

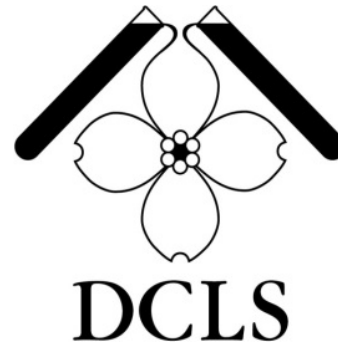
# Development and Implementation of a Quality System for Next-Generation Sequencing

---









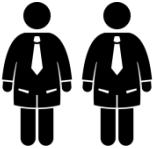
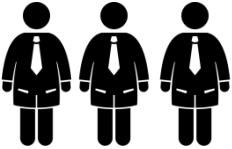
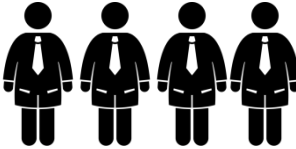
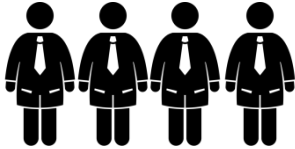



Lauren Turner, PhD

Lead Scientist

Virginia Division of Consolidated Laboratory Services



# DCLS' Phased Implementation of NGS

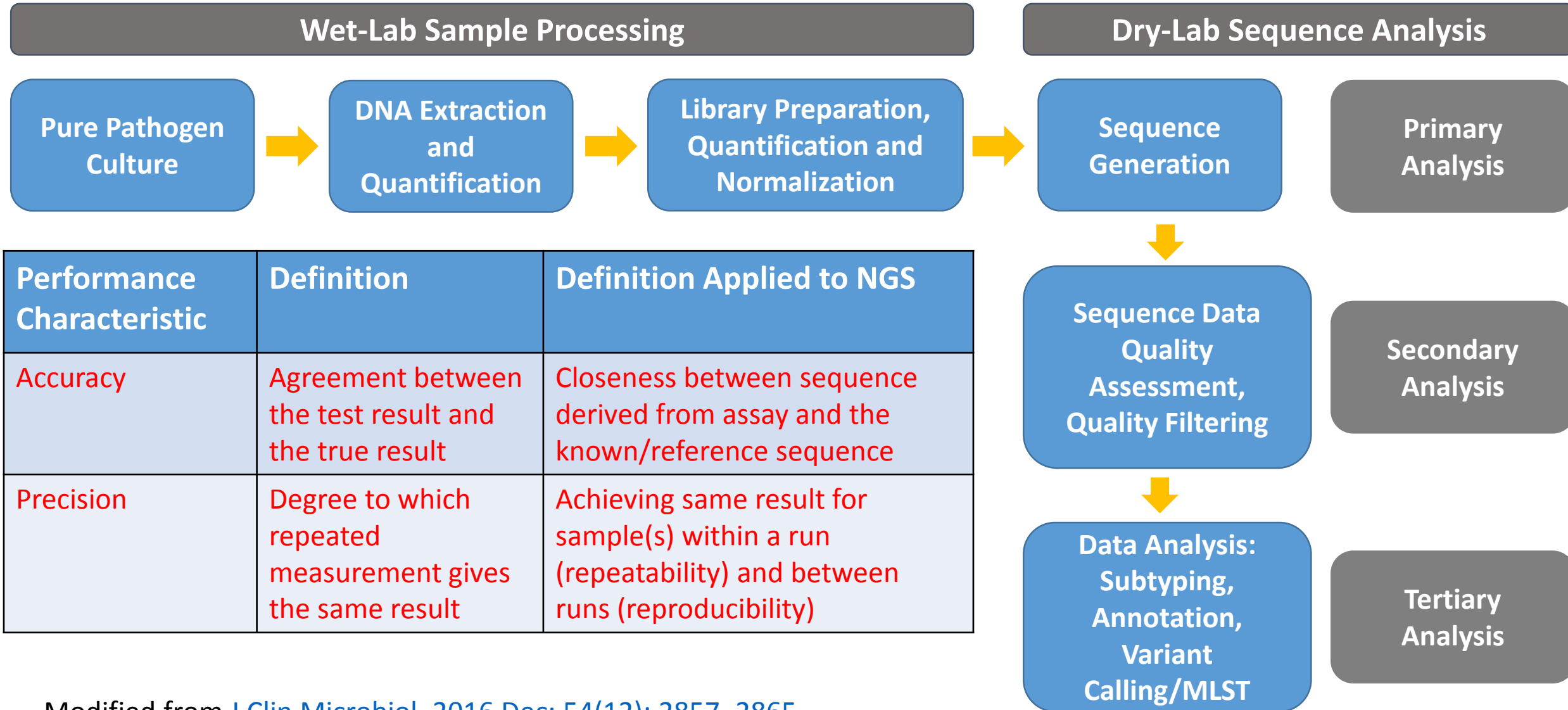
	2013	2014	2015	2016	2017	2018
<b>Programs</b>	<ul style="list-style-type: none"> <li>• <b>GenomeTrakr</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>GenomeTrakr</b></li> <li>• <b>PulseNet</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>GenomeTrakr</b></li> <li>• <b>PulseNet</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>GenomeTrakr</b></li> <li>• <b>PulseNet</b></li> <li>• <b>Bioinformatics</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>GenomeTrakr</b></li> <li>• <b>PulseNet</b></li> <li>• <b>TB</b></li> <li>• <b>Bioinformatics</b></li> </ul>	<ul style="list-style-type: none"> <li>• <b>GenomeTrakr</b></li> <li>• <b>PulseNet</b></li> <li>• <b>TB</b></li> <li>• <b>Bioinformatics</b></li> </ul>
<b>Samples Sequenced</b>	12	498	700	965	2,048	860 (as of 5/31)
<b>Instruments</b>						
<b>Testing Scientists</b>						
<b>Bioinformatics Scientists</b>						

# New Needs with the Expansion of NGS Operations

- Defining testing process and procedures
- Consistent test performance
- Management of samples across laboratory sections
- Management of reagents and instrument performance
- Verifying run information prior to sharing results



# NGS Workflow Components



Modified from [J Clin Microbiol. 2016 Dec; 54\(12\): 2857–2865.](#)

