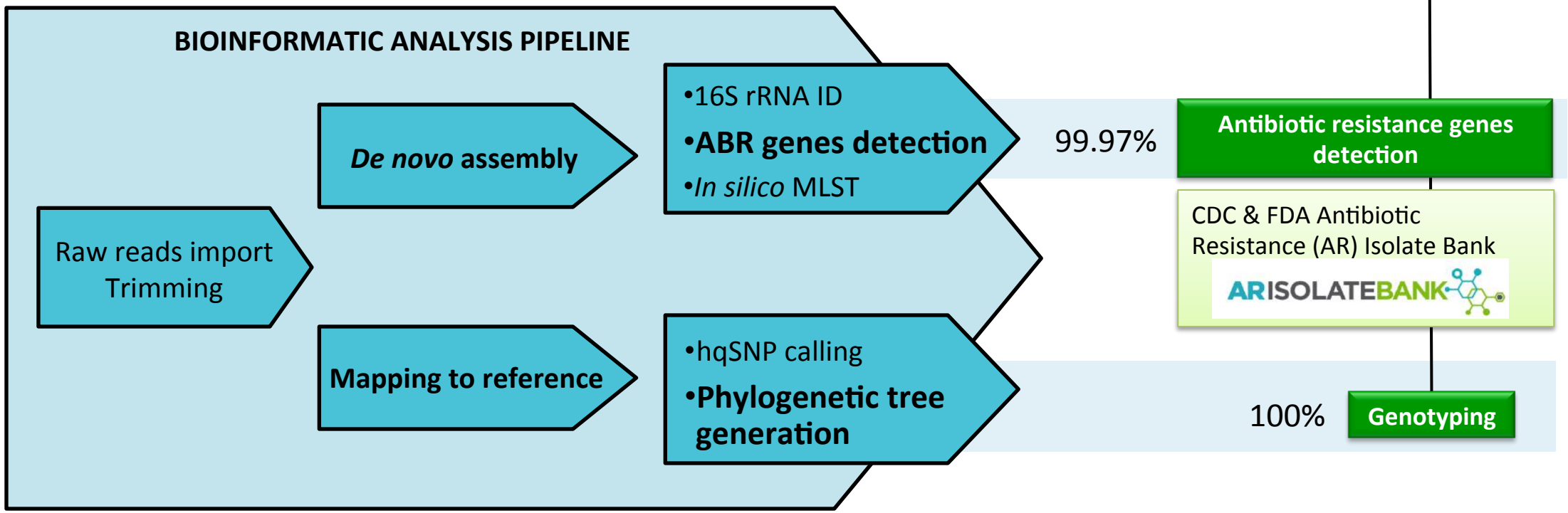


vs.



Bioinformatics pipeline accuracy





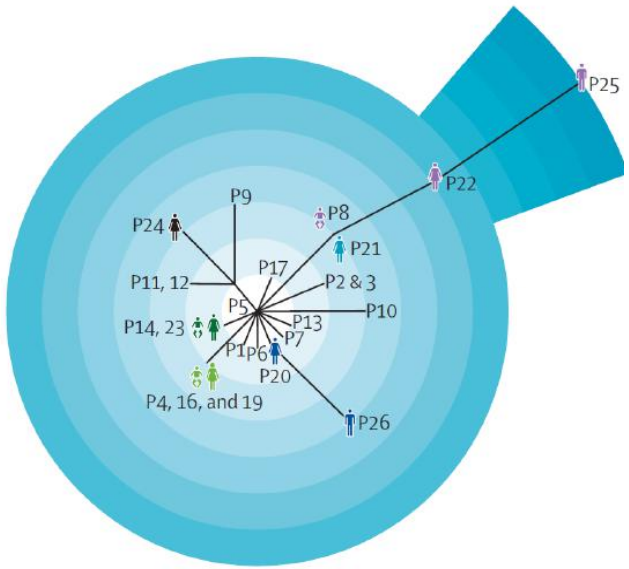
Phylogenetic bioinformatics pipeline accuracy:

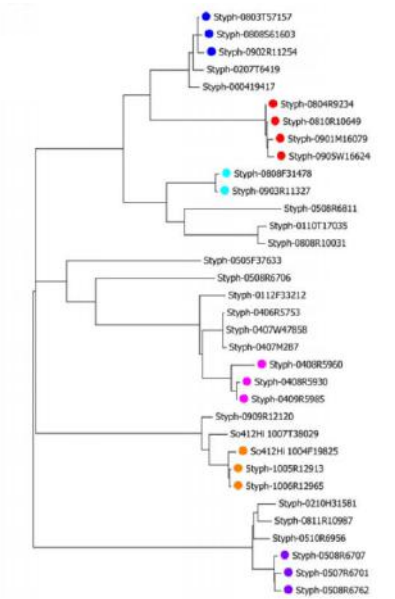
Raw WGS reads of bacterial isolates from outbreaks, previously characterized by WGS and epi data.

Phylogenetic analysis on the pipeline under validation

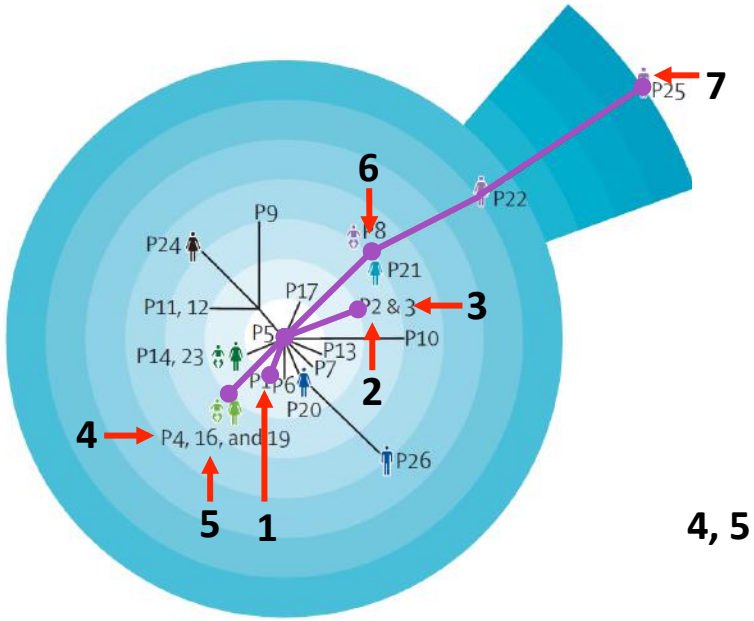
Accuracy criteria: previously suggested clustering is replicated

Study	Study 1. SR Harris et al. PMID: 23158674	Study 2. P Leekitcharoenphon et al. PMID: 24505344
Microorganism	Methicillin-resistant *Staphylococcus aureus*	*Salmonella enterica* serovar Typhimurium
Source of isolates	Human	Human
Number of isolates analyzed	7 outbreak isolates (1 outbreak cluster) + 2 epi unrelated isolates	9 outbreak isolates (4 outbreak clusters) + 2 epi unrelated isolates
Type of outbreak	Hospital-associated outbreak	Foodborne outbreaks

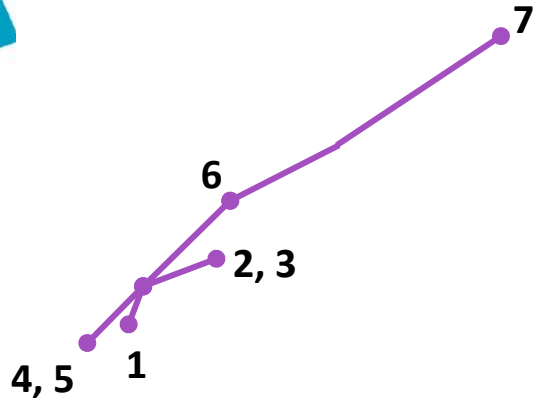




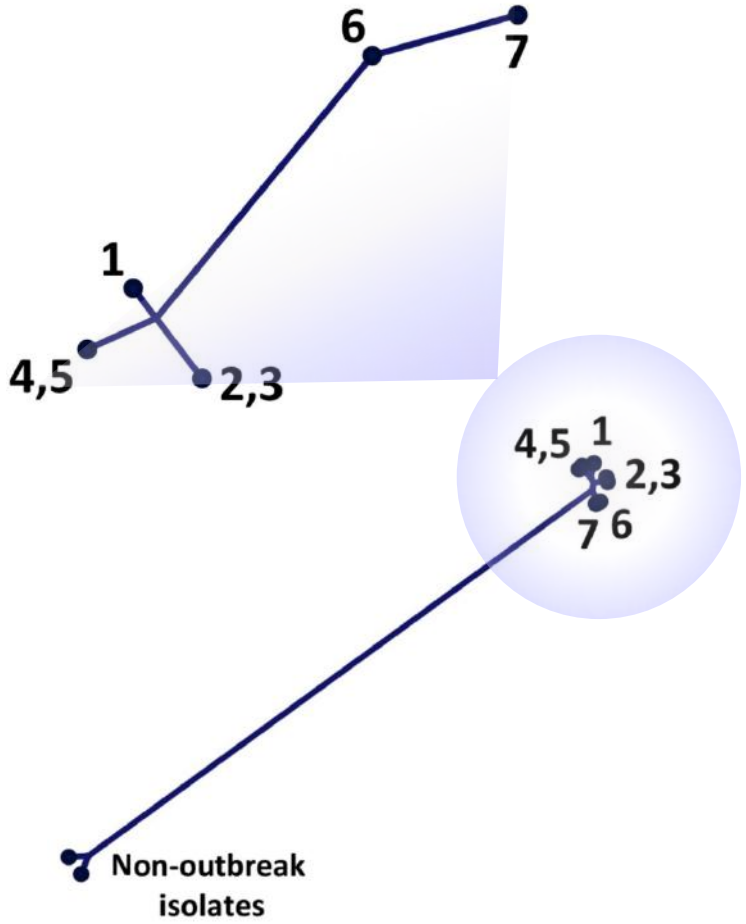
Study	Study 1. SR Harris et al. PMID: 23158674
Microorganism	Methicillin-resistant *Staphylococcus aureus*
Source of isolates	Human
Number of isolates analyzed	7 outbreak isolates (1 outbreak cluster) + 2 epi unrelated isolates
Type of outbreak	Hospital-associated outbreak



