

You are what you poop: Testing pathogen detection sequencing methods and systems on wastewater samples

Eric L Vaughn, MPH

Co-author: Joel T Steyer, PhD; Janis Doss, PhD; Wadih Bchara, MB (ASCP)

DC Department of Forensic Sciences – Public Health Laboratory

APHL ID Con, Atlanta, GA

March 14, 2023

Outline

- Wastewater testing in DC
- SARS-CoV-2 sequencing
- Going beyond COVID
- Oxford Nanopore (ONT)
- Jumpcode
- Illumina Respiratory Viral Oligos Panel
- Epidemiological applications
- Limitations and lessons learned
- Into the future

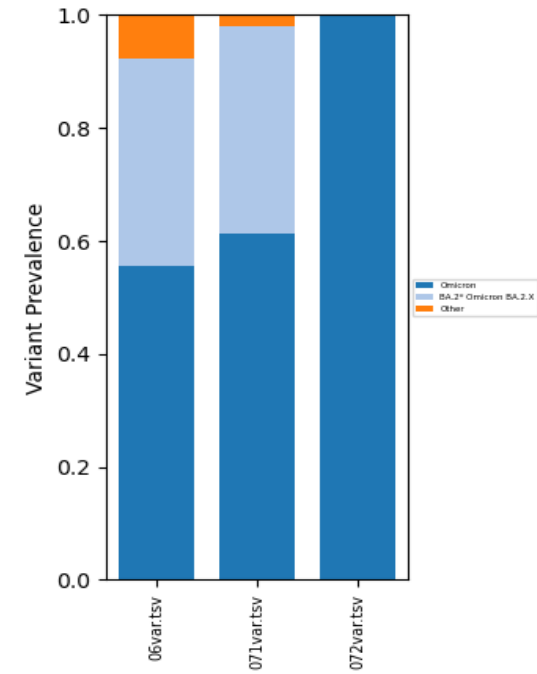
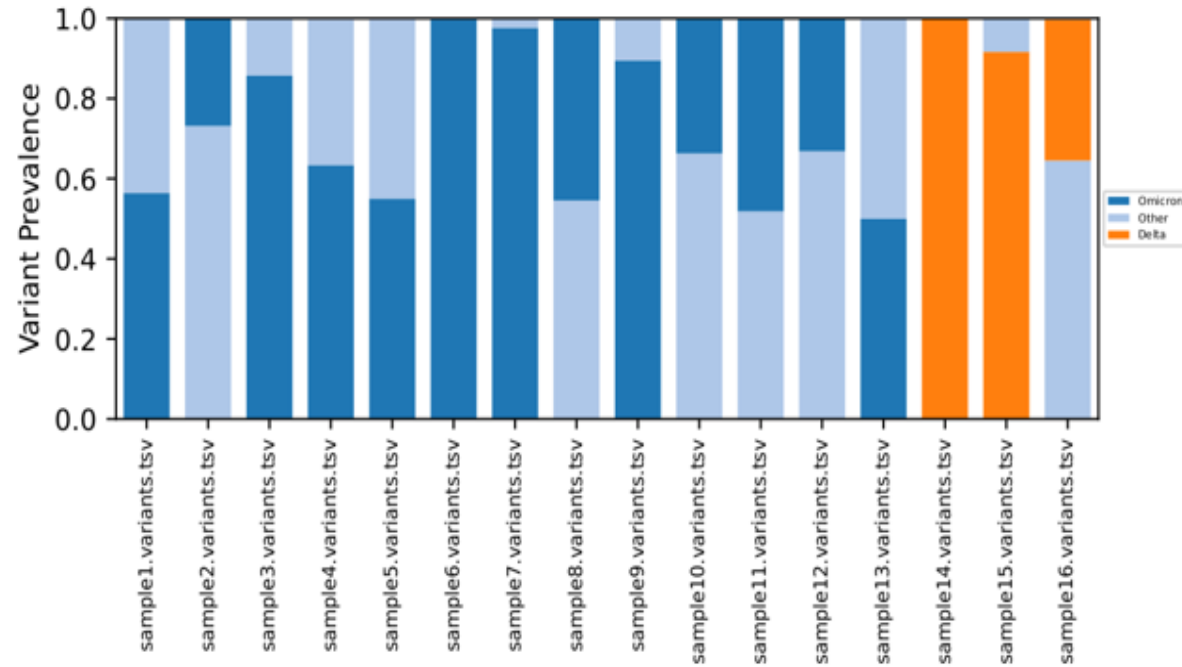
Wastewater testing in the District of Columbia

- Receive samples at the DFS Public Health Laboratory.
- RNA Extraction in the Molecular Diagnostics Unit (MDU).
- ddPCR testing in the MDU.
- Isolate proceeds to the NGS lab for sequencing.

SARS-CoV-2 in wastewater

- Receipt in NGS
- Sequenced with the same protocol for clinical isolates.
- cDNA synthesis
- ARTIC v4.1
- Illumina DNA Prep kit
- MiSeq Micro v2 kit

SARS-CoV-2 Sequencing Analysis



Going Beyond SARS-CoV-2

- What else is in the sample?
- Can we detect emerging infectious diseases?
- Can we detect foodborne organisms related to national and local clusters?
- What about overall surveillance?
- Does this include antimicrobial-resistance?