# DCLS' WGS Method and Data Acceptability Validation

- **Method-based validation of the entire wet-lab process:**
  - Bacterial culture
  - DNA extraction: Manual and QiaCube automated platform
  - DNA quantification
  - Library preparation, quantification and normalization
  - Pooling and massively parallel sequencing on the Illumina MiSeq
- **Performance characteristics:**
  - Internal assessment:
    - Run metrics: Clusters passing filter and Q30 score, FastQC checks
    - Estimated coverage meets external partner criteria
  - External assessment:
    - FDA: CFSAN SNP acceptability of validation panel samples re-sequenced by new method
    - CDC: SNP comparison of validation samples sequenced in parallel
    - Criteria: Coverage, Q-score or % mapped reads, SNPs

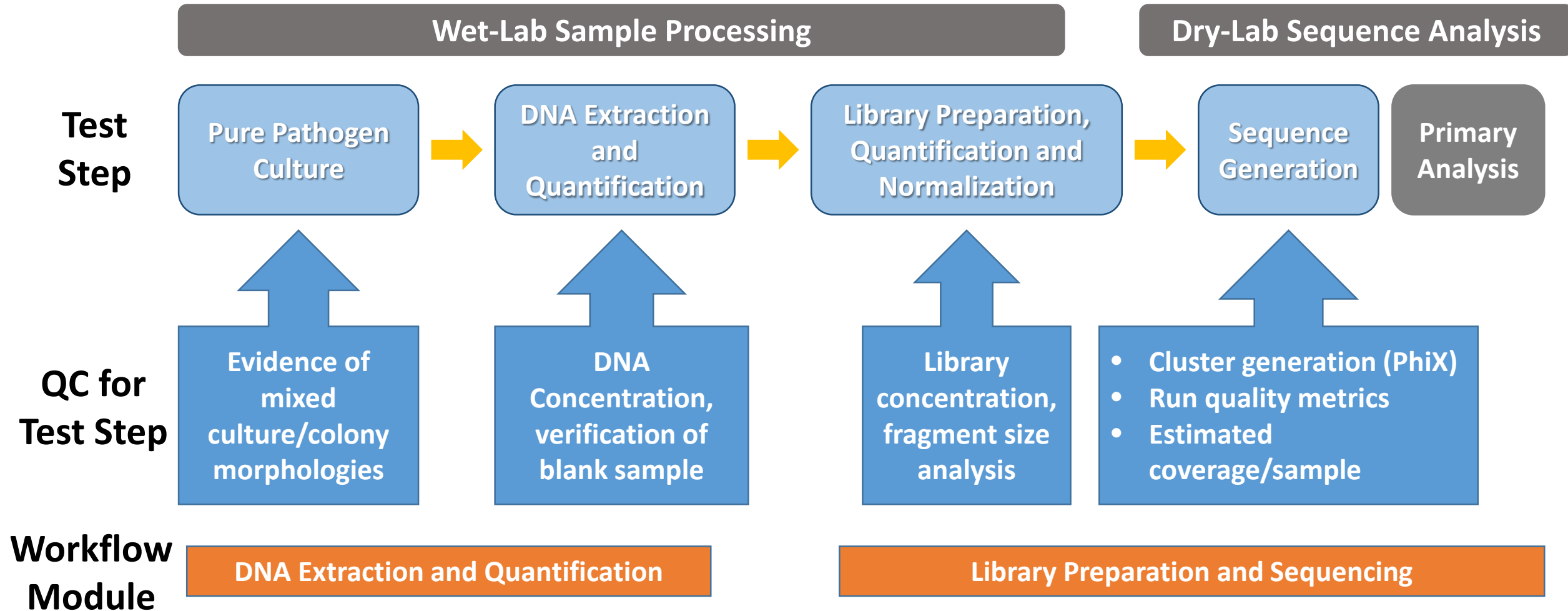
# Method Performance and Acceptability

- **Validation sample set:** Sample diversity representative of intended use
- **Accuracy:**

Organism	<i>N</i>	Genome Size	Coverage Requirement	% Isolates Meeting Coverage Requirement	Allowable SNPs	% Isolates Meeting Allowed SNP Threshold
<i>Salmonella</i>	6	5 MBp	>30	100	5	100
STEC	17	5 MBp	>40	100	5	100
<i>Campylobacter</i>	10	1.6 MBp	>20	100	<2	100
<i>Listeria</i>	6	3 MBp	>20	100	3	100

- **Precision/Reproducibility:**
  - 100% of 39 isolates tested in parallel yielded sequences with adequate coverage and sequence similarity

# NGS Wet-Lab QC



# NGS Wet-Lab Testing QA Documents

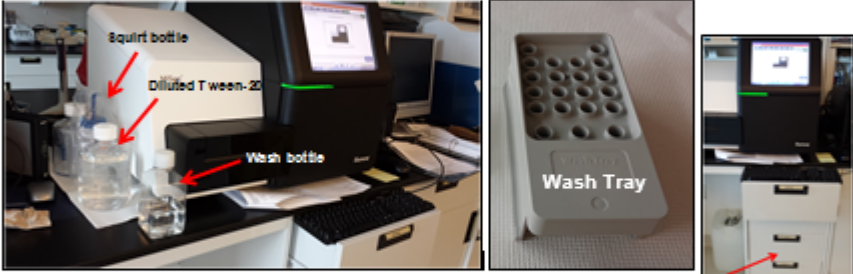
- 2 SOPs (DNA extraction for TB; WGS method)
- 10 Testing worksheets
- 7 Job aides (bench guides for quick reference)
- 4 Logs:
  - Sample management
  - Index rotation
  - Reagent inventories
  - Instrument maintenance
- 4 Training and competency checklists
- 4 Instructive documents:
  - Training maps: Extraction and library preparation
  - Second review of NGS data packets review
  - Sample submission instructions

DGS-XXXXX MiSeq post-run wash job aid 062515


A post-run wash is performed once the run is complete. The hands-on time is ~5 minutes and the run is ~20 minutes.

- Supplies and equipment**
  - (1) Wash tray
  - (2) Squirt bottle
  - (3) 10% Tween-20 solution in MilliQ water
  - (4) Wash bottle
  - (5) Wash buffer: 0.5% Tween-20 solution in MilliQ water. If more buffer is necessary, prepare 1L by adding 50 mL of 10% Tween-20 into 950 mL of MilliQ water. Prepared fresh immediately before use.

