/ CPHM-FRIDAY-260
June 20, 2025

Owen, Christopher N. & Poretsky, Rachel S.
Department of Biological Sciences, University of Illinois Chicago

University of Illinois Chicago
Tel. (312) 355-5102
cowen20@uic.edu

http://owen.science

Improving the Signal-to-Noise Ratio of Metagenomic Wastewater Surveillance for Human Pathogens and ARGs

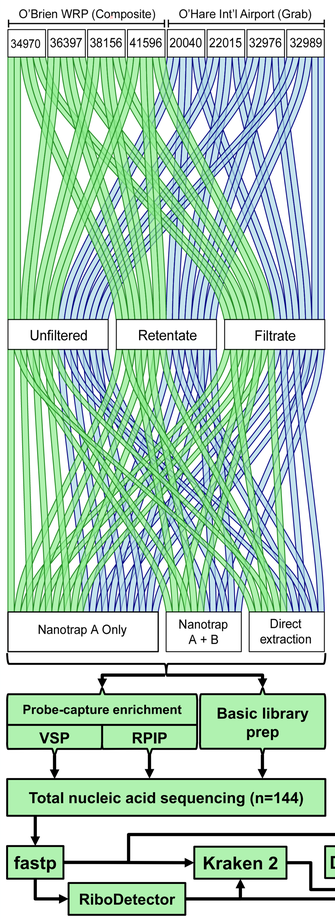
Background

Much of the genetic material found in wastewater originates in the viromes and microbiomes of the surrounding population. This genetic material can therefore be used to infer the presence of pathogens and functional genes such as antimicrobial resistance genes (ARGs) in the local community. However, most genetic sequences in wastewater are not relevant to public health. We subjected wastewater samples to a variety of processing methods intended to increase the relative abundance of public health-relevant sequences.

Methods

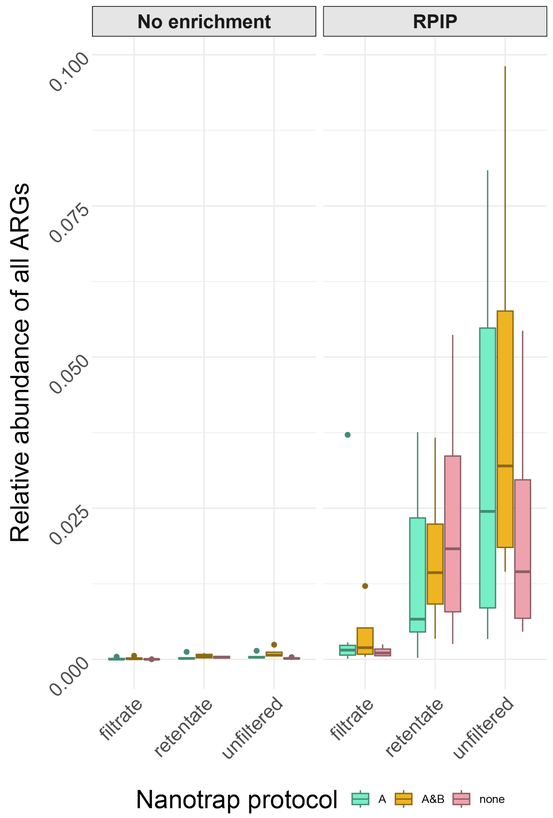
- 4 wastewater samples each from the Terrence J. O'Brien Water Reclamation Plant and O'Hare International Airport.
- Each sample was split into aliquots and subject to several treatments.
- First, we centrifugally filtered aliquots through 24-well plates with 0.45 µm pore size (above). We used both filtrate and resuspended retentate. An unfiltered aliquot of each sample served as a baseline.

- Next, we concentrated aliquots using Nanotrap Microbiome Particles (Ceres Nanosciences)—magnetic beads that adhere to surface structures of cells and viral capsids. We used 2 Nanotrap protocols, designated "A" and "A + B", as well as direct extraction as a baseline.
- Finally, we sequenced each aliquot 3 ways: using two hybridization capture panels plus untargeted total nucleic acid sequencing. The panels we used were the Viral Surveillance Panel (VSP) and Respiratory Pathogen ID and AMR Enrichment Panel (RPIP), both from Illumina.