Original tree

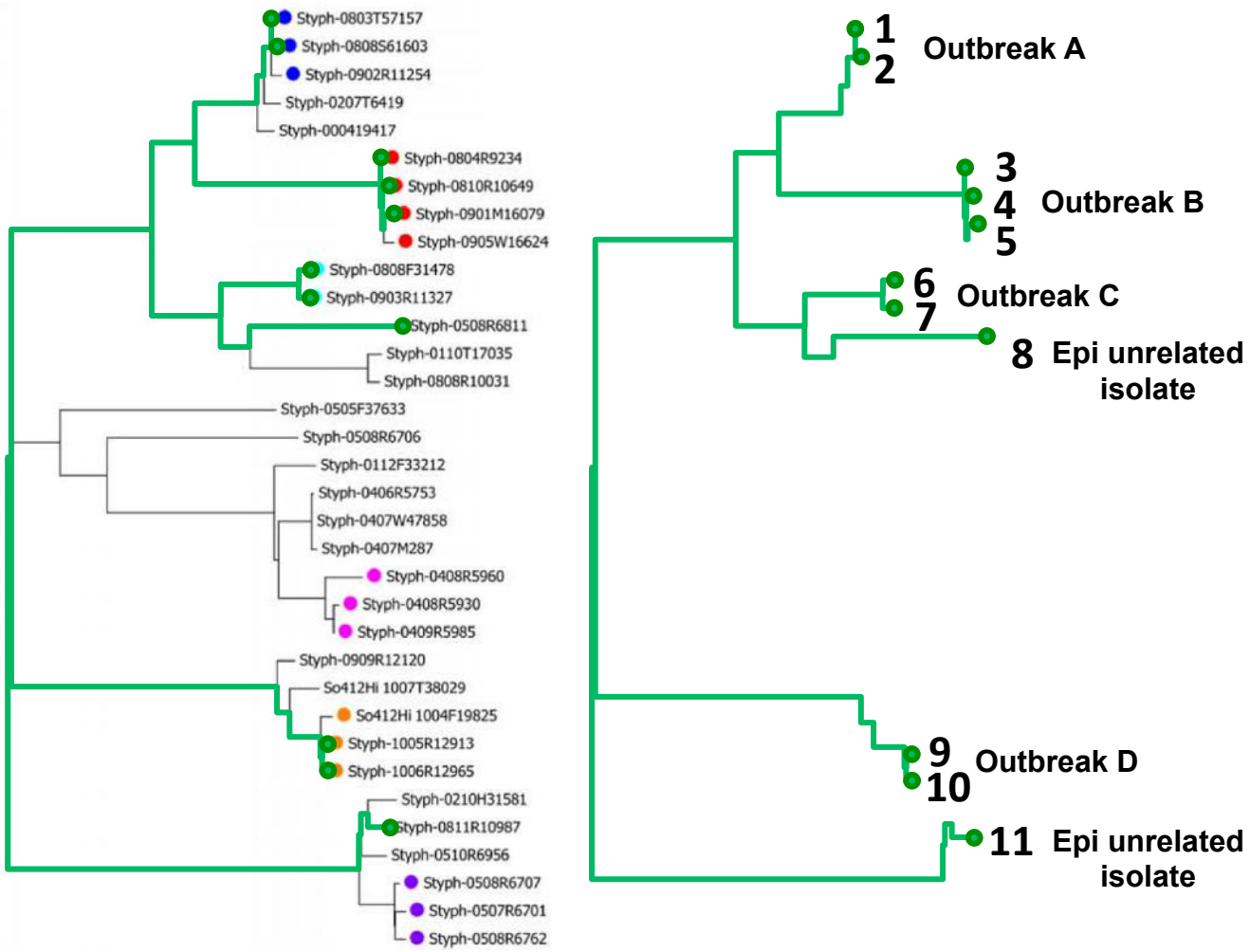


Validation tree

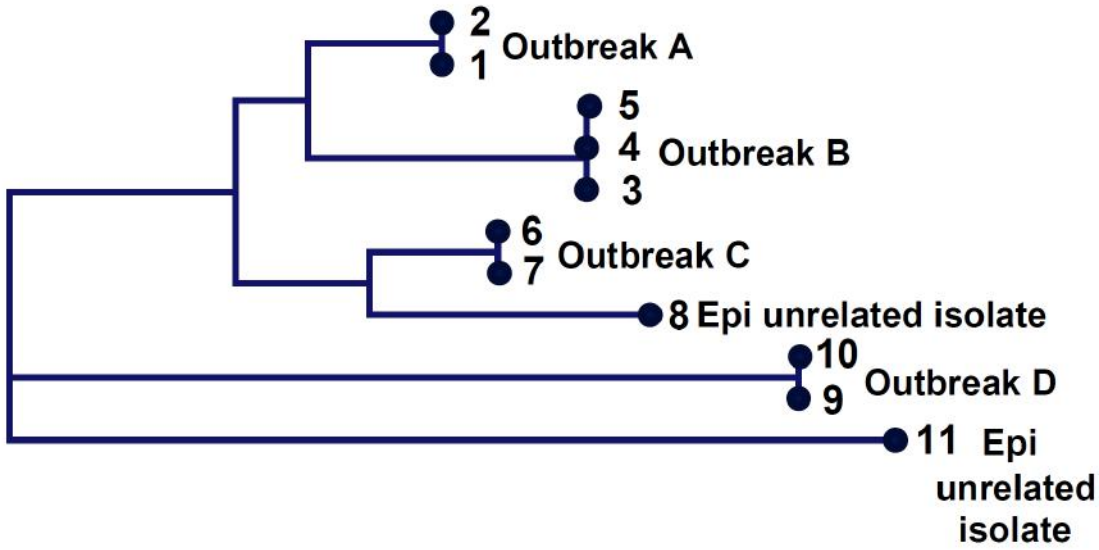


Study	Study 2. P Leekitcharoenphon et al. PMID: 24505344
Microorganism	*Salmonella enterica* serovar Typhimurium
Source of isolates	Human
Number of isolates analyzed	9 outbreak isolates (4 outbreak clusters) + 2 epi unrelated isolates
Type of outbreak	Foodborne outbreak

Original tree



Validation Tree



Between-laboratory comparison of genotyping pipelines

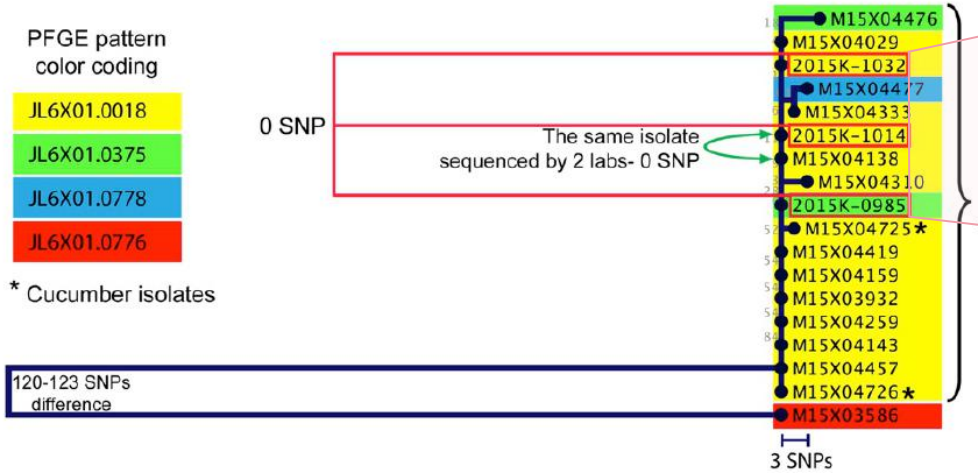
PLoS Currents
Rapid Sharing of Research in Progress

Version 1. PLoS Curr. 2016 November 22; 8:
ecurrents.outbreaks.1bb3e36e74bd5779bc43ac3a8dae52e6.
Published online 2016 November 22.
doi: 10.1371/currents.outbreaks.1bb3e36e74bd5779bc43ac3a8dae52e6
PMCID: PMC5145817
Research Article

Laboratory Investigation of *Salmonella enterica* serovar Poona Outbreak in California: Comparison of Pulsed-Field Gel Electrophoresis (PFGE) and Whole Genome Sequencing (WGS) Results

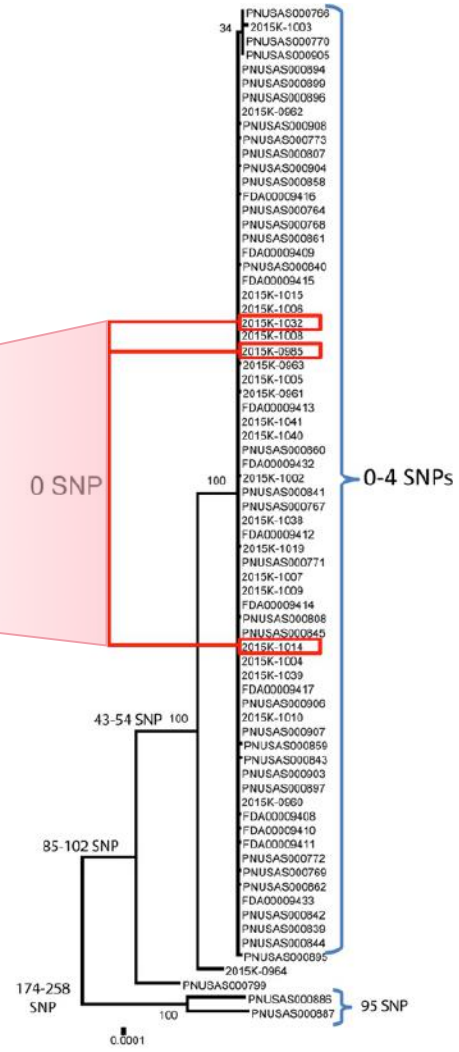
Varvara K. Kozyreva, John Crandall, Ashley Sabol, Alyssa Poe, Peng Zhang, Jeniffer Concepción-Acevedo,* Morgan N. Schroeder, Darlene Wagner, Jeffrey Higa, Eija Trees, and Vishnu Chaturvedi