Item (1) is in the drawer under the MiSeq, and Item (2) to (5) are on the shelf by the MiSeq. The WGS team will supply 10% Tween-20 and MilliQ water.



- starting a Post-run wash**
  - 2.1 When a run completes, it is indicated on the screen. Click "Next". Then click "Start Wash".





# Sample Management: Electronic Logs

B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	T
Organism	Serotype	Outbreak Code	Source type	WGS_ID	request receive date	DNA extraction date	Freezer location	Problems	DNA quant. value	Index 1	Index 2	Library Date	MiSeq run date	MiSeq Run #	Data share date
Salmonella	Enteritidis			PNUSAS036943		3/14/2018			28.1	N704	S502	4/10/2018	4/11/2018	VA-M04590-180411	4/17/2018
E. coli	non-O157			PNUSAE012698		3/14/2018		Library prep failed; Qubit value <0.50	59.3	N705	S502	4/10/2018	4/11/2018	VA-M04590-180411	N/A
E. coli	non-O157			PNUSAE012699		3/14/2018			58.9	N706	S502	4/10/2018	4/11/2018	VA-M04590-180411	4/17/2018
Salmonella	Enteritidis			PNUSAS037070		3/19/2018			23.6	N701	S503	4/12/2018	4/13/2018	VA-M04590-180413	4/17/2018
Salmonella	Enteritidis	SP#17259		PNUSAS037072		3/19/2018	32-H1		31.5	N705	S503	3/27/2018	3/28/2018	VA-M04590-180328	4/3/2018
Salmonella	Enteritidis	VA cluster		PNUSAS037073		3/19/2018	32-H2		29.8	N706	S503	3/27/2018	3/28/2018	VA-M04590-180328	4/3/2018
Salmonella	Enteritidis	VA cluster		PNUSAS037074		3/19/2018	32-H3		52	N701	S504	3/27/2018	3/28/2018	VA-M04590-180328	4/3/2018
Salmonella	Braenderup	1802MLJBP-1; SP# 17139		PNUSAS037075		3/19/2018	32-D2		16.7	N704	S517	3/20/2018	3/21/2018	VA-M04590-180321	3/26/2018
Salmonella	Enteritidis	SP#17259		PNUSAS037076		3/19/2018	32-H4		54.4	N702	S504	3/27/2018	3/28/2018	VA-M04590-180328	4/3/2018
Salmonella	Braenderup			PNUSAS037077		3/19/2018			5.34	N702	S503	4/12/2018	4/13/2018	VA-M04590-180413	4/17/2018
Salmonella	Enteritidis	KY cluster; SP# 17293		PNUSAS037078		3/19/2018	33-F1		27.1	N705	S503	4/5/2018	4/6/2018	VA-M04590-180406	4/16/2018
Salmonella	Enteritidis			PNUSAS037079		3/19/2018			32	N703	S503	4/12/2018	4/13/2018	VA-M04590-180413	4/17/2018
Salmonella	Typhimurium			PNUSAS037071		3/19/2018			33.8	N704	S503	4/12/2018	4/13/2018	VA-M04590-180413	4/17/2018
E. coli O157	O157			PNUSAE012716		3/19/2018			36.4	N705	S503	4/12/2018	4/13/2018	VA-M04590-180413	4/17/2018
Salmonella	Salmonella sp			PNUSAS037080		3/19/2018			8.98	N706	S503	4/12/2018	4/13/2018	VA-M04590-180413	4/17/2018
Salmonella	Typhimurium			PNUSAS037081		3/19/2018			25.8	N707	S503	4/12/2018	4/13/2018	VA-M04590-180413	4/17/2018
E. coli	non-O157			PNUSAE012717		3/19/2018			27.7	N708	S503	4/12/2018	4/13/2018	VA-M04590-180413	4/17/2018

## 1 Pre-analytical phase

- Sample prioritization
- Record of identifiers
- Availability of samples for library preparation
- Notation of issues and reordering samples
- Real-time record of test data for cross-reference against testing worksheets

## 2 DNA extraction analytical phase

## 3 Library preparation and sequencing analytical phase

# Reagent Management and Instrument Operability

- Monthly reagent inventory checks and ordering
- Logs for receipt of reagents and lot numbers, shipping inconsistencies and assignment of identifiers for “kit-ifying”
- Instrument maintenance logs for daily, with run, monthly and semiannual or annual maintenance tasks
- Troubleshooting logs
- Instrument service agreements and manufacturer preventive maintenance

Item	Storage Temp	Cat #	Supplier	Min. 1 month supply	Notes
Microseal B Adhesive Seals (100)	RT	MSB1001	BioRad	1 pack	
Microseal F Foil (100)	RT	MSF1001	BioRad	1 pack	
Lense paper (12 packs of 100 sheets)	RT	11-995	Fisher Scientific	1/2 pack	
1ml serological pipettes (50)	RT	S68228A	Fisher Scientific	25 pipettes	
NanoDrop CF-1 Calibration Fluid	RT	CF-1	Fisher Scientific	1 kit	
DNA Clean & Concentrator-5	RT	D4013	Zymo Research	1 kit	
MiSeq Disposable Wash Tubes (20 tubes)	RT	MS-102-9999	Illumina	1 bag	
EBT (4.8 mL)	RT	in PulseNet kit	Illumina	TBD	
AMPure beads 60 mL	4 °C	A63881	Beckman Coulter	15 mL	
Qubit dsDNA HS Kit (500 assays)	4 °C	Q32854	Life Technologies	1 kit	
NT Buffer in Nextera XT PulseNet Kit or Nextera XT DNA Library Preparation Kit Box 2 of 2	RT/4 °C	FC-133-1001 or FC-131-1024 (24-sample)	Illumina	TBD	
MiSeq Reagent Kit v2 Box 2 of 2 (300/500 cycle)	4 °C	MS-102-2002 for 300, MS-102-2003 for 500	Illumina	12	
Nextera XT DNA Library Preparation Kit Box 1 of 2 (24 samples) or PulseNet Kit (96 samples)	-20 °C	FC-131-1024 (24) OR FC-133-1001(96)	Illumina	8x 24-sample kits OR 2x 96-sample kits	
Nextera XT Index Kit 24 indexes (96 samples) or 96 indexes (192 samples)	-20 °C	FC-131-1001(24) or FC-131-1002(96) or FC-131-	Illumina	2x 24-index kits or 1x 96-index kit	
MiSeq Reagent Kit v2 Box 1 of 2 (500 cycle)	-20 °C	MS-102-2003	Illumina	12	
MiSeq Reagent Kit v2 Box 1 of 2 (300 cycle)	-20 °C	MS-102-2002	Illumina	TBD	
PhiX Control v3 (10 nM, 10 uL)	-20 °C	FC-110-3001	Illumina	1 vial	
Sodium Hydroxide aliquotes	-20 °C	(72082-100ML)	(Sigma) in-house	1/2 rack equivalent	
Index replacement caps	RT	15026762	Illumina	1	