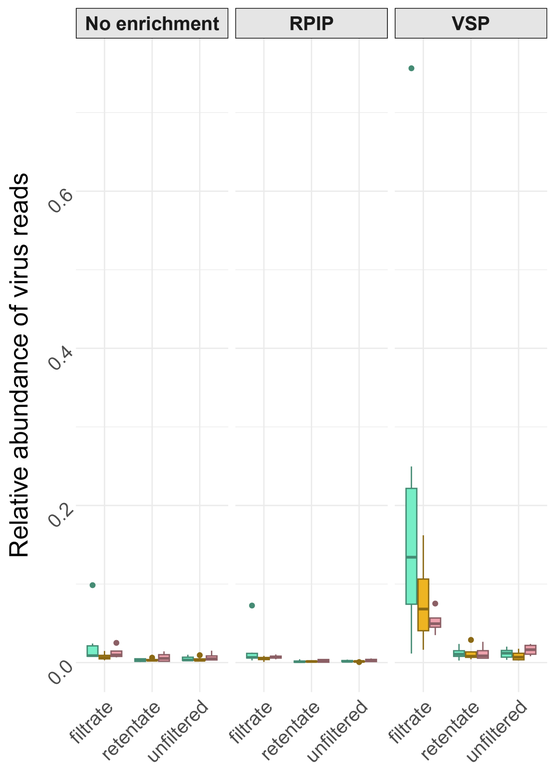


- VSP Targets full genomes of 66 human pathogenic viruses.
- RPIP targets > 280 respiratory pathogens (bacteria, fungi, and viruses), > 2000 AMR markers, and full genomes of SARS-CoV-2 and Influenza A and B.
- After performing read trimming and quality filtering with fastp, we assigned taxonomic IDs with Kraken 2 and identified ARG-like sequences with DeepARG. Ribosomal RNA reads were identified using RiboDetector, and statistical comparisons performed using DESeq2.

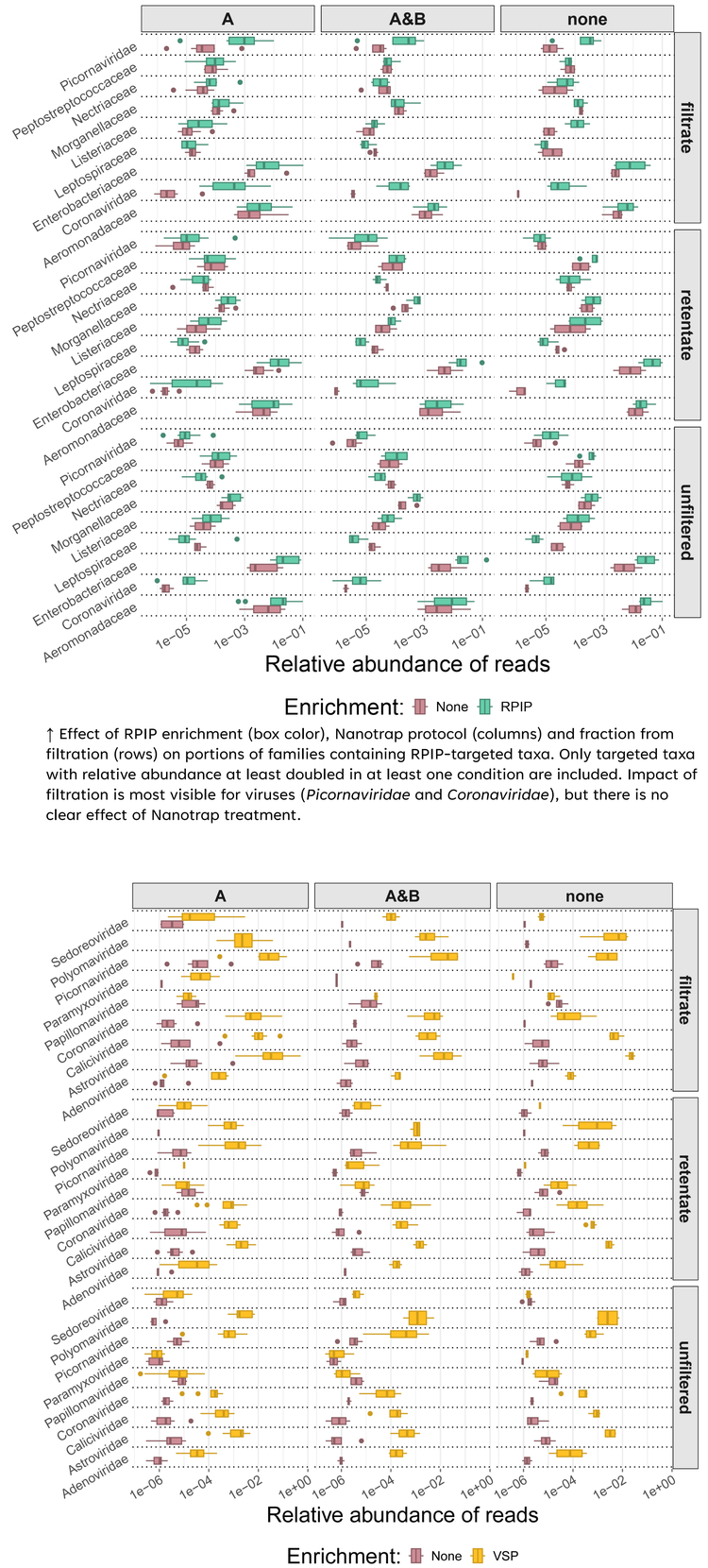
Results & Discussion



↑ Total portions of predicted ARG reads. RPIP enrichment increases the overall portions of ARG reads. Filtration decreases portions of ARG reads, likely because most are associated with cells that do not pass through the filter.