Microbial Diseases Lab CA



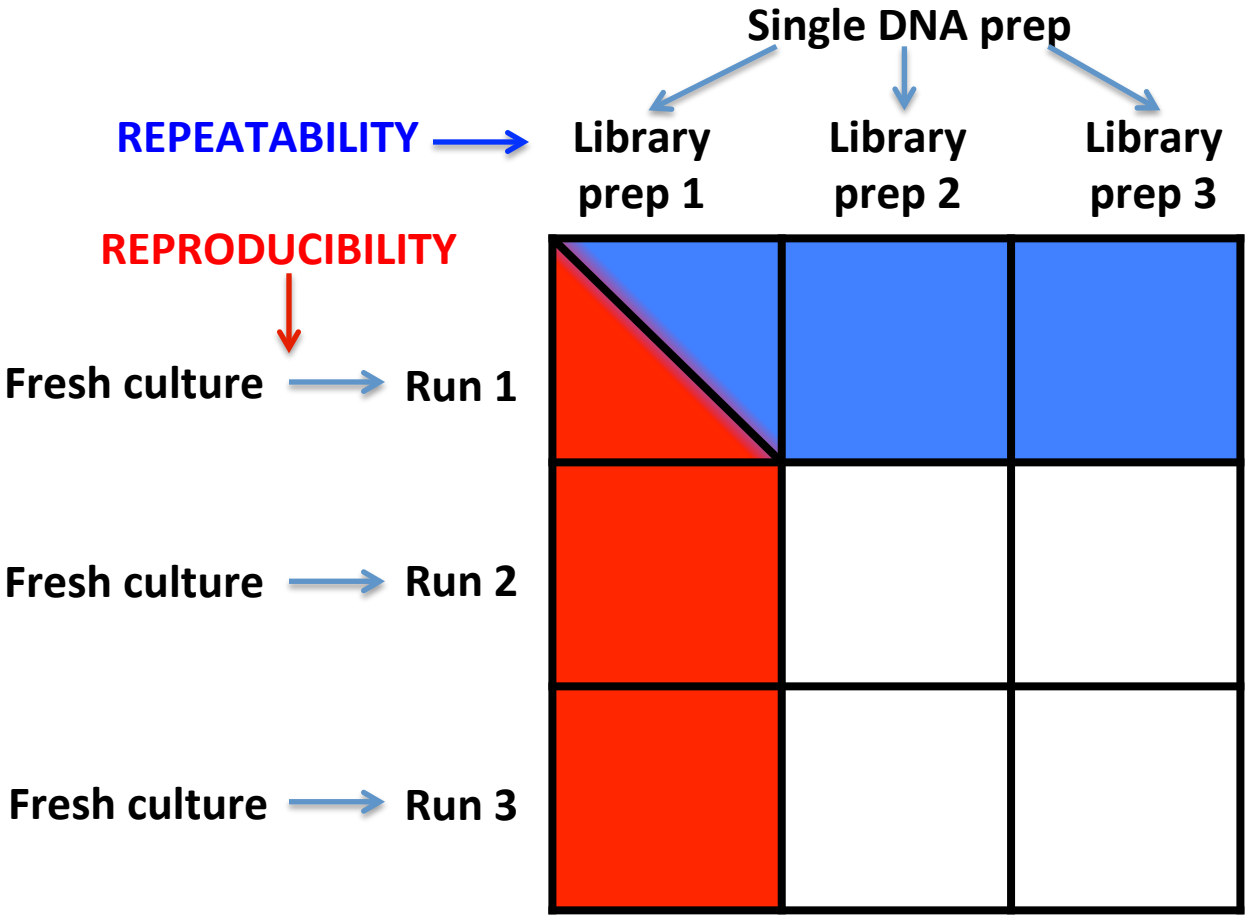
vs.

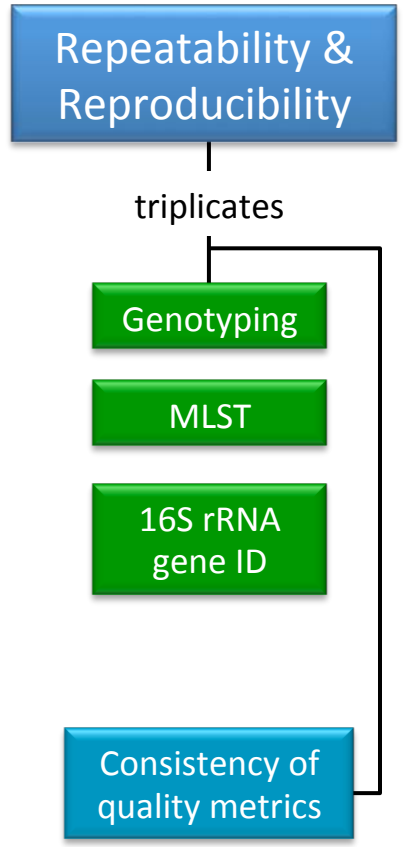
CDC



Repeatability (=precision within run=inter-assay agreement) –
The concordance of the assay results and quality metrics obtained for a sample tested multiple times within the same sequencing run.

Reproducibility (=precision between runs=intra-assay agreement) –
the consistency of the assay results and quality metrics for the same sample sequenced on different occasions by different operators.





Repeatability & Reproducibility

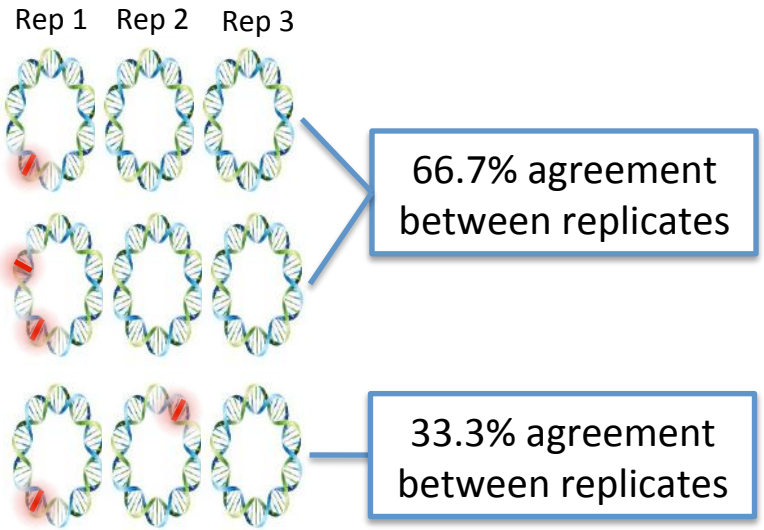
Repeatability and reproducibility of SNP calling–
between- and within-run precision of single nucleotide variant detection.

Evaluation of absolute precision **per replicate**

The whole genome of a single replicate = an independent test

Evaluation of precision relative to the genome size (**per base pair**)

Each nucleotide call in the WGS = an independent test (a percentage of the number of bp called in the genome)



$$\frac{\text{bp sequenced in agreement}}{\text{bp sequenced in total}}$$

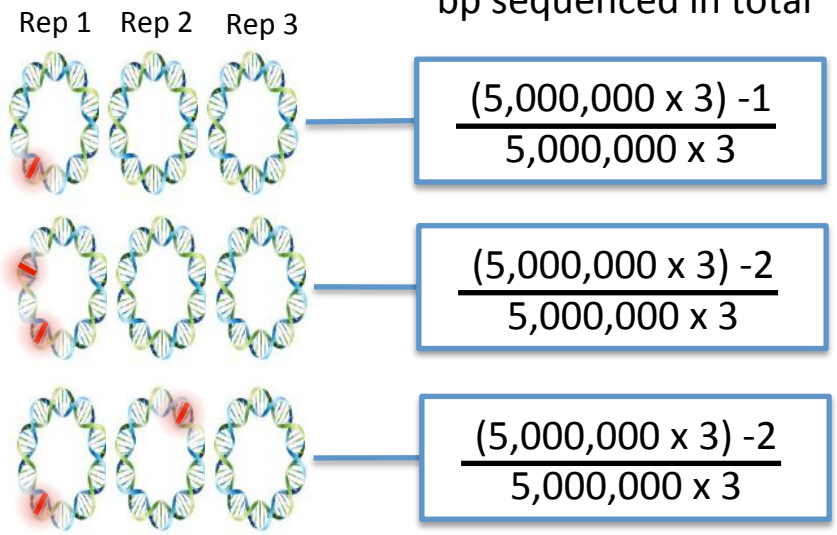


Image: https://epsrc.ukri.org/

Repeatability & Reproducibility

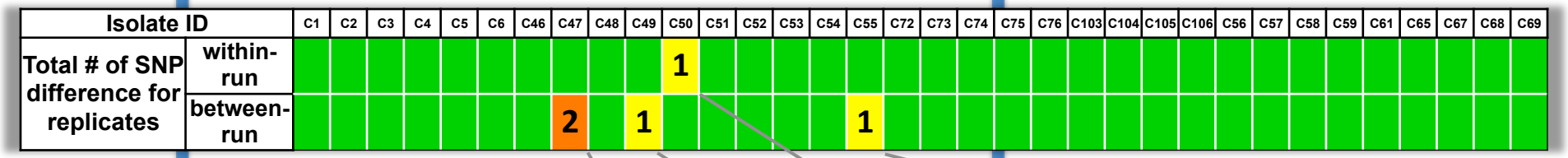
Repeatability and reproducibility of SNP calling–
between- and within-run precision of single nucleotide variant detection.