# Method Proficiency

- **Conceptual training:**

- Introductory webinars (WY/CO series)
- Illumina training videos: <https://support.illumina.com/training.html>
- Program-specific presentations

- **Hands-on training**

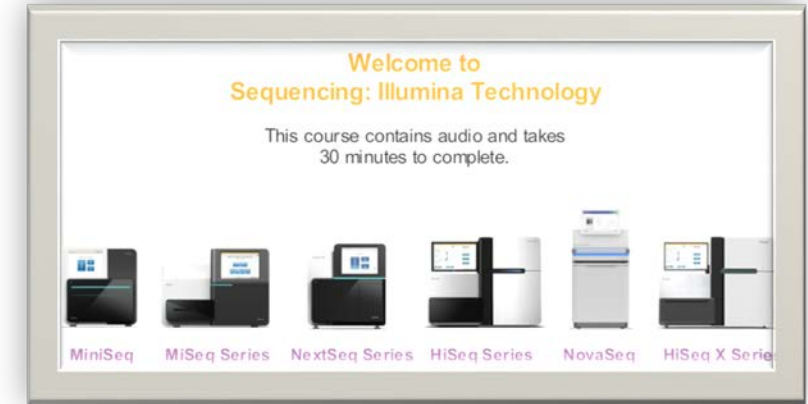
- Observation of experienced scientist(s)
- Observation by experienced scientist(s)
- Independent runs
- Final observed testing run

- **Ongoing competency:**

- Previous PT sample analysis and review by external partner

- **External Proficiency Assessment:**

- PulseNet/GenomeTrakr Harmonized PT Challenge (Annual)
- Global Microbial Identifier: wet-lab and informatics options (Annual)



 DCLS MDC	<b>Name:</b>					
	<b>Position:</b>					
	<b>Evaluation Cycle:</b>		<input type="checkbox"/> Initial Training and Competency			
	<b>Work Group:</b>		MDC- Whole Genome Sequencing by the Nextera XT Library Prep and Illumina MiSeq			
<b>Group Manager:</b>		Sean Kelly				
<b>WGS by Nextera XT and MiSeq</b>		<b>Areas of Assessment for competency:</b> ● Direct Observation (routine test performance) ● Document Review (reporting of results) ● Document Review (worksheets, QC, logs) ● Direct Observation (maintenance or function checks) ● Assessment of Test Performance (blind samples) ● Assessment of Problem Solving Skills (Q&A/troubleshooting)				
<b>Safety</b>		<b>Training Date</b>	<b>Trainer &amp; trainee initials</b>	<b>Assessment Tool/ Area</b>	<b>Assessment Date</b>	<b>Assessor Initials</b>
Read reference procedures and relevant DCLS SOPs		Evaluator Date and Initials:				
Completion Date:						
Read and review: Equipment manuals and manufacturer product inserts:		Evaluator Date and Initials:				
Completion Date:						
Review SDS for procedure, location MDC SDS notebooks.						●
Review precautions and waste management for the chemicals in the procedure						●
Review biological hazards associated with procedure, waste management for biological hazardous and non-hazardous waste.						●
Proper PPE and appropriate glove usage						●
<b>Overview WGS</b>						
Review WGS workflow, kit manuals, project presentations and Illumina videos.						●
Review organisms tested by WGS and DCLS points of contact to obtain cultures.						●