ONT metagenomic sequencing – methods and analysis

- RNA to cDNA synthesis
- Rapid sequencing kit
- MinION and GridION
- Epi2Me
- In-house data analysis pipeline

ONT metagenomic sequencing - AMR

Gene Clinically Relevant? ☐	Alignments	Average Accuracy	Taxon	CARD Model ?
				all models v
Neisseria meningitidis 16S rRNA mutation conferring resistance to spectinomycin	4786	87.2%	Neisseria meningitidis	rRNA mutation model
Helicobacter pylori 16S rRNA mutation conferring resistance to tetracycline	3537	85.9%	Helicobacter pylori 26695	rRNA mutation model
Salmonella enterica serovar Typhimurium 16S rRNA mutation in the rrsD gene conferring resistance to spectinomycin	2345	85.7%	Salmonella enterica subsp. salamae	rRNA mutation model
Neisseria gonorrhoeae 16S rRNA mutation conferring resistance to spectinomycin	1666	87.9%	Neisseria gonorrhoeae	rRNA mutation model
Pasteurella multocida 16S rRNA mutation conferring resistance to spectinomycin	840	86.6%	Pasteurella multocida 36950	rRNA mutation model
Escherichia coli 16S rRNA mutation in the rrsB gene conferring resistance to kanamycin A	388	85.6%	Escherichia coli K-12	rRNA mutation model
Escherichia coli 16S rRNA mutation in the rrsH gene conferring resistance to spectinomycin	388	85.1%	Escherichia coli K-12	rRNA mutation model
Escherichia coli 16S rRNA mutation conferring resistance to edeine	386	85.7%	Escherichia coli K-12	rRNA mutation model
Escherichia coli 16S rRNA mutation in the rrsB gene conferring resistance to neomycin	385	86.3%	Escherichia coli K-12	rRNA mutation model
Escherichia coli 16S rRNA mutation in the rrsB gene conferring resistance to spectinomycin	366	86.0%	Escherichia coli K-12	rRNA mutation model

Gene Clinically Relevant? ☐	Alignments	Average Accuracy	Taxon	CARD Model ?
				all models v
Mycobacterium tuberculosis 16S rRNA mutation conferring resistance to kanamycin	78	83.2%	Mycobacterium tuberculosis H37Rv	rRNA mutation model
Mycobacterium smegmatis 16S rRNA mutation in the rrsB gene conferring resistance to streptomycin	76	84.9%	Mycobacterium smegmatis str. MC2 155	rRNA mutation model
Mycobacterium smegmatis 16S rRNA mutation in the rrsB gene conferring resistance to neomycin	75	82.6%	Mycobacterium smegmatis str. MC2 155	rRNA mutation model
Mycobacterium tuberculosis 16S rRNA mutation conferring resistance to viomycin	72	83.9%	Mycobacterium tuberculosis H37Rv	rRNA mutation model
Mycobacterium smegmatis 16S rRNA mutation in the rrsB gene conferring resistance to hygromycin B	71	82.3%	Mycobacterium smegmatis str. MC2 155	rRNA mutation model
Mycobacterium smegmatis 16S rRNA mutation in the rrsB gene conferring resistance to viomycin	70	82.0%	Mycobacterium smegmatis str. MC2 155	rRNA mutation model
Mycobacterium smegmatis 16S rRNA mutation in the rrsB gene conferring resistance to kanamycin A	70	84.3%	Mycobacterium smegmatis str. MC2 155	rRNA mutation model
Mycobacterium tuberculosis 16S rRNA mutation conferring resistance to streptomycin	69	84.1%	Mycobacterium tuberculosis H37Rv	rRNA mutation model
mphG	17	93.1%	Photobacterium damsela subsp. damsela	protein homolog model
tetX	14	92.4%	Bacteroides fragilis	protein homolog model

ONT metagenomic sequencing - AMR

Gene Clinically Relevant? ☐	Alignments	Average Accuracy	Taxon	CARD Model [?]
				all models 
CblA-1	1	98.8%	mixed culture bacterium AX_gF3SD01_15	protein homolog model
tetQ	1	96.2%	Bacteroides fragilis	protein homolog model
aminocoumarin resistant alaS	1	73.5%	Escherichia coli str. K-12 substr. MG1655	protein homolog model
Mycobacterium leprae rpoB mutants conferring resistance to rifampicin	1	89.8%	Mycobacterium leprae TN	protein variant model
Escherichia coli EF-Tu mutants conferring resistance to Enacyloxin IIa	1	72.1%	Escherichia coli str. K-12 substr. MG1655	protein variant model
Mycobacterium tuberculosis rpoB mutants conferring resistance to rifampicin	1	72.1%	Mycobacterium tuberculosis H37Rv	protein variant model
Escherichia coli EF-Tu mutants conferring resistance to Pulvomycin	1	74.9%	Escherichia coli CFT073	protein variant model