Evaluation of absolute precision **per replicate**

The whole genome of a single replicate = an independent test

Evaluation of precision relative to the genome size (**per base pair**)

Each nucleotide call in the WGS = an independent test (a percentage of the number of bp called in the genome)



99.99997%
99.99998%
99.99999%
99.99999%

... of covered genome (sequenced in triplicate)

Precision per replicate:

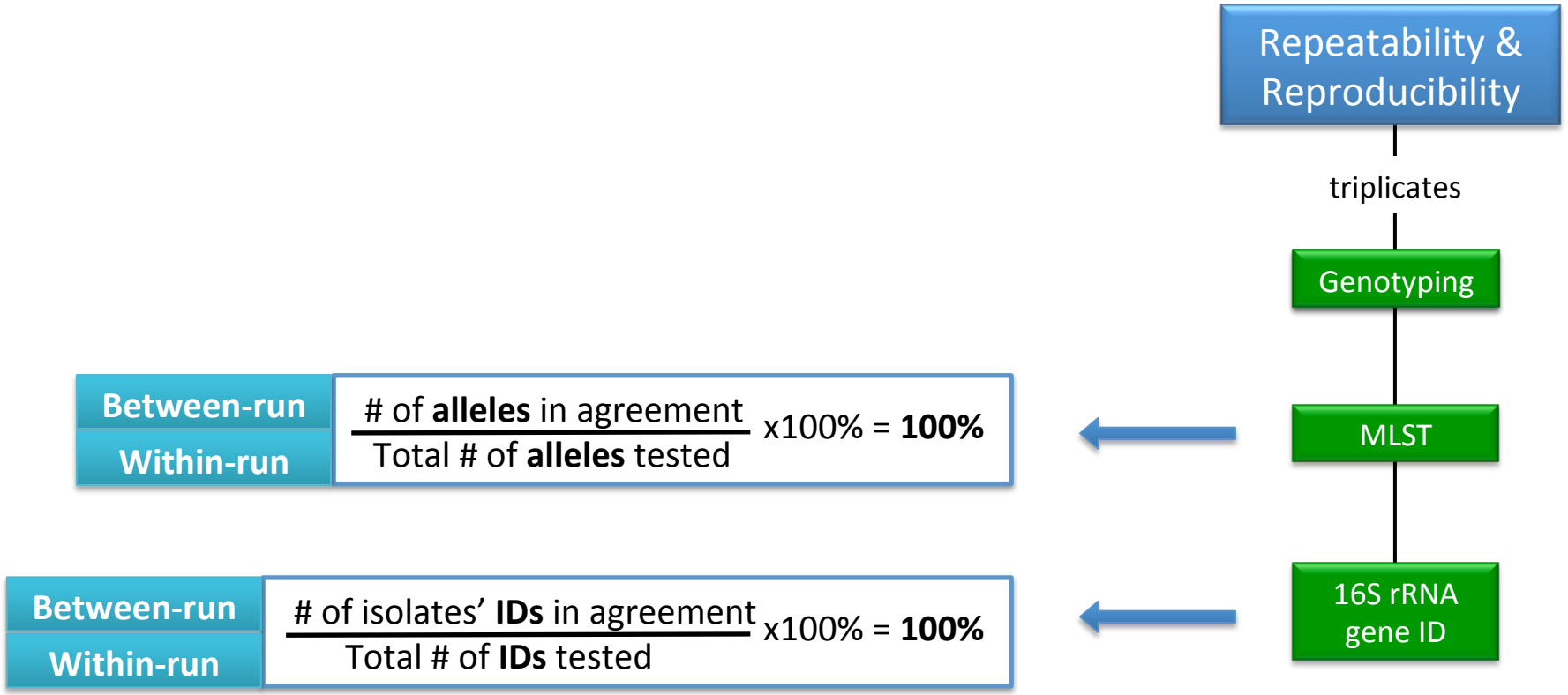
Within-run precision =
(101/102) x 100%
= 99.02%

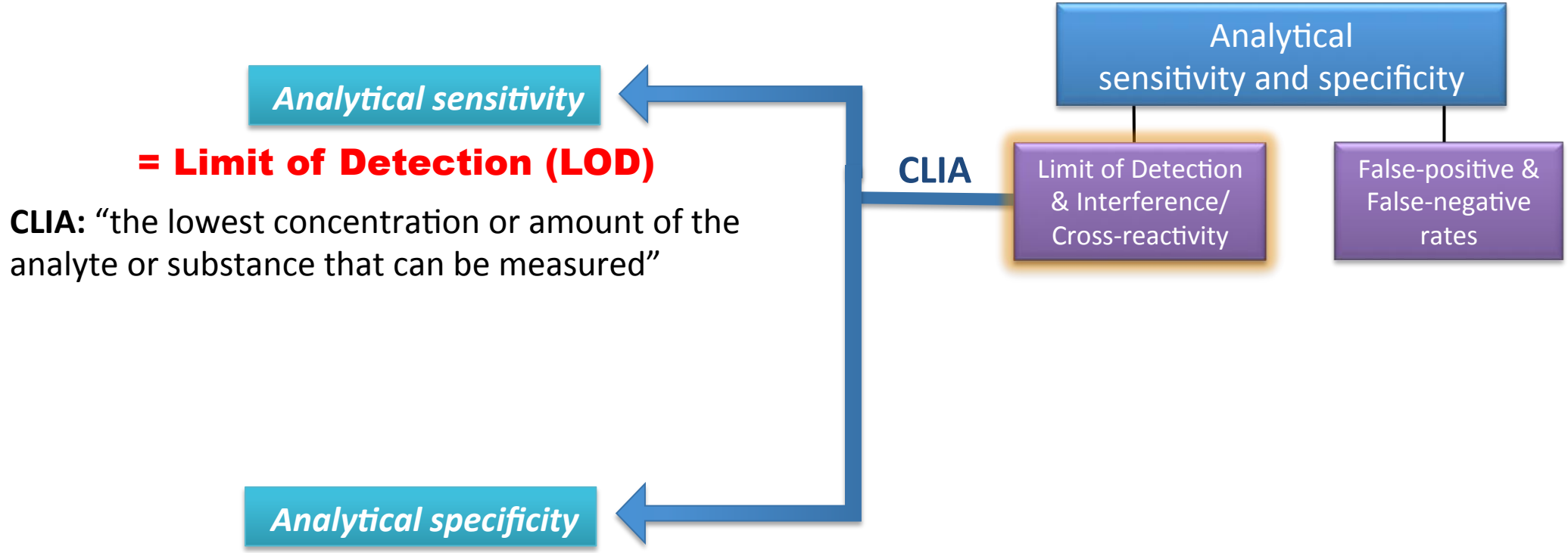
Between-run precision =
(99/102) x 100%
= 97.05%

Precision per base pair:

Within-run precision (avg) =
99.9999997%

Between-run precision (avg) =
99.999998%


Repeatability & Reproducibility
triplicates
Genotyping
MLST
16S rRNA gene ID
Between-run
Within-run
# of alleles in agreement / Total # of alleles tested x100% = 100%
Between-run
Within-run
# of isolates' IDs in agreement / Total # of IDs tested x100% = 100%



CLIA: "the lowest concentration or amount of the analyte or substance that can be measured"

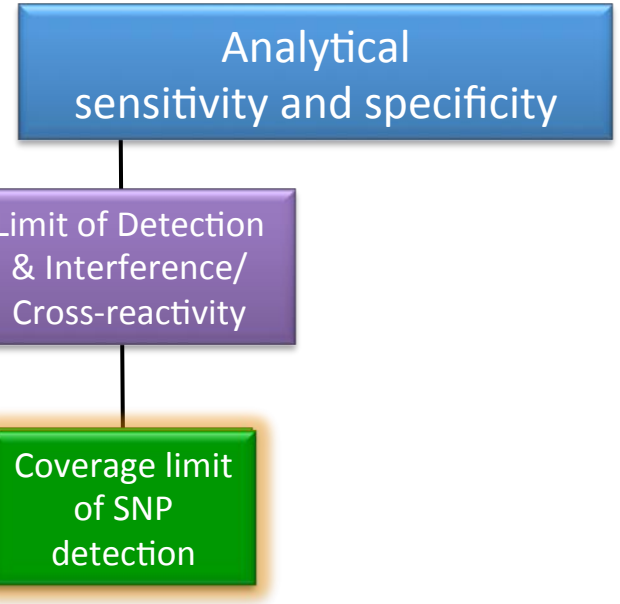
CLIA: "the extent to which the method measures the analyte for which it is reporting results"

+

document information regarding <u>interfering substances</u> (from product information, literature, or own testing)

- CLIA regulations 42CFR § 493.1253
- CLSI. MM09-A2, 2nd ed., 2014 CLSI, Wayne, PA

Analytical sensitivity of WGS

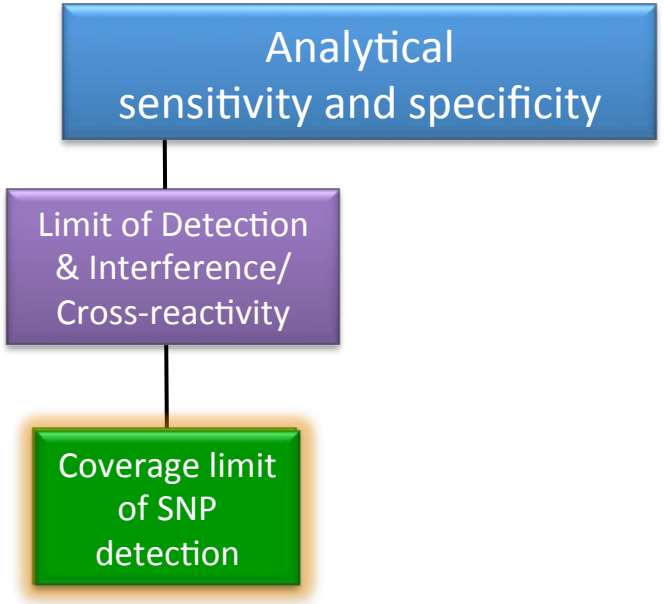


LOD = minimum genome coverage that allows accurate SNP detection

Downsampling

Species	Original coverage	Number of SNP detected between the validation sample and reference								
		At original coverage	60x	50x	40x	30x	20x	15x	10x	5x
Escherichia coli O157:H7	69.3x	5	5	5	5	5	5	4	4	4
Aeromonas hydrophilia	116.5x	1	1	1	1	1	1	1	1	1
Escherichia coli	85.2x	22	22	21	20	19	18	17	17	15
Staphylococcus aureus	145x	0	0	0	0	0	0	0	0	0
Salmonella enterica ser Typhimurium	86.6x	12	12	12	8	8	7	7	7	7
Enterococcus faecalis	58x	3	-	3	3	3	3	3	3	4
Legionella pneumophila	99.8x	2	2	2	2	2	2	2	2	2
Escherichia coli O121:H19	53x	0	-	0	0	0	0	0	0	0
Salmonella enterica ser Enteritidis	76.8x	0	0	0	0	0	0	0	0	0