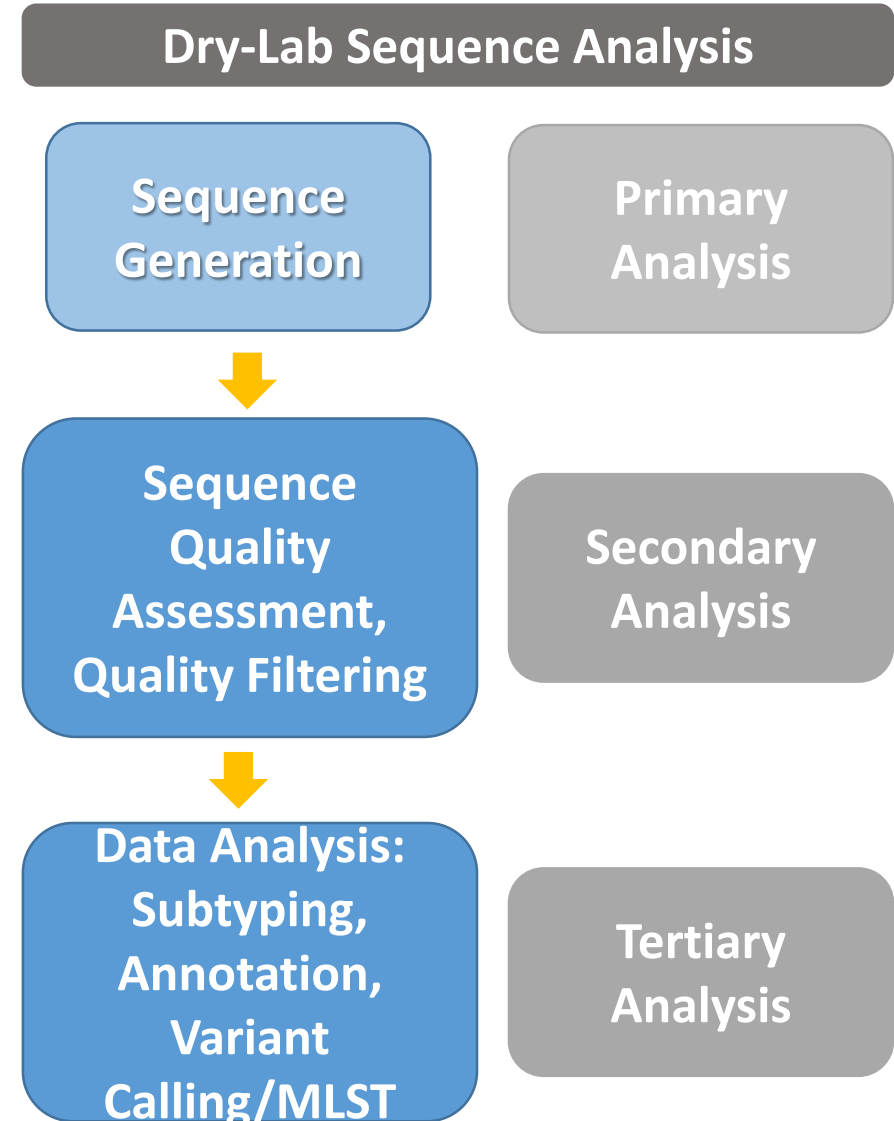
# Good Laboratory Practices for NGS

- ✓ Routine decontamination of equipment and surfaces
- ✓ Bench protectors
- ✓ Change gloves regularly
- ✓ Pre-amplification and post-amplification dedicated pipettes, tips, tubes, and ancillary consumables
- ✓ Define separate pre- and post-amplification workspaces
- ✓ Provide training and ongoing assessment of good pipetting practices



# Next Phase: NGS Data Analysis

- **Secondary Analysis:**
  - Data quality: Error rate (Q-scores) per sample
  - Taxonomic check: Organism verification
- **Tertiary Analysis:**
  - Reportable/actionable result
    - Subtype
    - Annotation and genotype
    - Variant analysis (relatedness in genomic epidemiology)



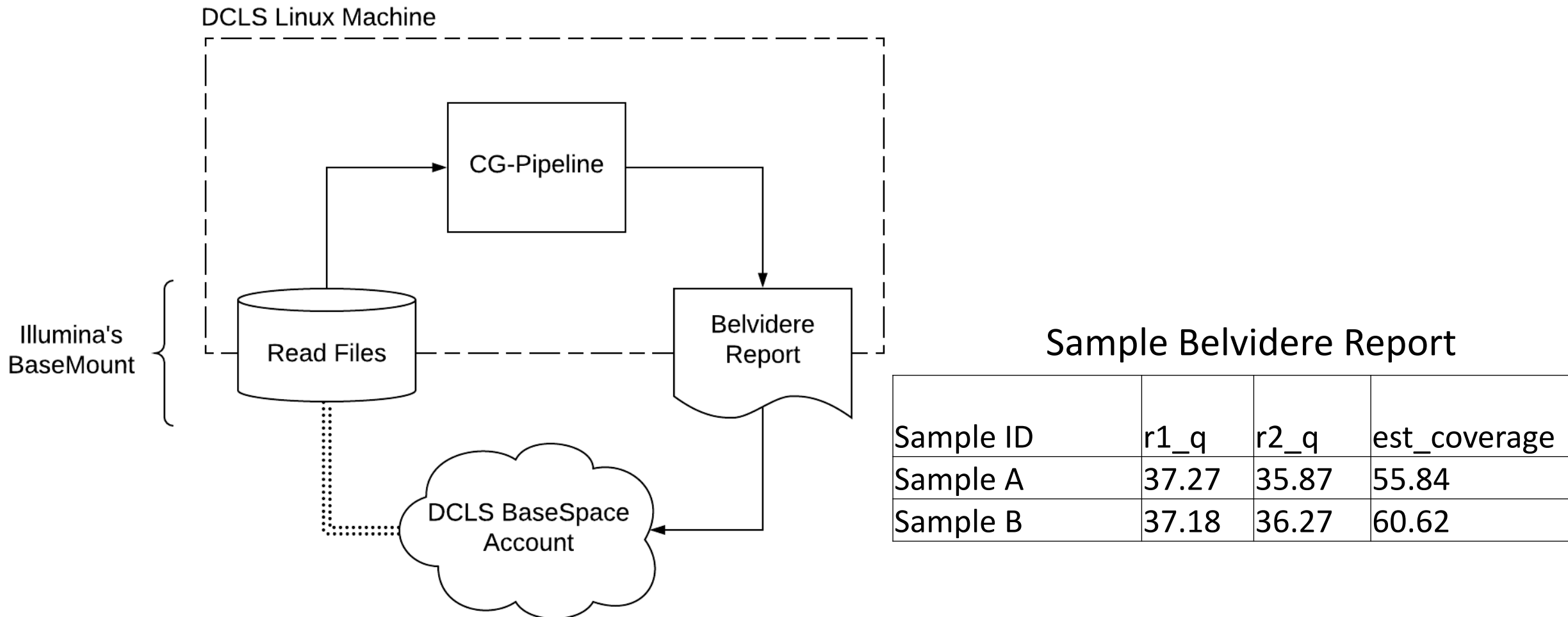
# Read Quality Assessment: Belvidere

- Ensuring sufficient read quality and sequencing coverage

## PulseNet Data Quality Criteria

	*Estimated Genome Coverage		
Quality of R1 and R2 reads (Avg)	<i>Listeria</i>	<i>Salmonella</i>	<i>E. coli</i>
30 - $\infty$	20x	30x	40x
29.0 - 29.99	30x	40x	50x
28.0 - 28.99	40x	50x	60x

# Read Quality Assessment: Belvidere



- Components of Belvidere: CDC's CG Pipeline and Illumina's Basemount
  - <https://github.com/lskatz/CG-Pipeline/>
  - <https://basemount.basespace.illumina.com/>