ONT metagenomic sequencing – emerging infections



Krona Search:

30 Max depth
11 Font size
+ Chart size
☒ Collapse
Snapshot
Link
?

Nucleocytoviricota
Magnitude: 9
Unassigned magnitude: 0
Count: 33
Unassigned: 1
Taxon: 2733007
Rank: phylum

1% of Viruses
0.0008% of Root

Viruses
Root

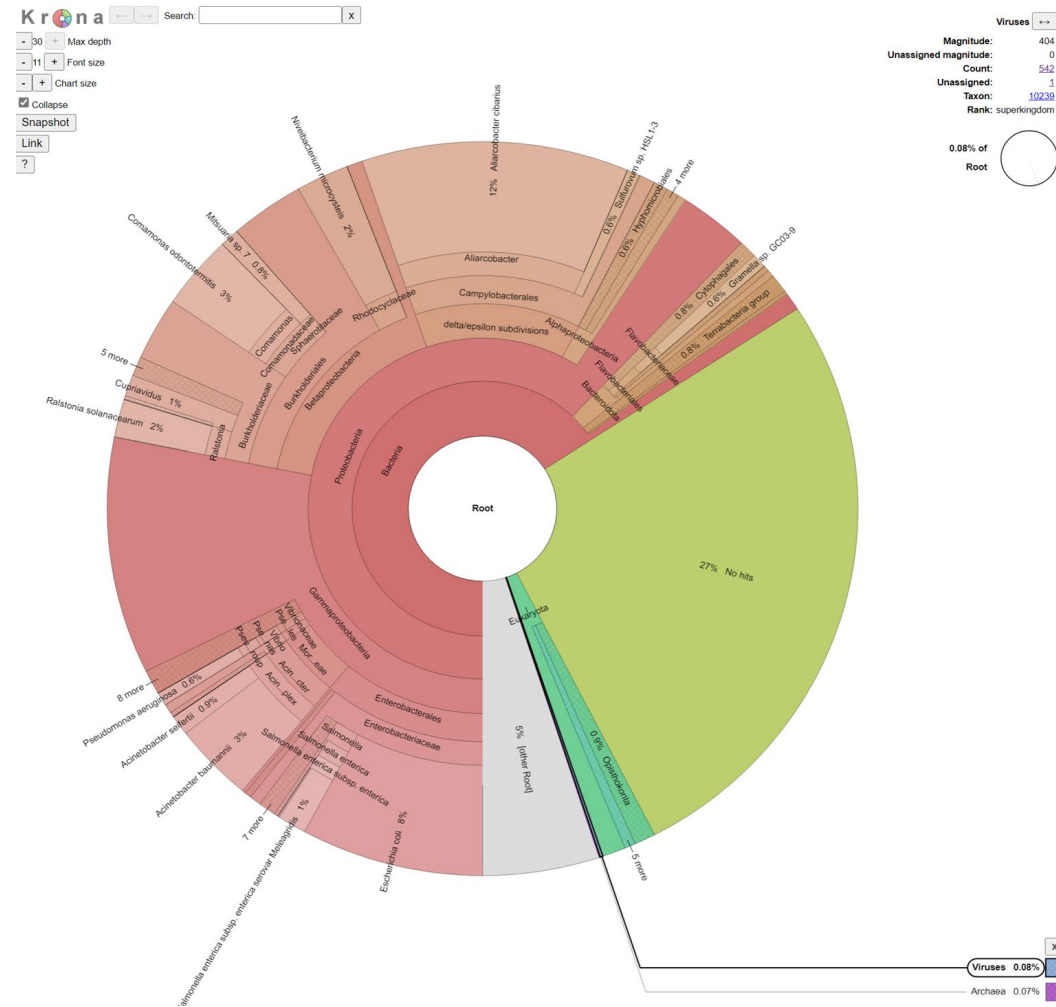
Tomato brown rugose fruit virus 36%
Tomato mosaic virus 15%
unclassified RNA viruses SH4-2016
Poviridae
Coronavirales
unclassified RNA viruses SH4-2016
Hepadnaviridae
Caudovirales
Herpesvirales
Nucleocytoviricota
unclassified Podoviridae
Bordetella phage BPP-1
Malvastrum yellow vein Chitwan betasatellite 0.1%
Monodnavirae 0.3%
Baculoviridae 0.5%
unclassified viruses 0.5%

2% Tomato mosaic virus
1% Cucumber green mottle mosaic virus
1% Pepper mild mottle virus
3 more
0.5% Baculoviridae
0.7% Propeovirales
0.1% Hantaan virus-like virus G1
0.1% Hantaan virus-like virus G2
0.5% Caliciviridae
1% unclassified Caliciviridae
0.9% Crass-like virus sh
7 more
3% unclassified Siphonviridae
0.8% Drillevirus
9 more
0.7% Septemtovirus kakheti25
1 more
1% Schottoviridae
2 more
1% Casjennaviridae
2 more
0.9% unclassified Podoviridae
0.7% unclassified Myoviridae
0.7% unclassified Myoviridae