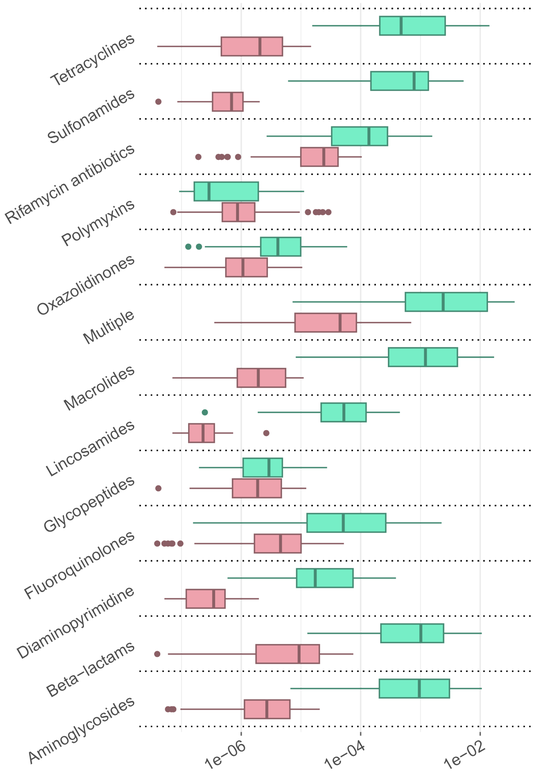


↑ Portion of all reads identified by Kraken2 as belonging to viruses. Filtration combined with VSP enrichment produce the greatest increase in viral reads.

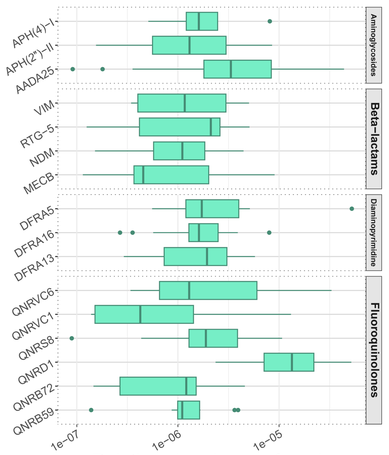


↑ Effect of RPIP enrichment (box color), Nanotrap protocol (columns) and fraction from filtration (rows) on portions of families containing RPIP-targeted taxa. Only targeted taxa with relative abundance at least doubled in at least one condition are included. Impact of filtration is most visible for viruses (*Picornaviridae* and *Coronaviridae*), but there is no clear effect of Nanotrap treatment.

↑ Effect of VSP enrichment (box color), Nanotrap protocol (columns) and fraction from filtration (rows) on portions of reads from families containing VSP-targeted taxa. Filtrates generally produce somewhat greater portions of viral reads (though this effect is not as pronounced in individual taxa), while there was not a clear consistent effect of Nanotrap treatment on these taxa.



↑ RPIP-targeted ARGs identified as having significant (based on DESeq2 analysis) differences in relative abundance between RPIP-enriched vs. unenriched samples, pooled by CARD resistance category.



↑ ARGs found in RPIP-enriched samples but absent from all unenriched samples.

Acknowledgements & References

We would like to thank Kevin J. Kunstman and Stefan J. Green of the RUSH Genomics and Microbiome Core Facility for their collaboration and the Rush University Center for Emerging Infectious Diseases for funding this research.

Arango-Argoty, G., Garner, E., Pruden, A. *et al.* DeepARG: a deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* **6**:23 (2018). doi: 10.1186/s40168-018-0401-z

Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**(17):i884-i890 (2018). doi: 10.1093/bioinformatics/bty560

Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:550 (2014). doi: 10.1186/s13059-014-0550-8

Wood, D.E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol* **20**:257 (2019). doi: 10.1186/s13059-019-1891-0

Zhi-Luo Deng, Philipp C Münch, René Mreches, Alice C McHardy, Rapid and accurate identification of ribosomal RNA sequences via deep learning, *Nucleic Acids Research*, **50**(10):e60 (2022). doi: 10.1093/nar/gkac112