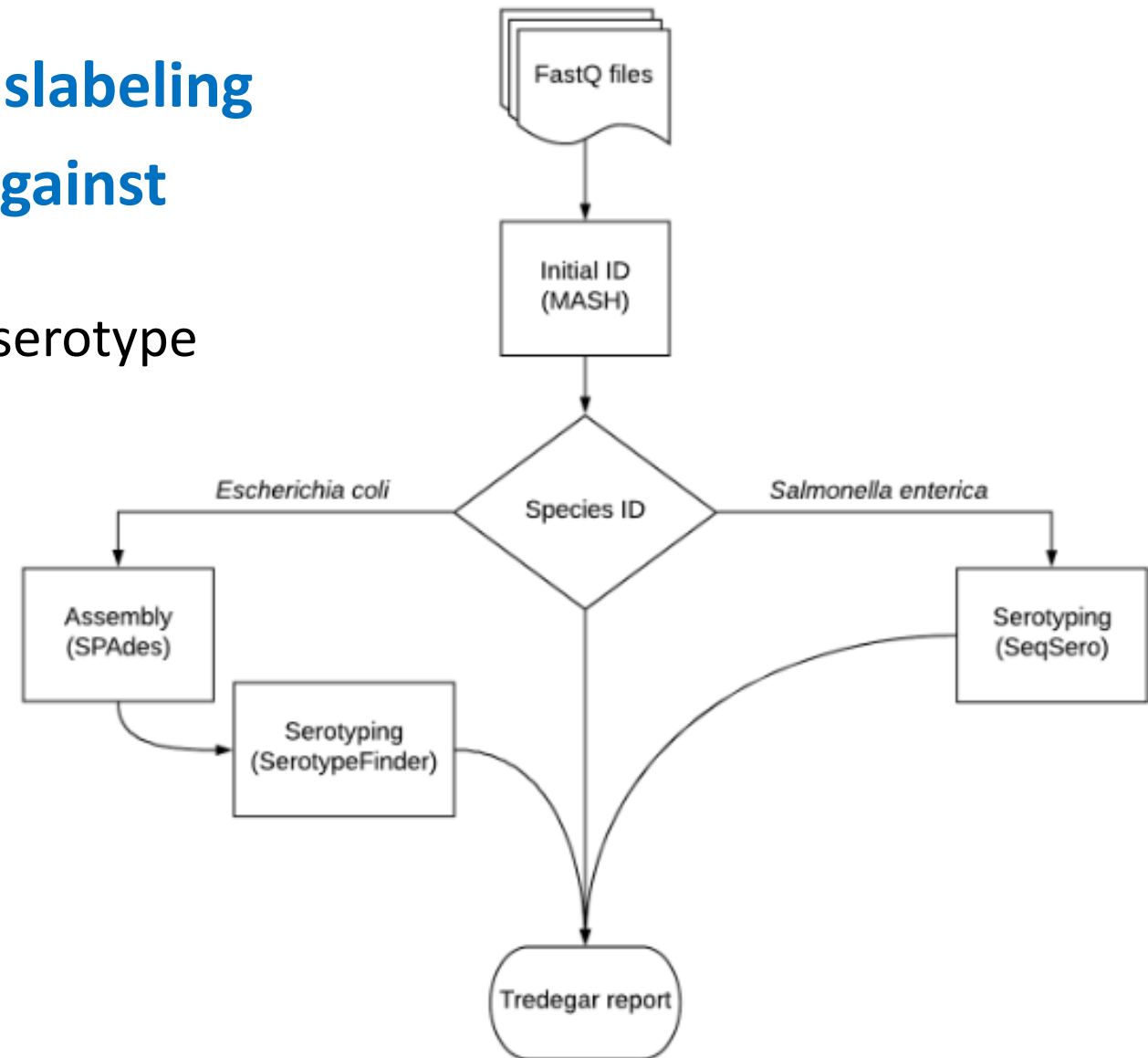
# Belvidere Validation

- Internal code review by second bioinformatics scientist
- Blinded sample set to assess pipeline accuracy and precision
  - 30 samples with 4 organisms (*Salmonella*, *Shigella*, STEC and *Listeria*) represented
  - Analyses run in replicate (4x) across 2 analysts



# Taxonomic Verification of Sequence Data: Tredegar

- Screening for isolate mix-up and mislabeling
- Species-level prediction by MASH against NCBI RefSeq Database
  - Shiga-toxin producing *E. coli* serotype prediction
  - *Salmonella* serotyping



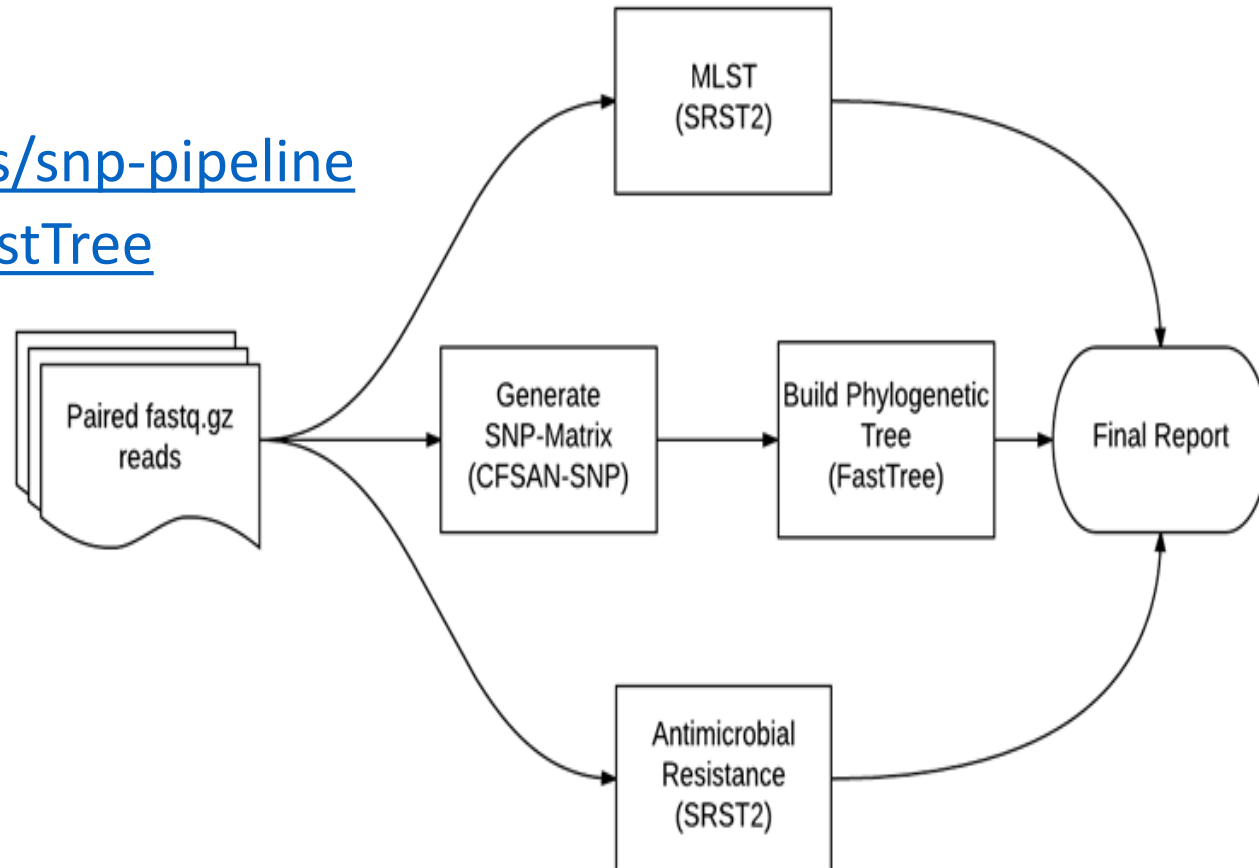
- Genome Biol. 2017 Jun 20;17(1):132. doi: 10.1186/s13059-016-0997-x.
- J Comput Biol. 2012 May; 19(5): 455–477. doi: 10.1089/cmb.2012.0021
- J Clin Microbiol. Aug; 53(8):2410-26. Doi: 10.1128/JCM.00008-15.
- J Clin Microbiol. May; 53(5):1685-92.

