

Contaminated Sequencing Data

When there's *Vibrio* in your *Salmonella*

Kevin G. Libuit, M.S.

Mid-Atlantic Bioinformatics Support Lead Scientist

Division of Consolidated Laboratory Services

PulseNet Request for Resequencing

DCLS flagged for contaminated *Salmonella* sequences

Can you please resequence the following *Salmonella* isolates? I included VA in the email since they originally performed the sequencing.

PNUSAS043725	SC__SALM18-1249	Contains <i>Salmonella</i> and <i>Vibrio</i>
PNUSAS043726	SC__SALM18-1259	Contains <i>Salmonella</i> and <i>Vibrio</i>
PNUSAS043351	SC__SALM18-1269	Contains <i>Salmonella</i> and <i>Vibrio</i>

PulseNet's Contamination Verification Process

Calculation Engine metrics raised the first flag

Serotype	WGS_id	AvgQuality	AvgReadCoverage	N50	NrContigs	NrNonACGT	Length
Mbandaka	PNUSAS043725	34	76 446996	783	1706	5263695	
Mbandaka	PNUSAS043726	34	81 228724	1756	5556	6151418	
Heidelberg	PNUSAS043351	35	97 231793	2617	6306	7907721	

Threshold values
NrContigs: <400
Length: 4.4 - 5.1 Mbp

PulseNet's Contaminate Verification Process

Metagenomics tools (MIDAS & Kraken) confirmed presence and identity of non-target genomic content

Standard: Kraken Classification %	*Kalamari: Kraken Classification %	Passed Kraken?
3.51%(U), 96.49%(C), 84.38%(G, Salmonella), 3.37%(G, Escherichia), 1.23%(G, Klebsiella), 1.99%(G, Vibrio), 84.08%(S, Salmonella enterica), 3.25%(S, Escherichia coli), 1.23%(S, Klebsiella pneumoniae), 1.93%(S, Vibrio vulnificus)	3.23%(U), 96.77%(C), 78.84%(G, Salmonella), 9.22%(G, Escherichia), 2.57%(G, Vibrio), 78.23%(S, Salmonella enterica), 8.60%(S, Escherichia coli), 1.96%(S, Vibrio vulnificus)	Y
*MIDAS Species Hits [species_id: read count, coverage, rel_abundance]		Passed MIDAS?
Salmonella_enterica_58156: 2093, 55.28x, 88.35% Vibrio_vulnificus_58105: 64, 1.45x, 2.31% Salmonella_enterica_53987: 58, 1.25x, 2.00% Salmonella_enterica_58266: 211, 4.49x, 7.18% Salmonella_bongori_55351: 5, 0.09x, 0.15%		N

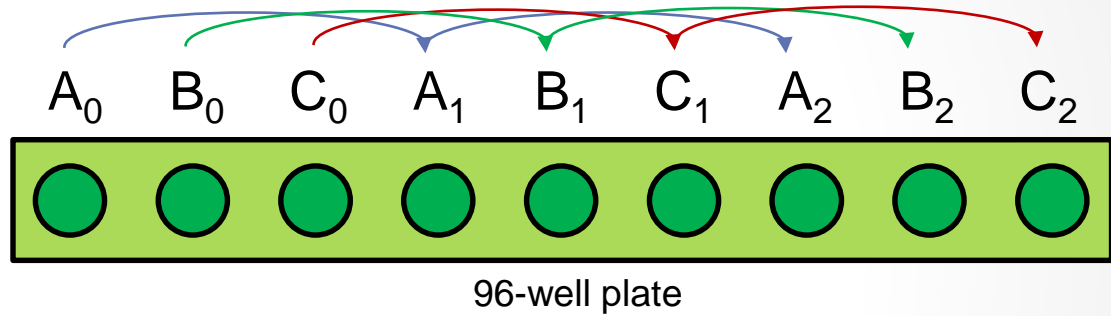
DCLS Response

1. Internal verification of CDC's results:
 - a) SPAdes, Quast (assembly metrics)
 - b) Kraken, MIDAS and MetaPhlan (contaminant ID)
2. Investigation of MiSeq Run
3. Review of wet-lab procedure to identify vulnerabilities
4. Corrective action to wet-lab protocol

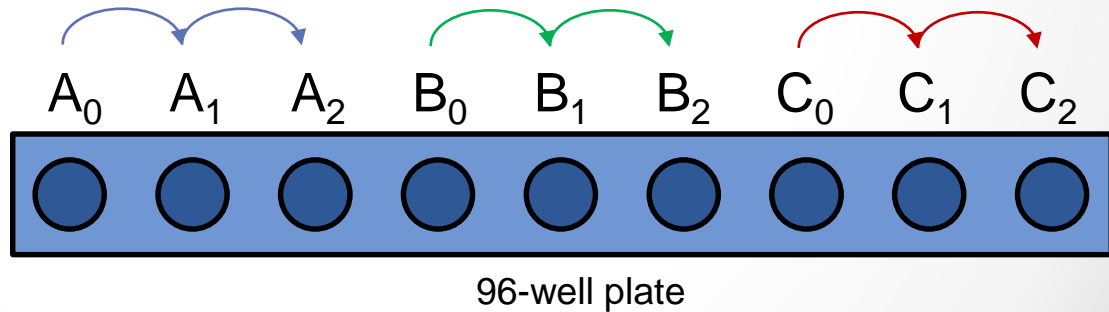


Library Prep – Serial Dilution Protocol

Old sample layout



Revised sample layout



Acknowledgements

- DCLS WGS team
- PulseNet Bioinformatics Support (Taylor Griswold)



Kevin G. Libuit, M.S.

Bioinformatics Regional Support Lead Scientist

Division of Consolidated Laboratory Services

Email: Kevin.Libuit@DGS.Virginia.gov