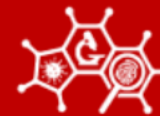
ONT metagenomic sequencing – emerging infections



ONT metagenomic sequencing – seasonal pathogens



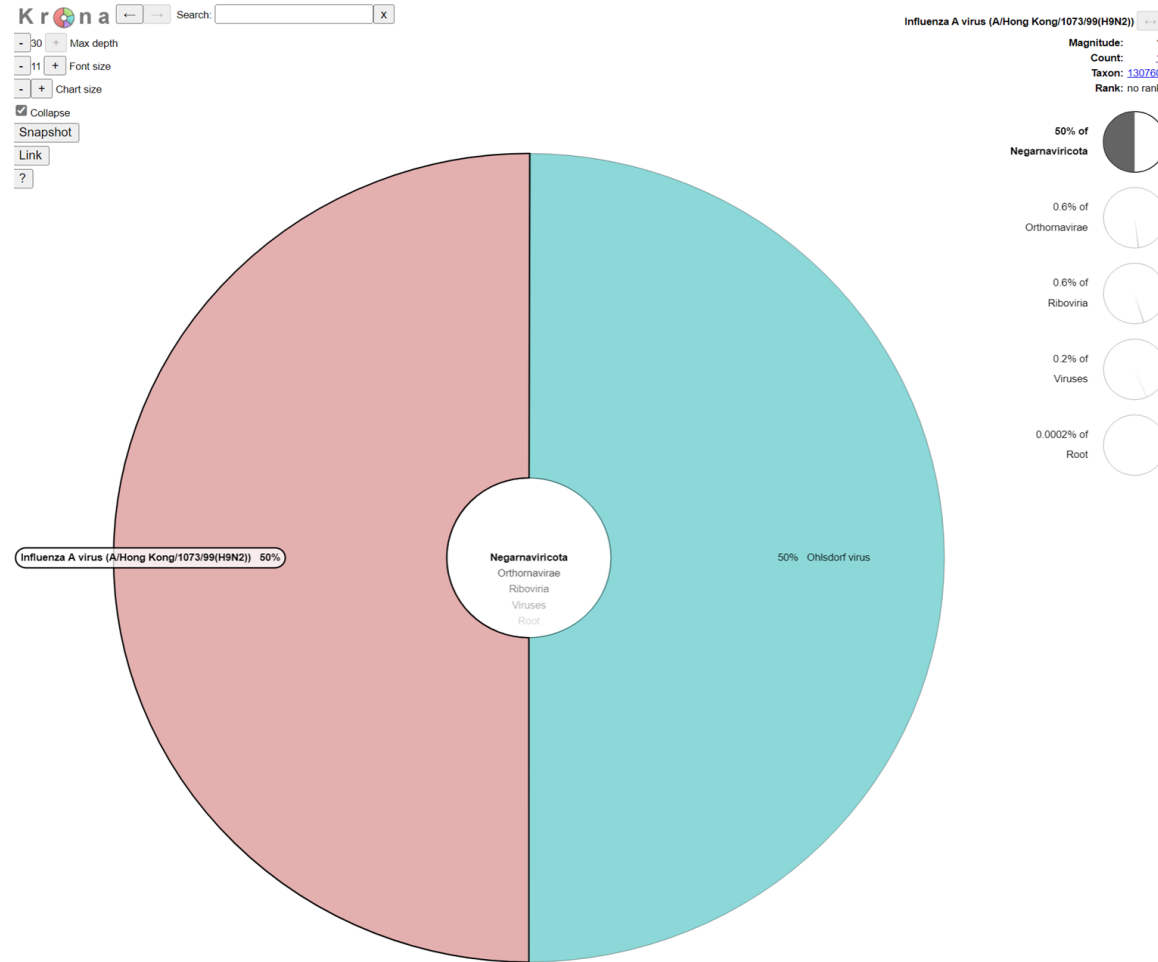
DFS.DC.GOV



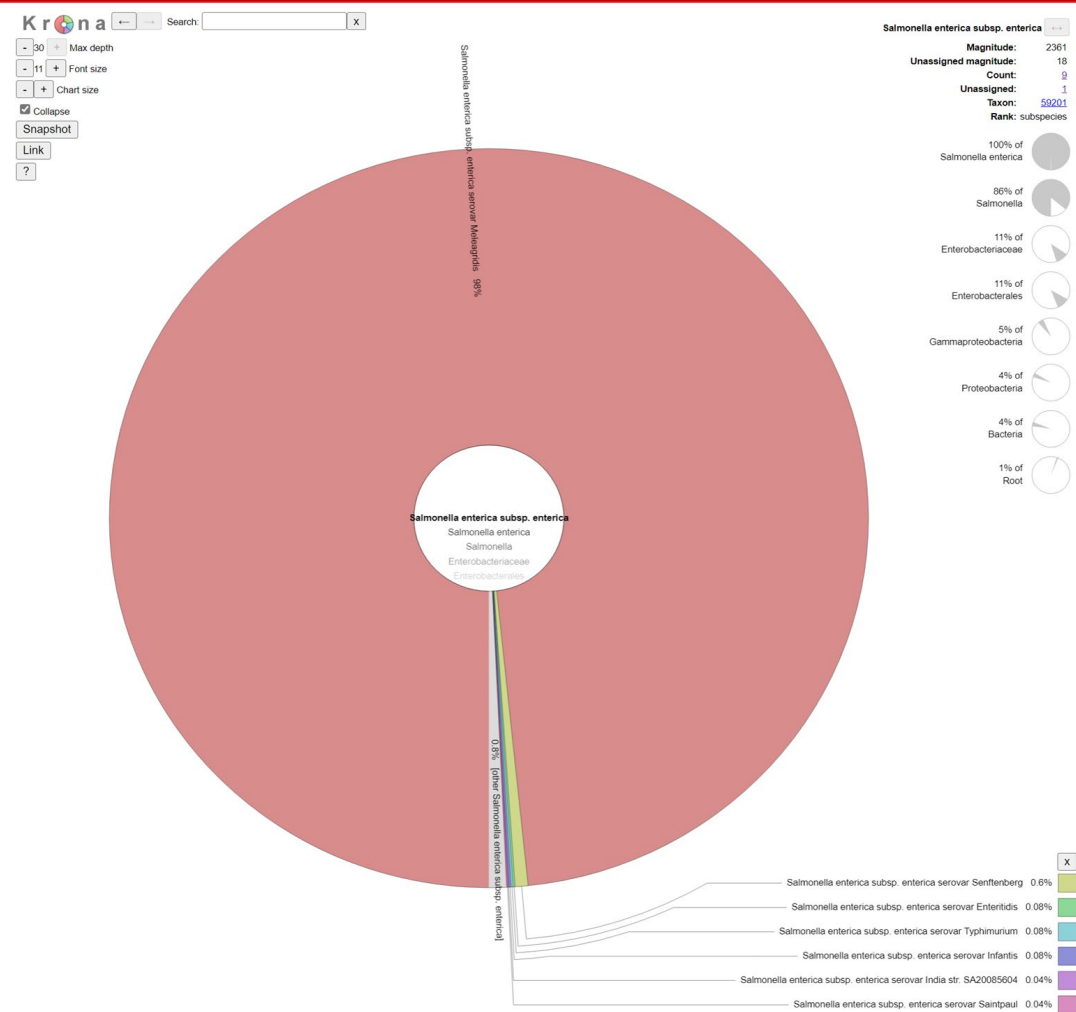
DISTRICT OF COLUMBIA
DFS
DEPARTMENT OF
FORENSIC SCIENCES

WE ARE WASHINGTON
GOVERNMENT OF THE DISTRICT OF COLUMBIA
DC MURIEL BOWSER, MAYOR