# Tredeggar Validation

- **Internal code review by second bioinformatics scientist**
- **Blinded sample set to assess pipeline accuracy and precision**
  - Challenge data set of 282 known isolates
  - Accuracy: Correct taxonomic, serologic ID
  - Precision: Multiple analyst comparison of dataset

# Subtyping and Genomic Epidemiology: GDIP

- Determining phylogenetic relationships during outbreak investigations
- Publically available: <https://github.com/dcls>
- Open source software and pipelines
  - <https://github.com/katholt/srst2>
  - <https://github.com/CFSAN-Biostatistics/snp-pipeline>
  - <https://github.com/PavelTorgashov/FastTree>



# GDIP Pipeline Validation

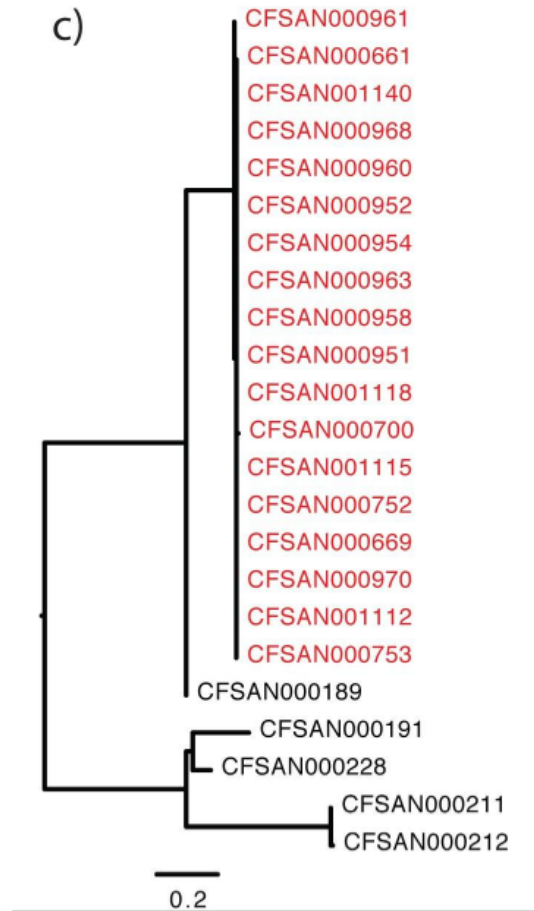
- **Evaluation of GDIP Pipeline output with**
  - Recapitulating known phylogenetic relationships
  - CDC/FDA Benchmark data sets



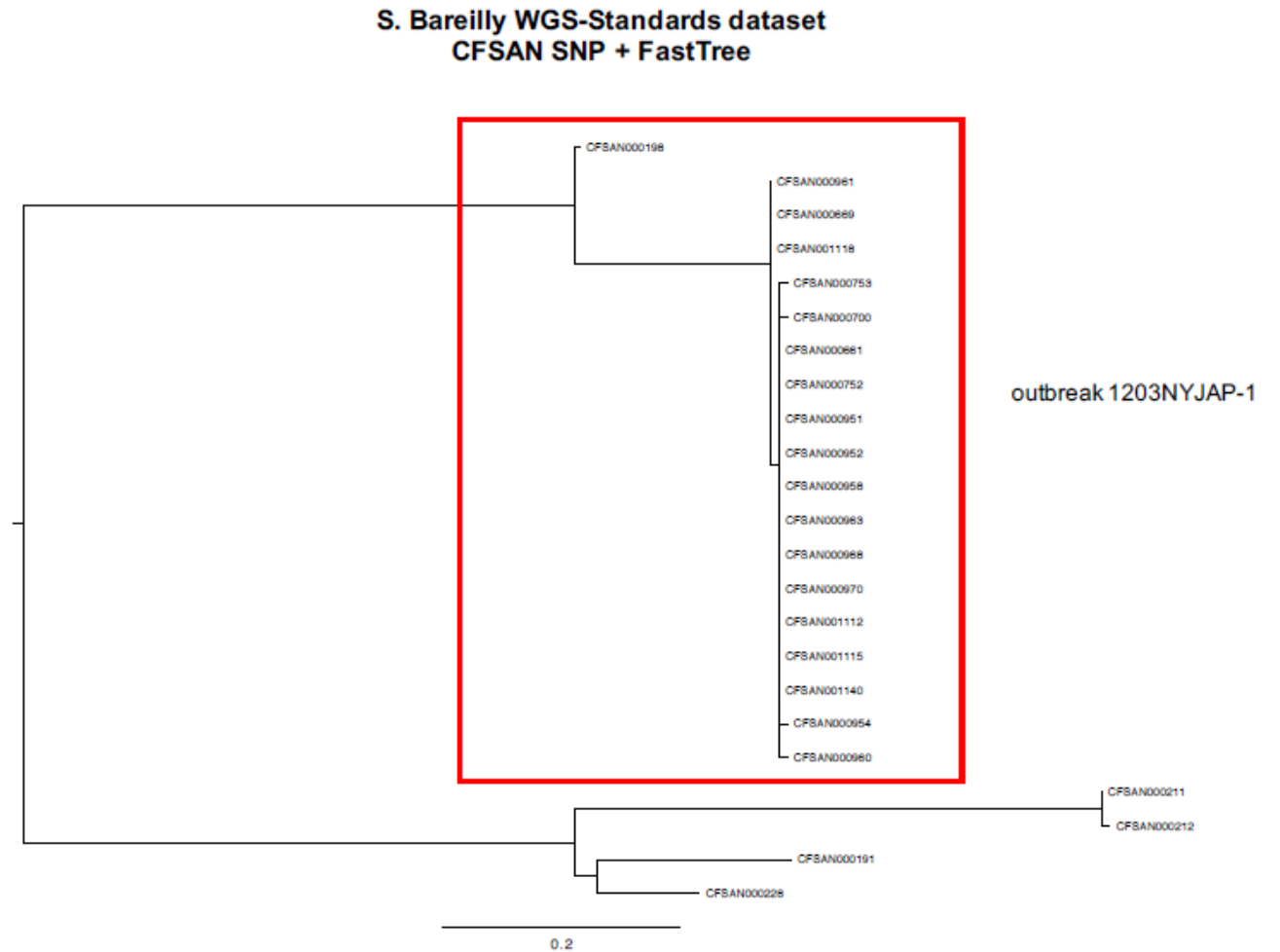
## Benchmark datasets for phylogenomic pipeline validation, applications for foodborne pathogen surveillance

