A Sample-to-results Workflow for Agnostic or Targeted Pathogen Sequencing from Wastewater

Kaylinnette Pinet[1], Patrick Acer[2], Anagha Kadam[1], Lauren Saunders[2], Brittany S. Sexton[1], Katelyn LaVigne[1], Matthew Angel[1], Patrick Andersen[2], Ganit Pricer[2], Ben Lepene[2], V K Chaithanya Ponnaluri[1], Keerthana Krishnan[1]
1. New England Biolabs, Ipswich, MA, USA | 2. Ceres Nanosciences, Manassas, VA, USA

Introduction

Cost-effective population-level pathogen surveillance can be achieved through agnostic or targeted sequencing of wastewater samples. Agnostic pathogen detection via metagenomic sequencing supports surveillance for unknown and novel pathogens. Targeted sequencing permits the monitoring of whole genomes or key genomic regions to track variants for a pathogen of interest. However, wastewater is a challenging substrate that requires optimized enrichment, extraction, and sequencing protocols. Here we present an effective and streamlined sample-to-results workflows for agnostic or targeted pathogen sequencing from wastewater.

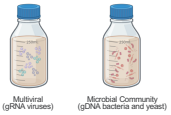
We employed Nanotrap® Microbiome Particles to concentrate intact viruses and bacteria from contrived wastewater samples in a semi-automated, high-throughput method. Nucleic acids were then isolated using the magnetic bead-based Monarch® Mag Viral DNA/RNA Extraction Kit. From these total nucleic acid extracts, we ran agnostic or targeted pathogen sequencing.

When applying an agnostic sequencing approach, we detected expected control pathogens and organic population compositions consistent with wastewater substrates. For the targeted pathogen sequencing workflows, we focused on influenza A (Flu A) and respiratory syncytial virus (RSV). The targeted sequencing of viral pathogens like Flu A and RSV from wastewater is technically challenging due to low viral titers, fragmented RNA, and high background from environmental and host nucleic acids. However, sequencing data following these robust workflows showed high genome coverage and accurate strain identification. Thus, these workflows provide scientists and public health labs with streamlined and cost-effective solutions for monitoring pathogens from population-level wastewater sources.

Methods

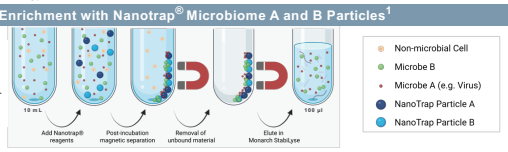
Wastewater Matrices[1]

Wastewater samples were spiked with either a custom multiviral particle mix or commercially available microbial community standard (Zymo Research, D6300). Viral particles were spiked at 10K or 1K copies per ml. The multiviral particle mix consisted of Flu A (H1N1), Flu B, RSV A, and RSV B sourced from ATCC or Zeptometrix. The microbial community standard was spiked at 0.01µl per ml of wastewater. This standard contains 8 bacteria and 2 yeasts; each bacteria consist of 12% of the total gDNA, while the yeast are each 2% of the total gDNA. The bacteria are *Listeria monocytogenes*, *Pseudomonas aeruginosa*, *Bacillus subtilis*, *Escherichia coli*, *Salmonella enterica*, *Lactobacillus fermentum*, *Enterococcus faecalis*, and *Staphylococcus aureus*. The yeast are *Saccharomyces cerevisiae* and *Cryptococcus neoformans*.



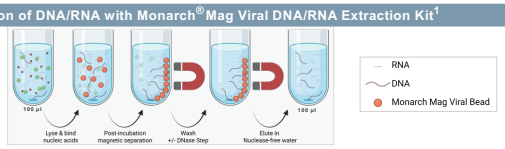
1. Enrichment with Nanotrap® Microbiome A and B Particles[1]

Three 10 mL replicates of each contrived wastewater sample were run through microbial concentration followed by bead-beating of microbial community spiked samples.



2. Extraction of DNA/RNA with Monarch® Mag Viral DNA/RNA Extraction Kit[1]

Total nucleic acids extracted from each 100 µl concentrate.