Analytical sensitivity of WGS



LOD = minimum genome coverage that allows accurate SNP detection

Downsampling

Species	Original coverage	N	LOD_{SNP}	detected between the validation sample and reference						
		At original coverage	60x	50x	40x	30x	20x	15x	10x	5x
Escherichia coli O157:H7	69.3x	5	5	5	5	5	5	4	4	4
Aeromonas hydrophilia	116.5x	1	1	1	1	1	1	1	1	1
Escherichia coli	85.2x	22	22	21	20	19	18	17	17	15
Staphylococcus aureus	145x	0	0	0	0	0	0	0	0	0
Salmonella enterica ser Typhimurium	86.6x	12	12	12	8	8	7	7	7	7
Enterococcus faecalis	58x	3	-	3	3	3	3	3	3	4
Legionella pneumophila	99.8x	2	2	2	2	2	2	2	2	2
Escherichia coli O121:H19	53x	0	-	0	0	0	0	0	0	0
Salmonella enterica ser Enteritidis	76.8x	0	0	0	0	0	0	0	0	0

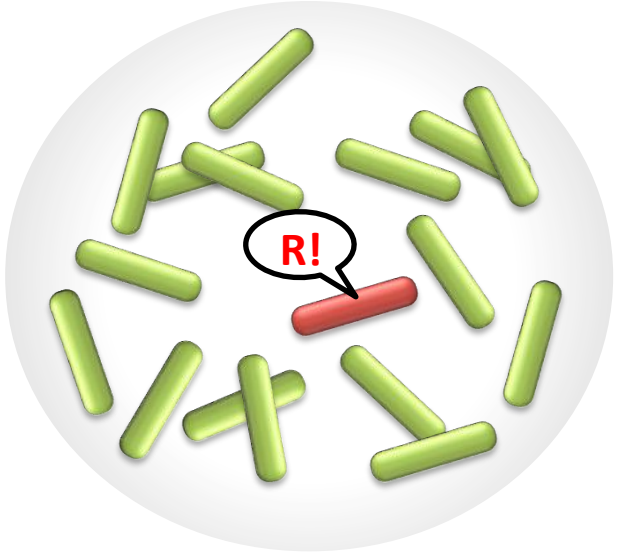
Analytical sensitivity of WGS

Analytical sensitivity and specificity

- Limit of Detection & Interference/ Cross-reactivity

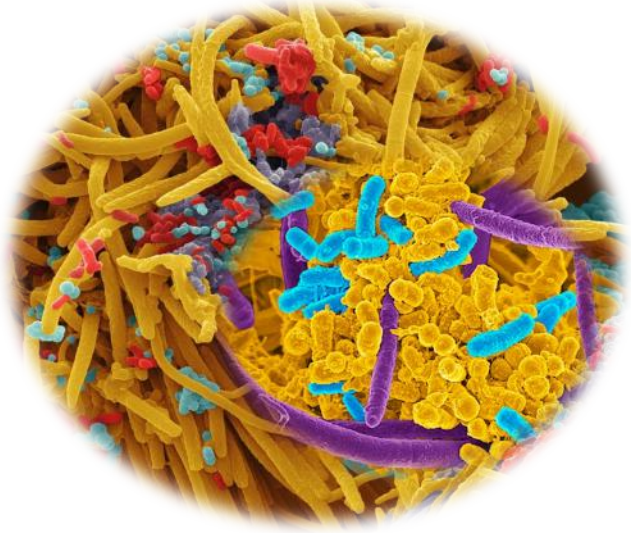
?

LOD for heteroresistance detection in *Mycobacterium tuberculosis*



?

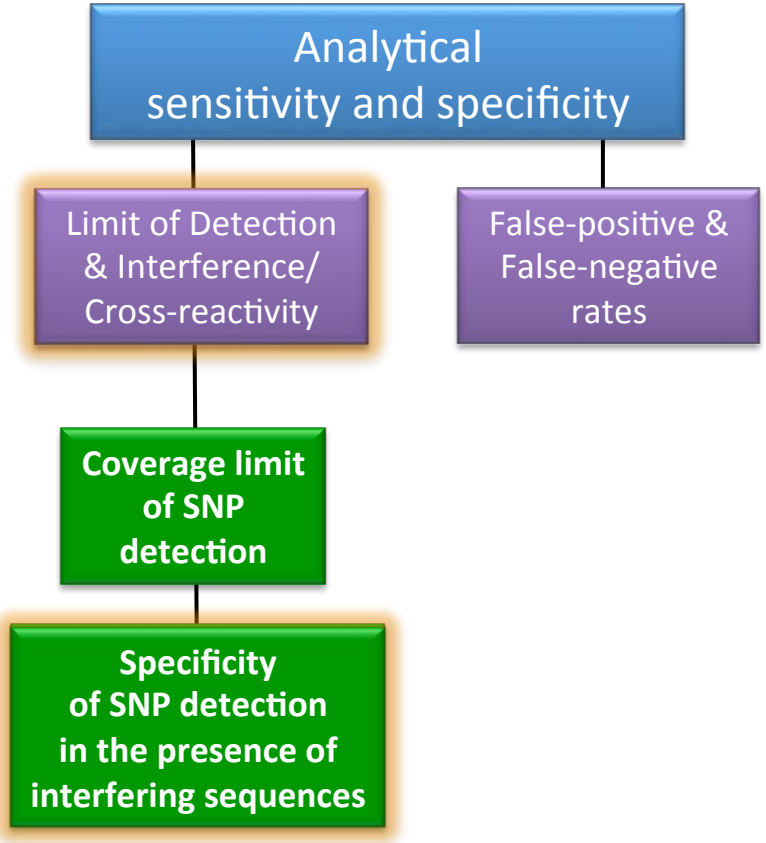
LOD for metagenomic pathogen detection



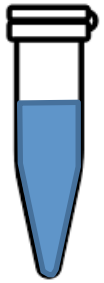
Heteroresistance in Mtb: Bernard C et al., J Antimicrob Chemother. 2016 Dec;71(12):3465-3472.
Electron microscopy: https://www.masterfile.com/, SPL/mediadrumworld.com

Analytical specificity of WGS

- the ability of an assay to detect only the intended target in the presence of potentially cross-reacting nucleotide sequences



Example of <u>modeled</u> interference/cross-reactivity from contaminating DNA:



	Number of SNPs differing w/reference
Pure genome of *E. coli*	22
Contamination 25% (*in silico*) w/ *Salmonella enterica*	22
Contamination 50% (*in silico*) w/ *Salmonella enterica*	24

1 SNP missed

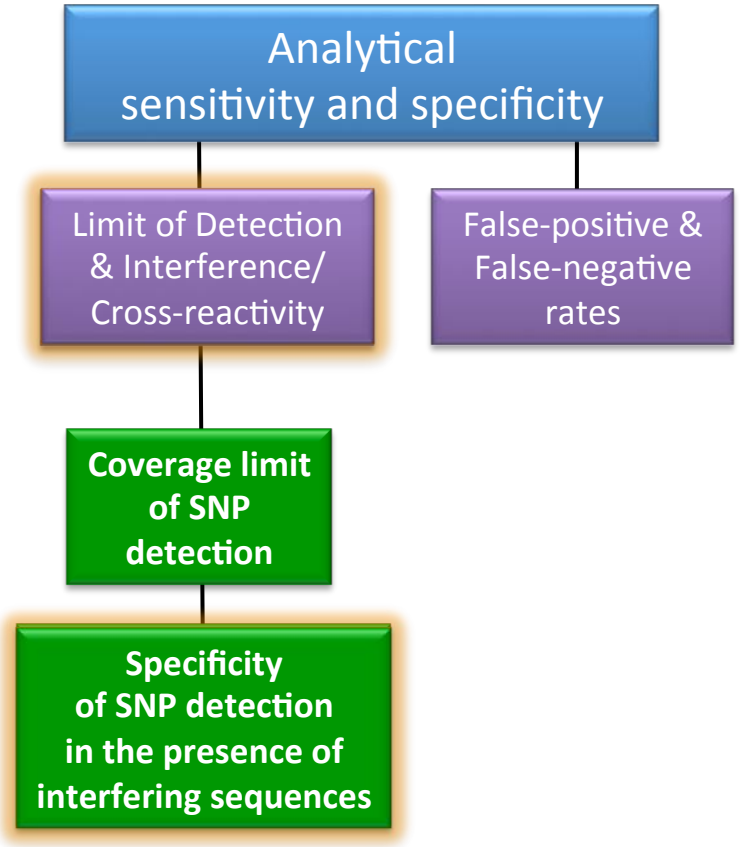
1 non-specific SNP detected

68% reads mapped
59.6% reads in pairs

Image: https://www.beckman.com/

Analytical specificity of WGS

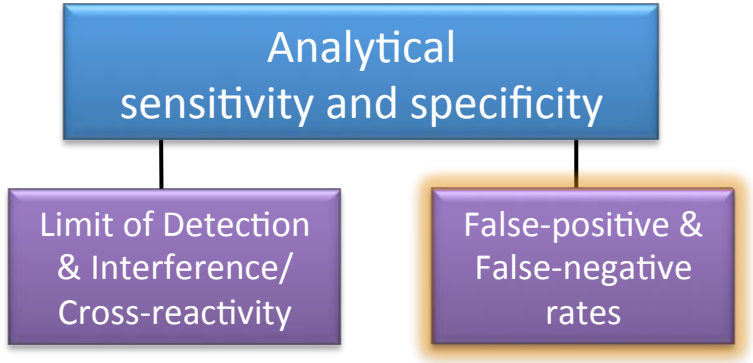
Samples		C3 *E.coli*	C3 *E.coli* + C75 *Salmonella enterica*	C3 *E.coli* + C1 *E.coli*	C3 *E.coli* + C54 *Acinetobacter baumannii*	C3 *E.coli* + C57 *Mycobacterium tuberculosis*	C3 *E.coli* + C5 *Staphylococcus aureus*
Mapping quality metrics	**% of Contamination reads**	no contamination	50%	50%	50%	50%	50%
	% of Mapped reads	99%	68%	91%	51%	67%	49%
	% of Not mapped reads	1%	32%	9%	49%	33%	51%
	% of Reads in pairs	90%	59.60%	81%	48.30%	60%	46.60%
	% of Broken paired reads	9%	8%	10%	3%	6%	3%
	% of reference covered	99%	99%	99%	99%	99%	99%
SNP calling specificity	**SNPs between sequence and reference**	22	24	22	40	22	36
	Number of FN SNPs	NA	1	0	0	0	0
	Number of FP SNPs	NA	1	0	18	0	14



$$\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\%$$

$$\text{Specificity} = \frac{TN}{TN + FP} \times 100\%$$

TP- True positive results
TN- True negative results
FP- False positive
FN- False negative

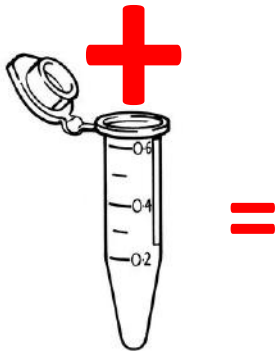


Often used in DIAGNOSTIC (or clinical) sensitivity and specificity calculations

DIAGNOSTIC SENSITIVITY & SPECIFICITY-
ability to predict the disease or condition in a person:

vs.