ONT metagenomic sequencing – seasonal pathogens



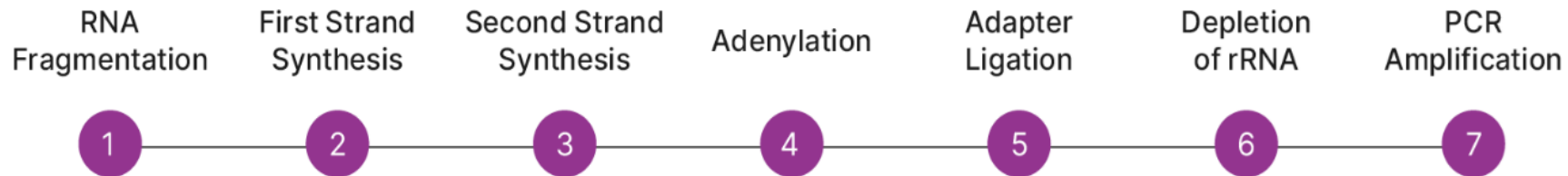
ONT metagenomic sequencing – other areas of interest



Jumpcode – CRISPRclean rRNA Kit

- Library prep kit designed to deplete human, mouse, rat, and pan bacteria genes.
- Ideal for samples containing eukaryotic and bacterial rRNA.
- Illumina MiSeq reagent v2 kit
- In-house data analysis pipeline