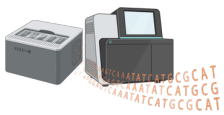
3. NEBNext Sequencing Library Preparation

- **Agnostic Pathogen Sequencing (Nanopore):**
Depletion of uninformative high-background RNA followed by cDNA synthesis where DNA samples are inputs for transposase-based tagmentation and amplification incorporating ONT native barcodes. Following amplification, Oxford Nanopore Technologies® (ONT) library preparation and sequencing were performed.

- **Targeted RSV Sequencing (Illumina):**
RNA was amplified using LunaScript® One-Step RT-PCR and NEBNext® RSV Primer Pools (Mixes 1 & 2). Libraries were prepared with the UltraExpress® FS DNA Kit and sequenced on Illumina.

- **Targeted Influenza A Sequencing (Nanopore):**
Targeted cDNA synthesis and amplification were performed following the NEBNext Flu A Integrated Indexing DNA Library Prep for ONT sequencing workflow[2]. Indexed amplicons were pooled prior to A-tailing and Native Adapter ligation and sequenced on an R10 flow cell.

- **Analysis:**
Reads were trimmed with Seqtk[3] then processed with Minimap2 or Bowtie 2[4] (alignment), Kraken 2 (taxonomy), and Samtools (coverage). Viral content from multiviral agnostic sequencing samples was also assessed with TaxTriage[5].



Results

Pathogen Detection from Wastewater with qPCR

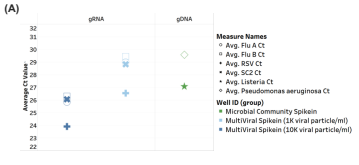


Figure 1. Wastewater was prepared, microbes enriched, and nucleic acids extracted as described in the Methods. qPCR reactions were run on a BioRad CFX96 Real-Time PCR Detection System and Ct values were measured in duplicate. (A) Average Ct values across 3 extraction replicates and 2 qPCR replicates for detecting and quantifying known pathogen targets spiked into the wastewater samples. The 4 gRNA qPCR targets were Flu A, Flu B, RSV, and SARS-CoV-2. The 2 gDNA qPCR targets were *L. monocytogenes* and *P. aeruginosa*.

References

1. Figure created in https://BioRender.com
2. Kaylinnette Pinet, Abhinaya Srikanth, Peter Thielen, New England Biolabs, Juliet Bonnevie 2024. NEBNext iiMS Influenza A DNA Library Prep for Oxford Nanopore. protocols.io
3. H. Li, Seqtk: a fast and lightweight tool for processing FASTA or FASTQ sequences, 2013. https://github.com/lh3/seqtk/
4. Langmead, B., Salzberg, S. Fast gapped-read alignment with Bowtie 2. Nat Methods 9, 357–359 (2012). https://doi.org/10.1038/nmeth.1923
5. Brian Merritt, Jeremy D Ratcliff, Stanley Ta, Gunars Osis, Matthew R. Mauldin, Peter M Thielen. TaxTriage: An Open-Source Metagenomic Sequencing Data Analysis Pipeline Enabling Putative Pathogen Detection. bioRxiv 2025.07.16.664785; doi: https://doi.org/10.1101/2025.07.16.664785

Results Continued

Detection of DNA or RNA Pathogen via Rapid Agnostic Sequencing

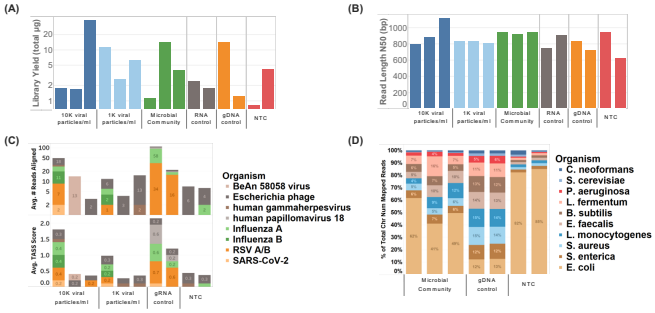


Figure 2. Wastewater was prepared, microbes enriched, and nucleic acids extracted as described in Methods. Libraries were prepared from wastewater extracts and gRNA/gDNA controls with a transposase-based agnostic library prep for Oxford Nanopore Technologies (ONT) sequencing on a ONT PromethION flow cell on a ONT PromethION 2 Solo (P2 Solo) instrument. Reads were mapped to reference genomes with Minimap2 or a Kraken database with TaxTriage. (A) Transposase-based agnostic library yields (log scale) quantified using TapeStation High Sensitivity D5000 reagents. Sufficient ONT library yields were generated for sequencing. (B) Sequenced N50 read length across all samples were ~800 bp. (C) Read counts and TASS (Threat Agnostic Sentinel Surveillance) TaxTriage confidence scores for the detected organisms in the multiviral spiked wastewater samples and pre-extracted gRNA controls. (D) Percent of reads aligned to expected organisms from microbial community standard spiked wastewater samples and pre-extracted gDNA control libraries.