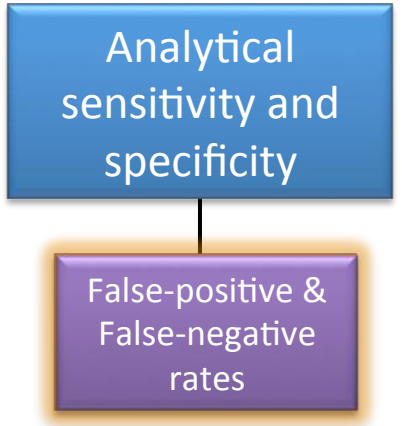
Analytical Sensitivity & Specificity as false-positive & false-negative rates of WGS assays



=

Saah AJ et al. Ann Intern Med. 1997 Jan 1;126(1):91-4.
Image: http://zombie-authors.blogspot.com

Analytical Sensitivity & Specificity as false-positive & false-negative rates of WGS assays



Analytical sensitivity- the likelihood that a WGS assay will detect sequence variation when it is present (this value reflects a false negative rate of the test).

$$\text{Analytical sensitivity} = \frac{TP}{TP + FN} \times 100\%$$

Analytical specificity- the probability that a WGS assay will not detect sequence variations when none are present (this value reflects a test's false positive rate).

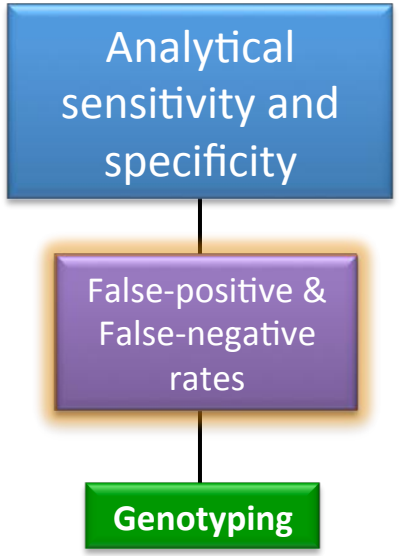
$$\text{Analytical specificity} = \frac{TN}{TN + FP} \times 100\%$$

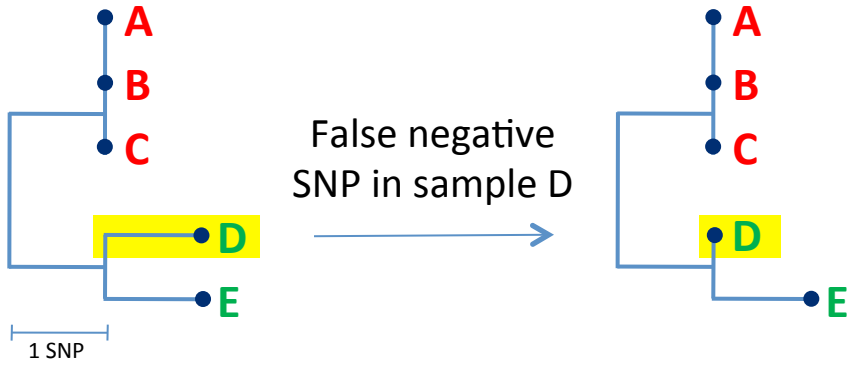
TP- True positive results
TN- True negative results
FP- False positive
FN- False negative

Gargis AS et al., J Clin Microbiol. 2016 Dec;54(12):2857-2865.
Pont-Kingdon G et al., Arch Pathol Lab Med. 2012 Jan;136(1):41-6.

Analytical sensitivity in genotyping assay is the likelihood that all the SNPs differing between the isolates will be detected.

False negative result - missed sequence variations, e.g.:

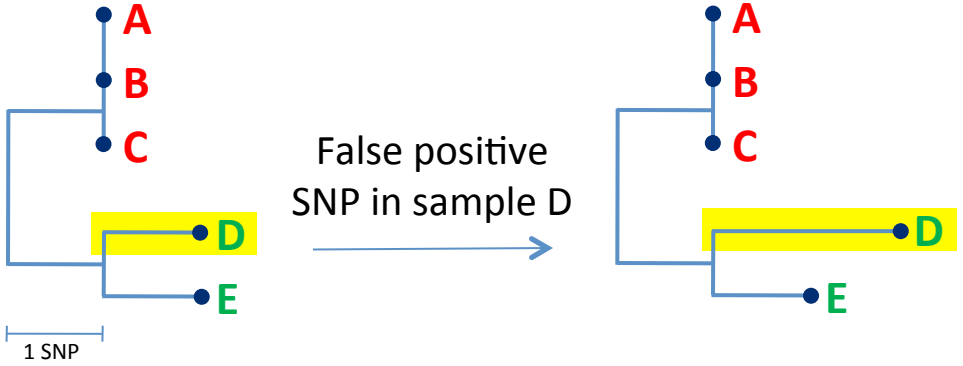




Reference tree Validation tree

Analytical specificity in genotyping assay is the likelihood that variation between the isolates (SNP) will not be detected when none are present.

False positive result – false sequence variations (introduced either during library prep/sequencing, or data analysis), e.g.:



Reference tree Validation tree

10- *Enterobacteriaceae*
5- Gram-positive cocci isolates
5- Gram-negative non-fermenting bacterial isolates
9- *Mycobacterium tuberculosis*
5- representatives of miscellaneous species

Validation Set
34 bacterial isolates

WHOLE GENOME SEQUENCING VALIDATION IN PUBLIC HEALTH MICROBIOLOGY LAB SETTINGS

Accuracy

- MiSeq Illumina platform accuracy
  - 99.999%
- Test accuracy
  - Genotyping — 100%
  - MLST — 100%
  - 16S rRNA gene ID — 100%
  - Antibiotic resistance genes detection — 100%
- Bioinformatics pipeline accuracy
  - 100%

Reproducibility/ Repeatability

triplicates

- Genotyping: per replicate per base pair
  - **within run / between run:**
  - 99.02% / 97.05%
  - 99.9999997% / 99.999998%
- MLST — 100%
- 16S rRNA gene ID — 100%

Analytical sensitivity and specificity

- False-positive & False-negative rates
  - Genotyping — 100%
  - MLST — 100%
- Limit of Detection & Interference/ Cross-reactivity
  - Coverage limit of SNP detection — 60x
  - Specificity of SNP detection in the presence of interfering sequences ✓

WGS implementation in CLIA settings

ASSAY DEVELOPMENT
Wet-Lab / Dry-Lab

- Workflow development & optimization
- Empirically determined optimal assay conditions
- Documentation: SOP, worksheets, QC logs, calibration sheets, etc.
- Intended use, Acceptance and rejection criteria

ASSAY VALIDATION
Platform – Bioinformatics - Assay

- Validation Plan
- Ability to sequence targets of interest
- Accurate results of SNP detection / genotype / ID / resistance genes / etc
- Determine QC metrics and checkpoints