Ruth E. Timme<sup>1</sup>, Hugh Rand<sup>1</sup>, Martin Shumway<sup>2</sup>, Eija K. Trees<sup>3</sup>, Mustafa Simmons<sup>4</sup>, Richa Agarwala<sup>2</sup>, Steven Davis<sup>1</sup>, Glenn E. Tillman<sup>4</sup>, Stephanie Defibaugh-Chavez<sup>5</sup>, Heather A. Carleton<sup>3</sup>, William A. Klimke<sup>2</sup> and Lee S. Katz<sup>3,6</sup>

- Salmonella dataset expected result:



- DCLS GDIP output:

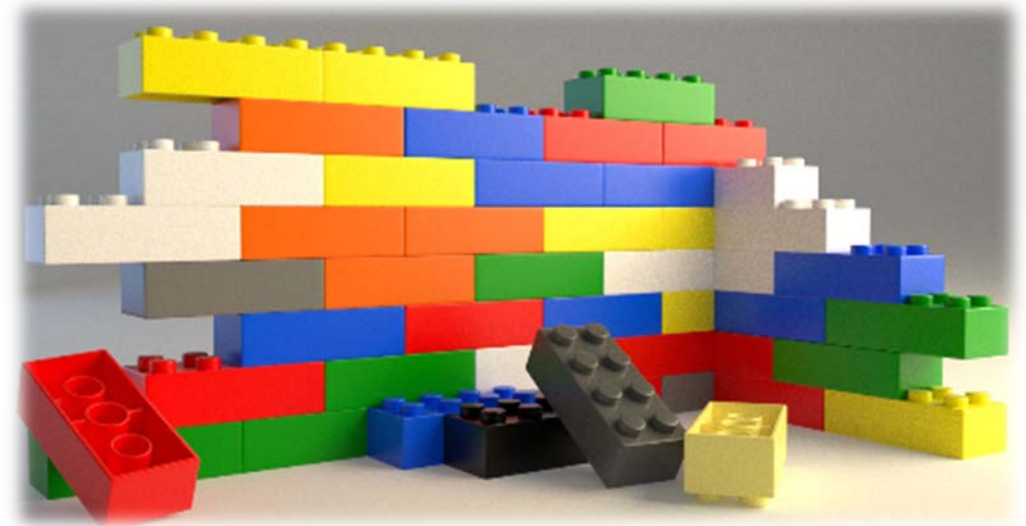


# Quality Considerations for Analytical Pipelines

- **Code review by second bioinformatics scientists**
- **Version control of validated pipeline**
  - “Locked down” locally, principal of least privileges
  - Version documentation (Github)
- **Usage documentation and result interpretation**
  - Code, SOPs, work guides as needed
- **Plan for verification of new pipeline versions**
- **Periodic checks of QC samples (lab determined or CRM)**
- **Reports/outputs that document pipeline/software versions**
- **Development of bioinformatics best practices, e.g. a “Command Tool Kit” with reusable code for common NGS analysis functions**

# Challenges to DCLS' NGS QC/QA Implementation

- Developed in the course of testing within a non-NGS QMS
- Institution of QA elements reactively versus proactively
- QA/QC guidelines in development
- Epidemiology → clinical purpose, will QC/QA fit?



# Solutions moving forward

- Evaluate positive and negative controls throughout the testing process
- Restructure electronic data management
- Reporting formats and disclaimers for internal use, epidemiology vs. clinical
- Continuing engagement with IT Groups



# Acknowledgements

- **DCLS Bioinformatics Scientists**
  - Kevin Libuit, APHL Bioinformatics Fellow
  - Gretchen Wilson
- **DCLS Molecular Detection and Characterization**
  - Sean Kelly
  - Leigh Emma Lion
  - Aki Harada
  - Arian Nassiri
  - MK Dickinson
- **DCLS Epidemiology Support Group**
  - Stephanie Dela Cruz
  - Francis Tannor
  - Shannon Levesque
  - Robert Barnes
- **CDC Office of Advanced Molecular Detection**
- **CDC PulseNet**
- **FDA GenomeTrakr**
- **CDC Division of TB Elimination**
- **APHL**
- **Federal grant support:**
  - CDC Epidemiology and Laboratory Capacity for Infectious Diseases: NU50CK000387-04-01
  - FDA ISO/IEC 17025 and WGS: 1U18FD006188-01