Targeted RSV Sequencing

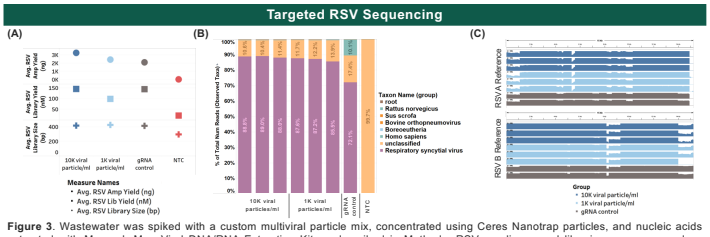


Figure 3. Wastewater was spiked with a custom multiviral particle mix, concentrated using Ceres Nanotrap particles, and nucleic acids extracted with Monarch Mag Viral DNA/RNA Extraction Kit as described in Methods. RSV amplicons and libraries were prepared as described in Methods (Targeted RSV Sequencing). Reads were mapped to reference genomes with Bowtie 2 after down sampling to 2M PE with Seqtk. (A) Average amplicon yields, library yields, and library sizes (n = 2-3). (B) Percent of reads assigned to top 25 taxa hits for each wastewater sample, as well as a positive control and NTC. (C) IGV visualization of genome coverage across RSV A or RSV B reference genomes, comparing enriched wastewater samples and extraction controls (y-axis on a 0–1000 log scale).

Targeted Influenza A Sequencing

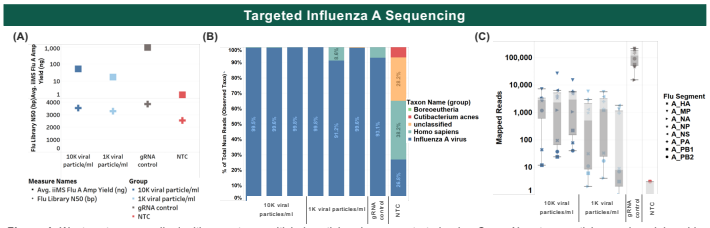


Figure 4. Wastewater was spiked with a custom multiviral particle mix, concentrated using Ceres Nanotrap particles, and nucleic acids extracted with Monarch Mag Viral DNA/RNA Extraction Kit it as described in Methods. Flu A amplicons were generated, pooled, and ONT libraries were prepared as described in Methods (Targeted Influenza A Sequencing). Reads were mapped to reference genomes with Minimap2. (A) Average amplicon yields and library N50 values (n = 2-3). (B) Percent of reads assigned to top 25 taxa hits for each wastewater sample, as well as a positive control and NTC. (C) IGV visualization of genome coverage across an Influenza A composite reference genome, comparing enriched wastewater samples and extraction controls (y-axis on a 0–1000 log scale).

Conclusions

- **Efficient Pipelines for Rapid Agnostic or Targeted Pathogen Sequencing:**
Combining Nanotrap Microbiome A and B particles with Monarch nucleic acid extractions and NEBNext library preparation enables scalable, automation-friendly workflows for wastewater surveillance through agnostic or targeted sequencing.

- **Agnostic sequencing can be applied to pathogenic microbial community assessment:**
Combining agnostic sequencing with a metagenomic analysis pipeline can illuminate microbial community compositions and reveal the presence of pathogenic species.

- **Robust Sequencing for RSV and Influenza A:**
Amplicon-based approaches using dual primer pools or variant-tolerant schemes yield high genome coverage, even from difficult wastewater matrices.

- **NEB is focused on developing variant-tolerant primer schemes and protocols for the targeted pathogen sequencing:**
Our amplicon-based targeted sequencing methods for RSV or Influenza A provide high genome coverage.

Acknowledgements

Applied Physics laboratory at Johns Hopkins University and New England Biolabs sequencing core.