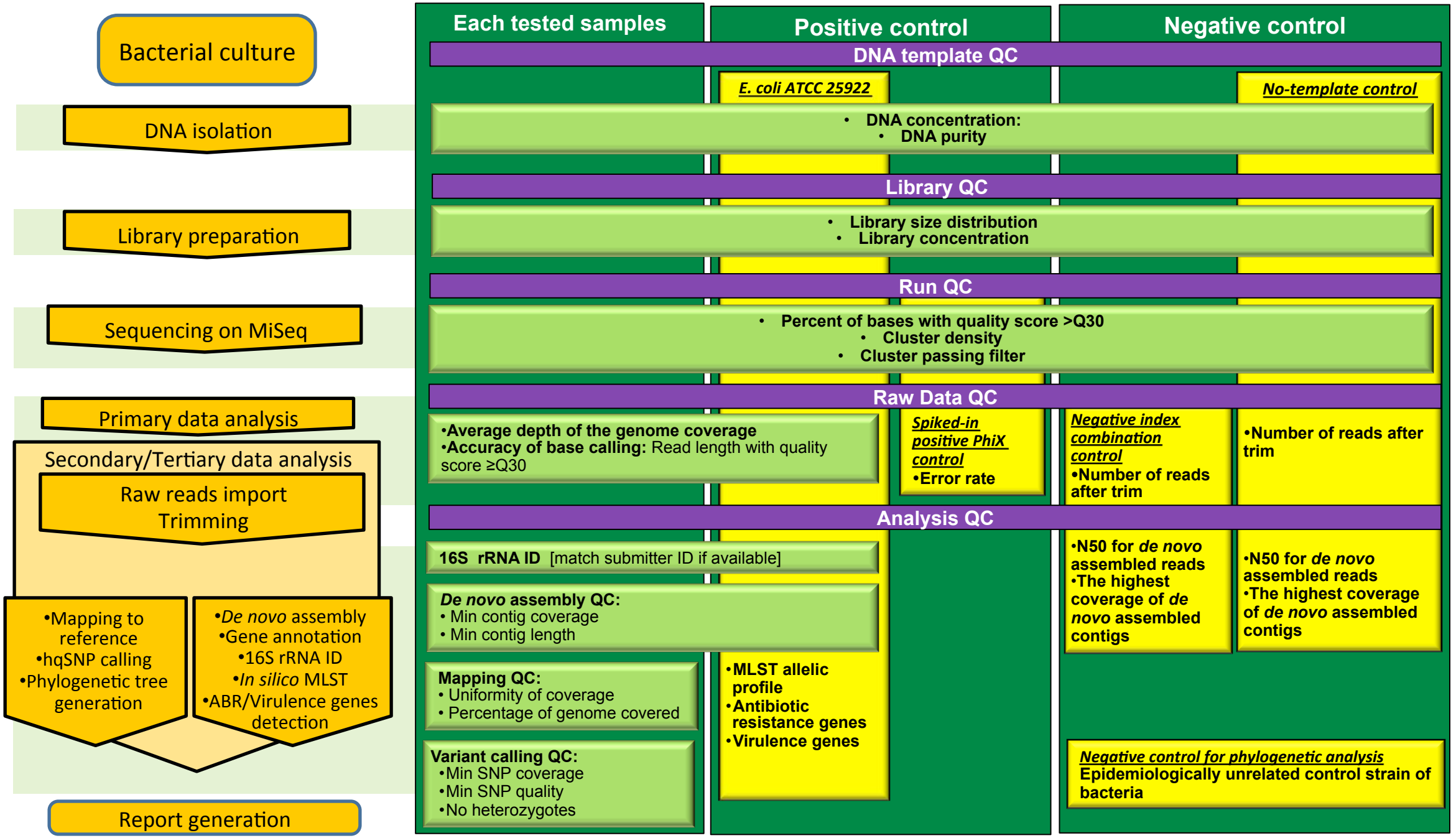
QUALITY MANAGEMENT

- Documented quality management plan
- Routine QC of all reagents
- Equipment preventive maintenance and calibration
- Proficiency testing (PT)
- Competency assessment
- Re-validation/confirmation of the performance for assay modifications
- Instrument correlation
- Corrective actions

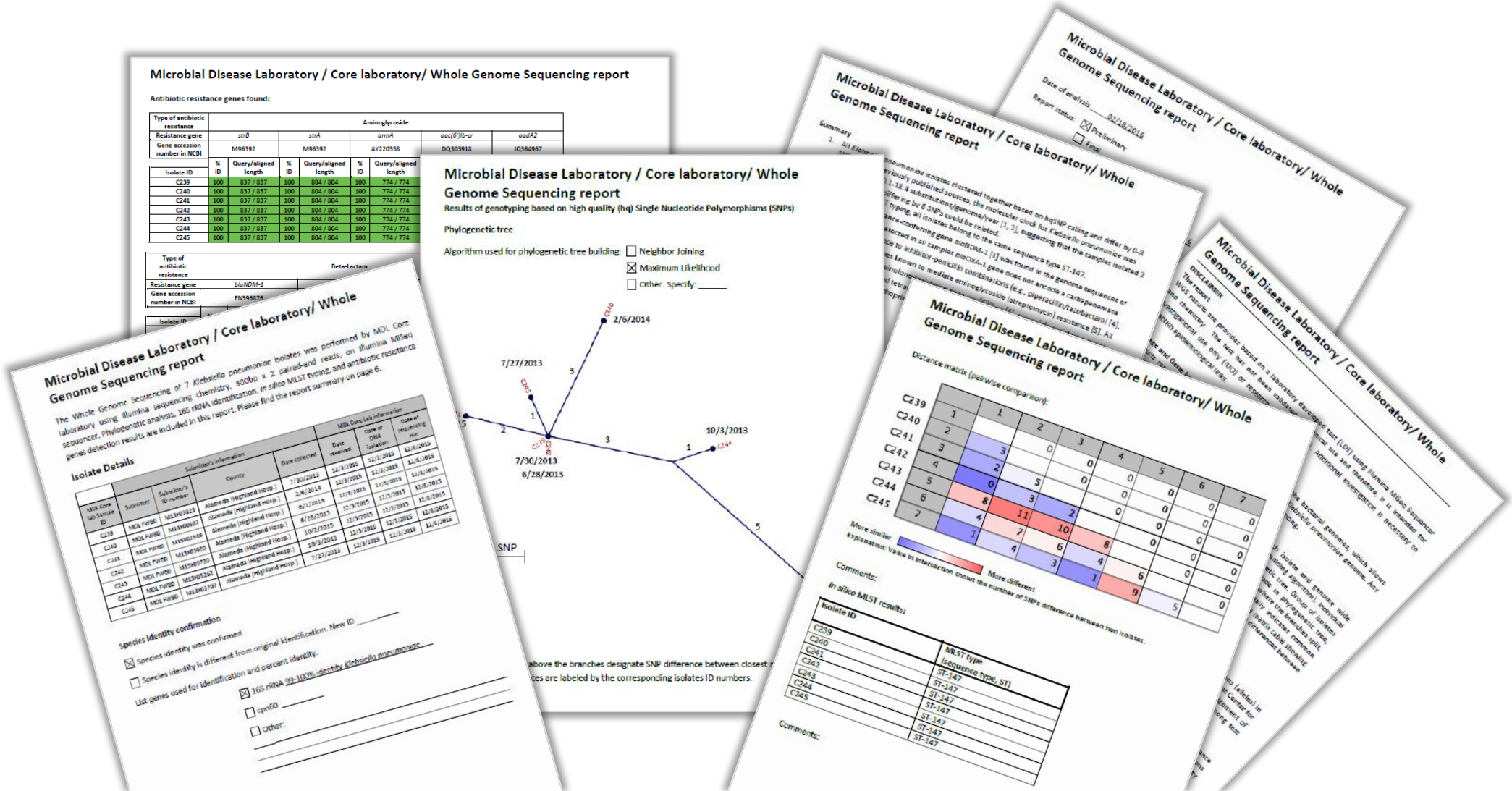
QUALITY WGS

Modified from Gargis AS et al., J Clin Microbiol. 2016 Dec;54(12):2857-2865.

WGS QUALITY CONTROL SCHEME



REPORTING LANGUAGE




CLIA
INSPECTION

10- *Enterobacteriaceae*
5- Gram-positive cocci isolates
5- Gram-negative non-fermenting bacterial isolates
9- *Mycobacterium tuberculosis*
5- representatives of miscellaneous species

Validation Set
34 bacterial isolates

WHOLE GENOME SEQUENCING VALIDATION IN PUBLIC HEALTH MICROBIOLOGY LAB SETTINGS

Accuracy

- MiSeq Illumina platform accuracy — 99.999%
- Test accuracy
  - Genotyping — 100%
  - MLST — 100%
  - 16S rRNA gene ID — 100%
  - Antibiotic resistance genes detection — 100%
- Bioinformatics pipeline accuracy — 100%

Reproducibility/ Repeatability

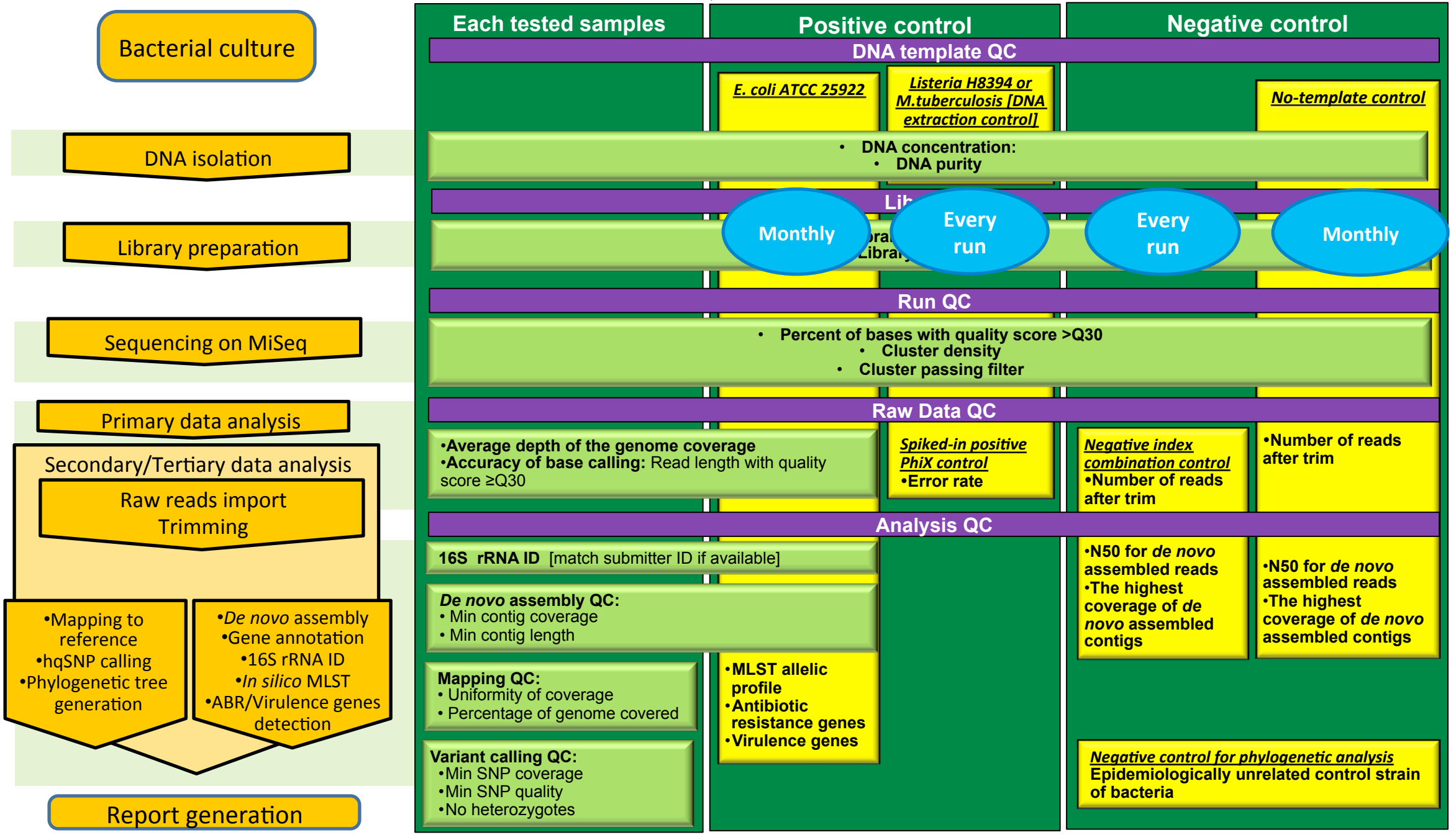
triplicates

- Genotyping: per replicate per base pair — **within run / between run:** 99.02% / 97.05%; 99.9999997% / 99.999998%
- MLST — 100%
- 16S rRNA gene ID — 100%