CRISPRclean Plus Workflow



🕒 Hands-on time: 3.5 Hours | Assay time: 9 Hours

Jumpcode: Analysis – non-depleted sample



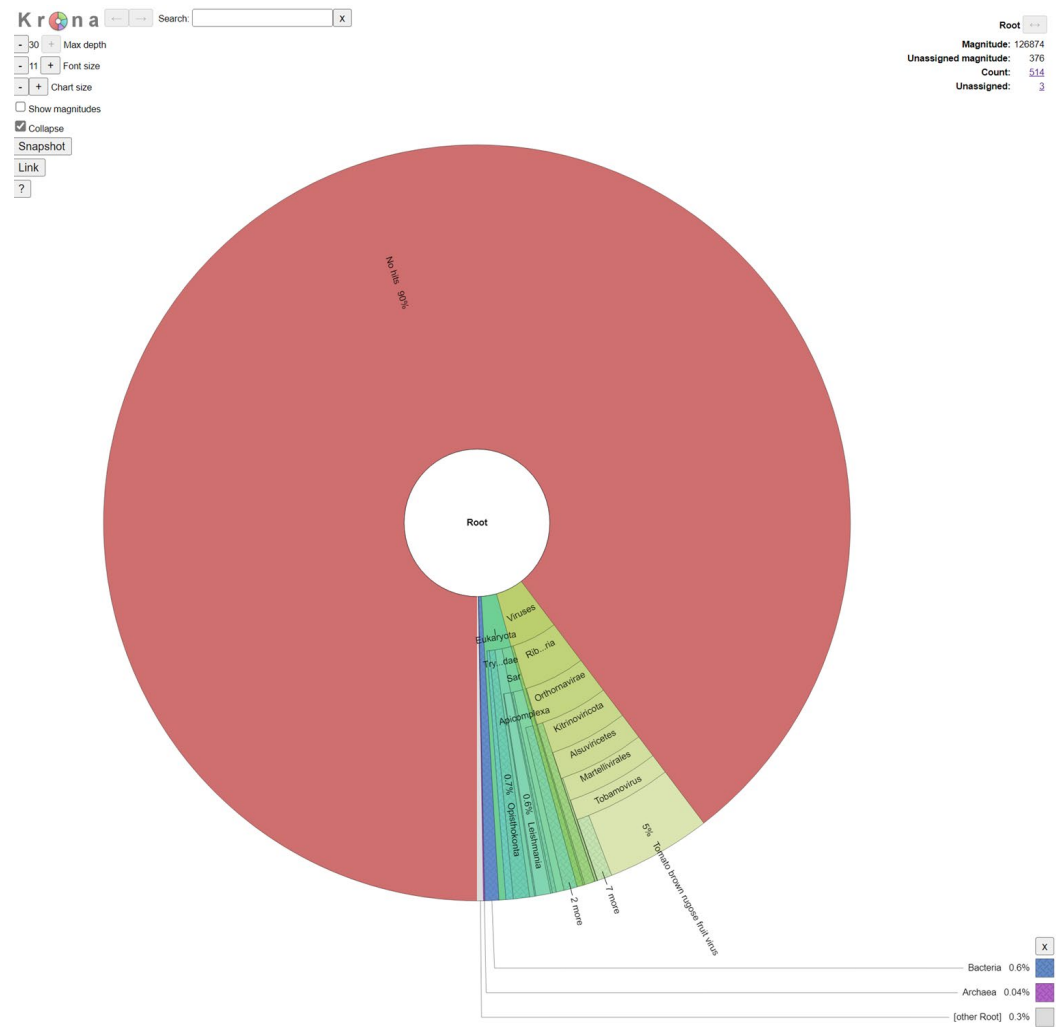
% of Root	Taxonomic Rank	Scientific Name
32.97	Unclassified	unclassified
67.03	Root	root
61.49	Root - 1	cellular organisms
58.87	Domain	Bacteria
2.31	Domain	Eukaryota
0.06	Domain	Archaea
0.08	Domain*	Viruses

Jumpcode: Analysis – human, mouse, rat only



% of Root	Taxonomic Rank	Scientific Name
27.66	Unclassified	unclassified
72.34	Root	root
70.47	Root - 1	cellular organisms
68.8	Domain	Bacteria
1.39	Domain	Eukaryota
0.15	Domain	Archaea
1.78	Domain*	Viruses

Jumpcode: Analysis – complete depletion



% of Root	Taxonomic Rank	Scientific Name
89.01	Unclassified	unclassified
10.99	Root	root
6.24	Domain*	Viruses
4.7	Root - 1	cellular organisms
3.45	Domain	Eukaryota
0.93	Domain	Bacteria
0.08	Domain	Archaea

Illumina Respiratory Virus Oligo Panel

- RNA denaturation and cDNA synthesis
- Illumina RNA Prep with Enrichment, (L) Tagmentation Kit
- Illumina Respiratory Virus Oligo panel v2
- Illumina MiSeq Reagent v2 Kit
- DRAGEN RNA Pathogen Detection