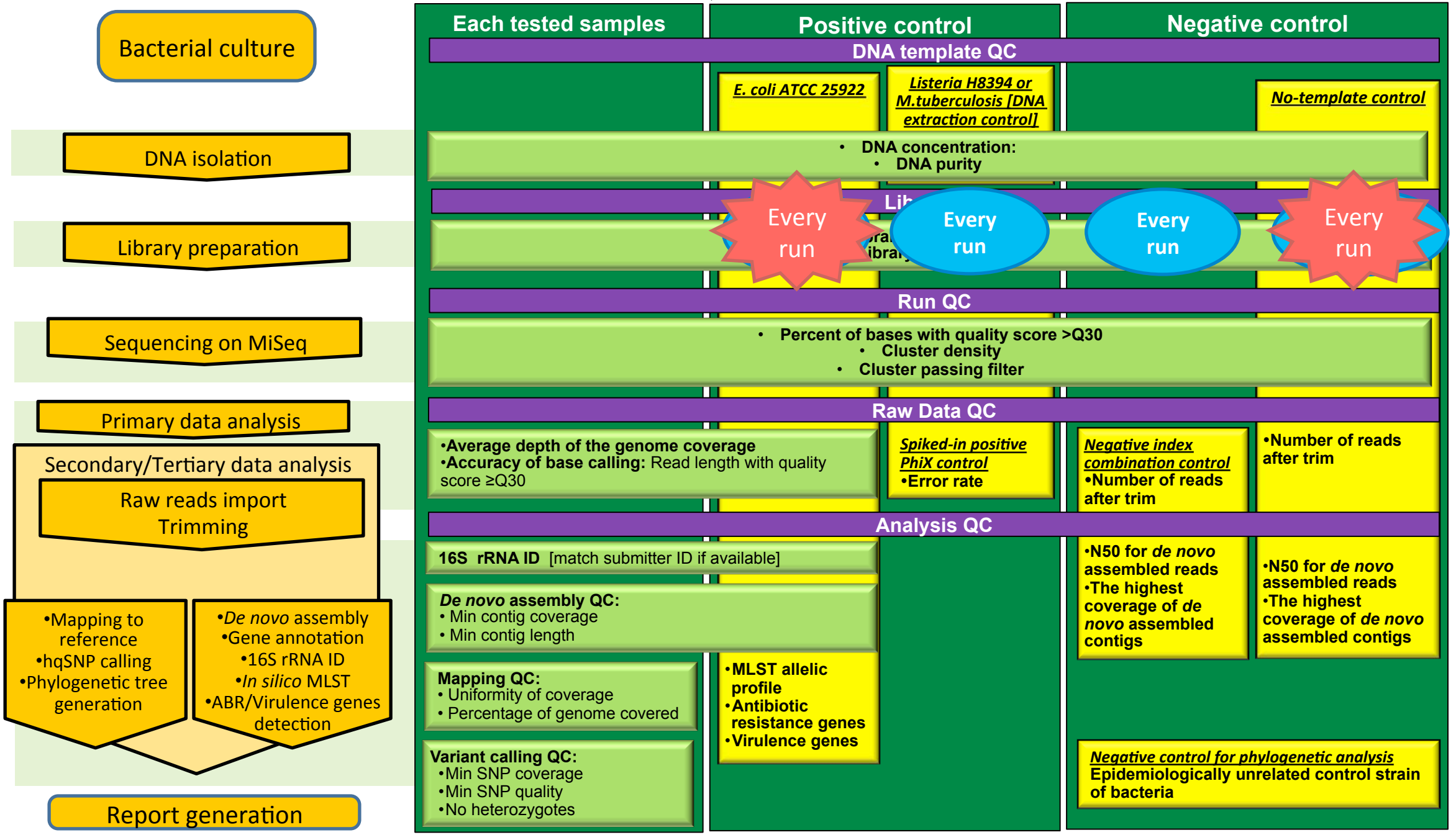
Analytical sensitivity and specificity

- False-positive & False-negative rates
  - Genotyping — 100%
  - MLST — 100%
- Limit of Detection & Interference/ Cross-reactivity
  - Coverage limit of SNP detection — 60x
  - Specificity of SNP detection in the presence of interfering sequences ✓

WGS QUALITY CONTROL SCHEME



WGS QUALITY CONTROL SCHEME



Re-Validation / Confirmation of the Performance for the modified components of the WGS pipeline

- An amendment of Illumina MiSeq v.2 chemistry sequencing kits
- A new processing algorithm for highly-contagious pathogens
- Added virulence genes analysis
- Added Kmer species identification analysis
- Developed a custom script to create an automated bioinformatics pipeline
- … Change is the only constant …

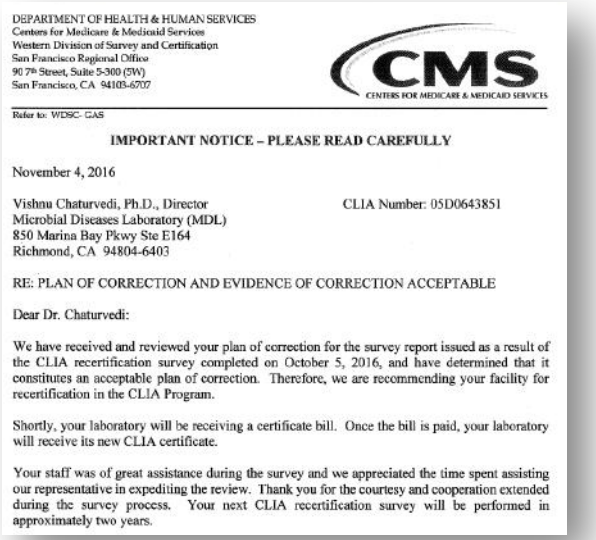
More validations to come…

Summary

- ✓ **Established the performance specifications** for WGS in the application to public health microbiology in accordance with CLIA guidelines for the LDTs
- ✓ **Developed quality assurance (QA) and quality control (QC) measures** for WGS.
- ✓ **Developed a validation set** of pathogenic bacteria for further validations of new WGS instruments/platforms, QC purposes, and multi-laboratory comparisons.
- ✓ **Developed a modular template for the validation** and re-validation of ‘wet bench' and ‘dry bench' components of WGS.
- ✓ **Designed laboratory reports** for end users with or without WGS expertise.

✓ Successfully passed CLIA inspection!

- ✓ Validation report is available as a supplementary material to the published manuscript in Journal of Clinical Microbiology, 2017 Aug; 55(8):2502-2520 (PMID: 28592550).



DEPARTMENT OF HEALTH & HUMAN SERVICES
Centers for Medicare & Medicaid Services
Western Division of Survey and Certification
San Francisco Regional Office
90 7th Street, Suite 5-300 (5W)
San Francisco, CA 94103-6707

CMS
CENTERS FOR MEDICARE & MEDICAID SERVICES

Refer to: WDSC- GAS

IMPORTANT NOTICE – PLEASE READ CAREFULLY

November 4, 2016

Vishnu Chaturvedi, Ph.D., Director
Microbial Diseases Laboratory (MDL)
850 Marina Bay Pkwy Ste E164
Richmond, CA 94804-6403

CLIA Number: 05D0643851

RE: PLAN OF CORRECTION AND EVIDENCE OF CORRECTION ACCEPTABLE

Dear Dr. Chaturvedi:

We have received and reviewed your plan of correction for the survey report issued as a result of the CLIA recertification survey completed on October 5, 2016, and have determined that it constitutes an acceptable plan of correction. Therefore, we are recommending your facility for recertification in the CLIA Program.

Shortly, your laboratory will be receiving a certificate bill. Once the bill is paid, your laboratory will receive its new CLIA certificate.

Your staff was of great assistance during the survey and we appreciated the time spent assisting our representative in expediting the review. Thank you for the courtesy and cooperation extended during the survey process. Your next CLIA recertification survey will be performed in approximately two years.

x 2

ACKNOWLEDGEMENTS

MDL Laboratory Director (former)
Dr. Vishnu Chaturvedi

Core laboratory
Chau-Linda Truong (former member)
Dr. Rituparna Mukhopadhyay
John Crandall
Dr. Matthew Sylvester
Dr. Zhirong Li

Food- and Waterborne Diseases Section
Dr. Stephanie Abromaitis
Greg Inami
Yue Qing Zhao
Francine Arroyo
Beverly Kaneko

Bacterial Diseases Section
Dr. Peng Zhang
Margot Graves (former member)
Robin Hogue (former member)

Mycobacteriology and Mycology Section
Dr. Ed Desmond
Terry Weber
Grace Lin

High Risk Pathogens Section
Dr. Jennifer Kyle
Mahtab Shahkarami
Alyssa Poe

Immunodiagnostics Section
Joseph Morgan
Frank Ni (former member)
Linda Sae-Jang

Compliance
Gillian Edwards

Collaborators outside of CDPH:

Dr. Alexander Greninger, UCSF/ University of Washington

Dr. Eija Trees, PulseNet Next Generation Subtyping Methods Unit, CDC
Dr. Heather Carleton, Enteric Diseases Laboratory Branch, CDC


KEEP
CALM
AND
VALIDATE