Viruses Detected by Respiratory Virus Oligo Panel

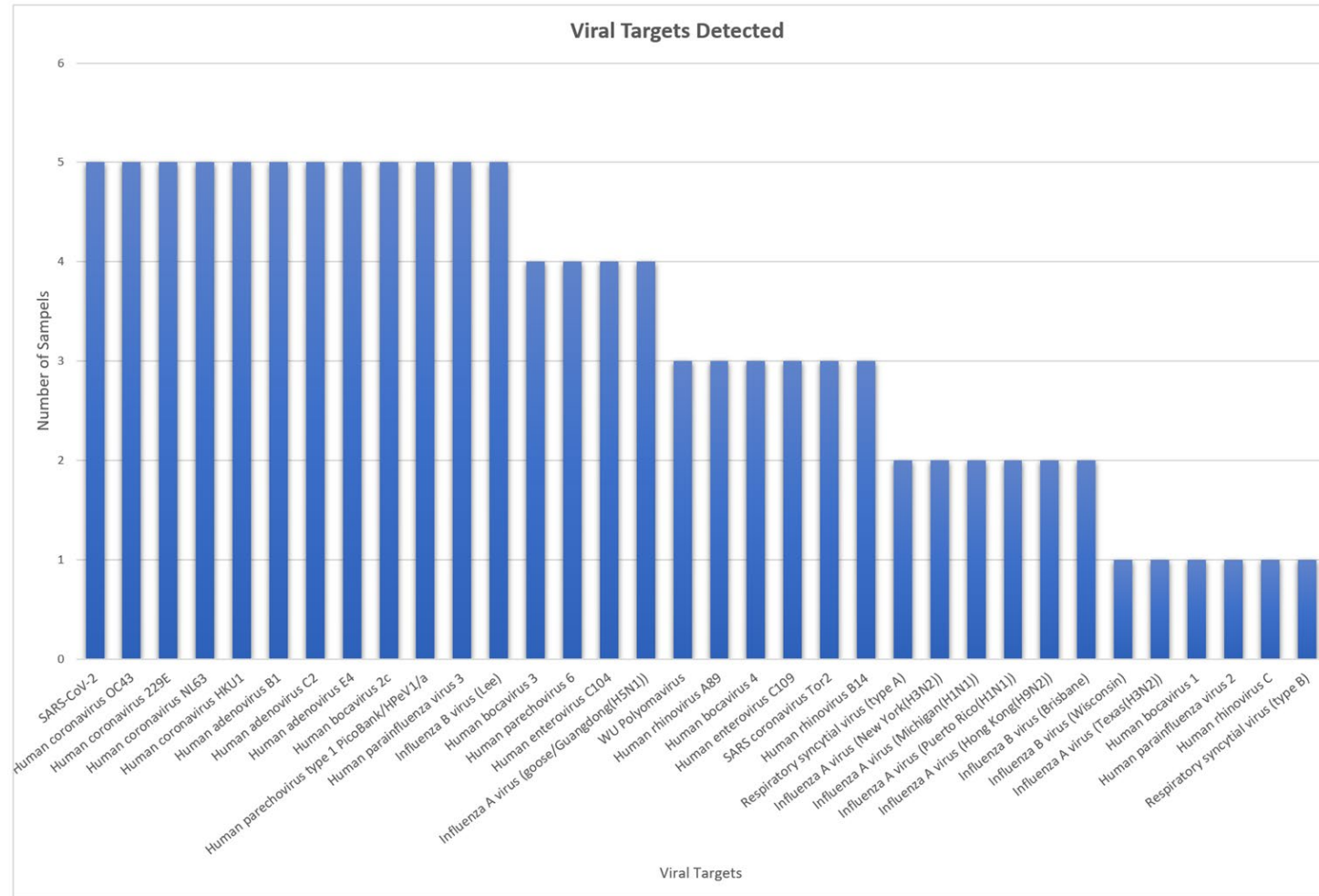
Human coronavirus 229E
Human coronavirus NL63
Human coronavirus OC43
Human coronavirus HKU1
SARS-CoV-2
Human adenovirus B1
Human adenovirus C2
Human adenovirus E4
Human bocavirus 1 (Primate
bocaparvovirus 1 isolate st2)
Human bocavirus 2c PK isolate PK-5510
Human bocavirus 3

Human parainfluenza virus 1
Human parainfluenza virus 2
Human parainfluenza virus 3
Human parainfluenza virus 4a
Human metapneumovirus (CAN97-83)
Respiratory syncytial virus (type A)
Human Respiratory syncytial virus 9320
(type B)
Influenza A virus (A/Puerto
Rico/8/1934(H1N1))
Influenza A virus
(A/Korea/426/1968(H2N2))
Influenza A virus (A/New
York/392/2004(H3N2))
Influenza A virus
(A/goose/Guangdong/1/1996(H5N1))

Human bocavirus 4 NI strain HBoV4-NI-
385
KI polyomavirus Stockholm 60
WU Polyomavirus
Human parechovirus type 1
PicoBank/HPeV1/a
Human parechovirus 6
Human rhinovirus A89
Human rhinovirus C (strain 024)
Human rhinovirus B14
Human enterovirus C104 strain: AK11
Human enterovirus C109 isolate
NICA08-4327

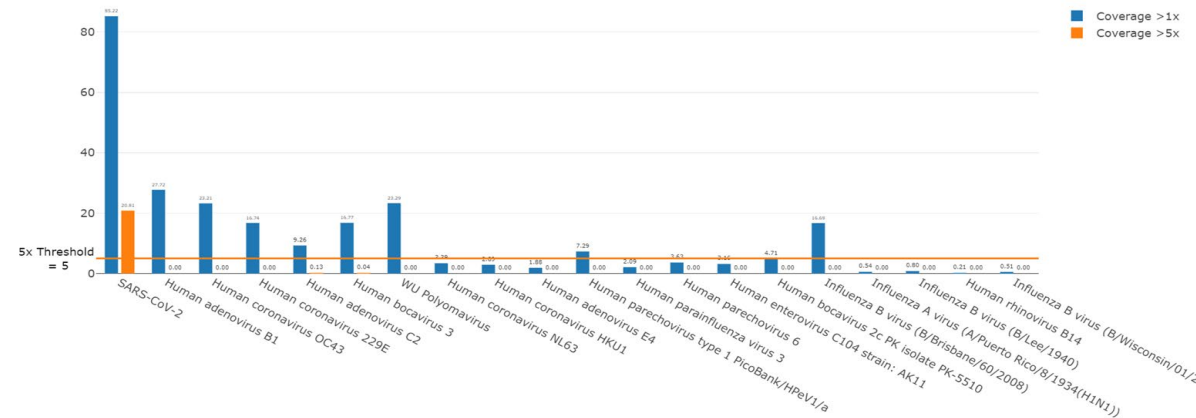
Influenza A virus (A/Zhejiang/DTID-
ZJU01/2013(H7N9))
Influenza A virus (A/Hong
Kong/1073/99(H9N2))
Influenza A virus
(A/Texas/50/2012(H3N2))
Influenza A virus
(A/Michigan/45/2015(H1N1))
Influenza B virus (B/Lee/1940)
Influenza B virus (B/Wisconsin/01/2010)
Influenza B virus (B/Brisbane/60/2008)
Influenza B virus (B/Colorado/06/2017)
Influenza B virus
(B/Washington/02/2019)
Human control genes

Illumina Respiratory Virus Oligo Panel - Analysis

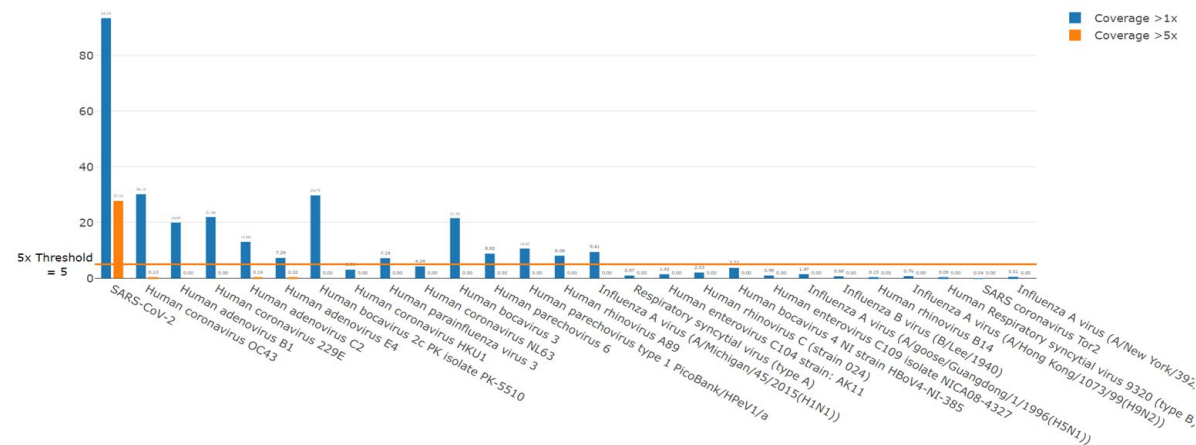


Illumina Respiratory Virus Oligo Panel - Analysis

Percentage of viral genome with coverage >1x, >5x



Percentage of viral genome with coverage >1x, >5x



Epidemiological Applications

- Look for emerging pathogens within samples and across the District.
- Aid in cluster detection
- Identify the increase of seasonal pathogens, particularly around schools and long-term healthcare facilities that are close to sample collection sites.

Limitations

- Metagenomic:
 - Overall presence of pathogen in the community not established.
 - Large number of no-calls
 - Starting sample quality issues
- Jumpcode:
 - Not enough samples and reagents to test alongside ONT process and compare results.
 - Potential loss of data for targets that were not intended to be depleted.
- Illumina:
 - Not generalized; only get the targets intended.
 - Not necessarily whole genome sequencing.

Looking ahead

- Where do we go from here?
- How can we quantify the data to show community prevalence for pathogens of interest?
- Do we keep testing other alternatives and methods?
- How do we develop a reporting mechanism?
- Do we need more partnerships above PHL and DC Health?
- What are the ethical implications of the work?

Acknowledgements

- DC DFS Public Health Laboratory
 - Jocelyn Hauser, PhD
 - David Payne, PhD
- Next Generation Sequencing Core
 - Wadih Bchara
 - Janis Doss, PhD
 - Joel Steyer, PhD
 - Ikenna Uzomah
 - Taylor Boyd
- DC Department of Health
- Molecular Diagnostics Laboratory
 - Elizabeth Zelaya
 - Shauntelle Byfield
 - Monica Mann
 - Andrea Cabello
 - Karon Robinson
 - Onyinye Modika
 - Angela Park
 - Leanna Sealey
 - Vashti Irani
 - Tesfagebriel Hagos
 - Channyn Williams

Thank You!