Supplemental information for

Deep untargeted wastewater metagenomic sequencing from sewersheds across the United States

Lennart J. Justen*[1, 2, 3], Clayton Rushford*[4], Olivia S. Hershey[5], Róisín Floyd-O'Sullivan[5], Simon L. Grimm[5], William J. Bradshaw[5], Harmon Bhasin[5], Daniel P. Rice[5], Katherine Stansifer[5], Jo D. Faraguna[5], Michael R. McLaren[5], Alessandro Zulli[5], Alejandro Tovar-Mendez[6], Emma E. Copen[6], Kristen K. Shelton[7], Ayaaz Amirali[8], Sherin Kannoly[9], Sofia Pesantez[9], Aiden Stanciu[9], Iñigo Caballero Quiroga[9], Leopolda Silvera[10], Nicole Greenwood[11], Barbra Bongiovi[12], Austin Walkins[12], Ryan Love[13], Scott Lening[13], Kaylyn Patterson[14], Theresa Johnston[14], Sandra Hernandez[15], Aymara Benitez[15], Billie Jo McCarley[15], Samantha Engelage[16], Suguna Pillay[16], Cindy Calender[17], Brent Herring[18], Carey Robinson[19], Monett Wastewater Treatment Plant, Columbia Missouri Wastewater Treatment Plant, Daniel Cunningham-Bryant[20], Gordon Adams[3,21], Jillian Paull[3,22], Jamie Devlin[21], Vamsi Thiriveedhi[23], Sarah E. Turbett[21,23,24], Jacob E. Lemieux[3,21,24], Rose S. Kantor[25], David H. O'Connor[26], John J. Dennehy[9,27], Rachel Poretsky[28], Jason A. Rothman[29], Helena M. Solo-Gabriele[8], Jason R. Vogel[7], Pardis C. Sabeti[3,22,30,31], Jeff Kaufman†[5], Marc C. Johnson†[6]

*These authors contributed equally to this work.
†Corresponding authors: jeff@securebio.org, marcjohnson@health.missouri.edu

Affiliations
[1]Massachusetts Institute of Technology, Media Lab, Cambridge, MA, USA
[2]The Charles Stark Draper Laboratory, Inc., Draper Scholar, Cambridge, MA, USA
[3]Broad Institute of MIT and Harvard, Cambridge, MA, USA
[4]University of Missouri, Department of Veterinary Pathobiology, Columbia, MO, USA
[5]SecureBio, Cambridge, MA, USA
[6]University of Missouri, Department of Molecular Microbiology and Immunology, Columbia, MO, USA
[7]University of Oklahoma, School of Civil Engineering and Environmental Science, Norman, OK, USA
[8]University of Miami, Department of Chemical, Environmental, and Materials Engineering, Miami, FL, USA
[9]Queens College of The City University of New York, Biology Department, Flushing, NY, USA
[10]NYC Health + Hospitals, New York, NY, USA
[11]Riverside Water Quality Control Plant, Riverside, CA, USA
[12]City of Boise, Boise, ID, USA
[13]Inland Empire Utilities Agency, Chino, CA, USA
[14]Metropolitan Water Reclamation District of Greater Chicago, Chicago, IL, USA
[15]Miami-Dade Water and Sewer Department, Miami, FL, USA
[16]City of Palo Alto, Palo Alto, CA, USA
[17]City of Kansas City Missouri, KC Health, Communicable Disease and Prevention Division, Kansas City, MO
[18]City of Kansas City Missouri, KC Water, Water and Wastewater Operations, Kansas City, MO
[19]City of Kansas City Missouri, KC Water, Wastewater Treatment Division, Kansas City, MO

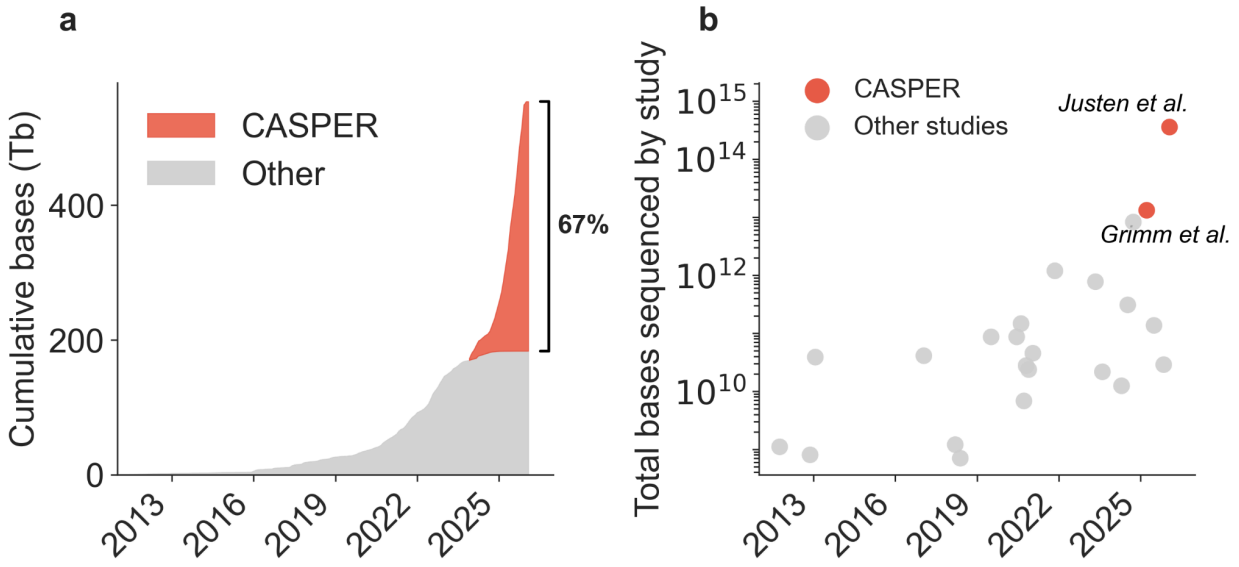
[20]The Charles Stark Draper Laboratory, Inc., Cambridge, MA, USA
[21]Massachusetts General Hospital, Division of Infectious Diseases, Boston, MA, USA
[22]Howard Hughes Medical Institute, Chevy Chase, MD, USA
[23]Massachusetts General Hospital, Department of Pathology, Boston, MA, USA
[24]Harvard Medical School, Boston, MA, USA
[25]Lawrence Livermore National Laboratory, Physical and Life Sciences Directorate, Livermore, CA, USA
[26]University of Wisconsin-Madison, Department of Pathology and Laboratory Medicine, Madison, WI, USA
[27]The Graduate Center of The City University of New York, New York, NY, USA
[28]University of Illinois Chicago, Department of Biological Sciences, Chicago, IL, USA
[29]University of California, Riverside, Department of Microbiology and Plant Pathology, Riverside, CA, USA
[30]Harvard University, Harvard T.H. Chan School of Public Health, Department of Immunology and Infectious Diseases, Boston, MA, USA
[31]Harvard University, Faculty of Arts and Sciences, Department of Organismic and Evolutionary Biology, Cambridge, MA, USA

Table of contents

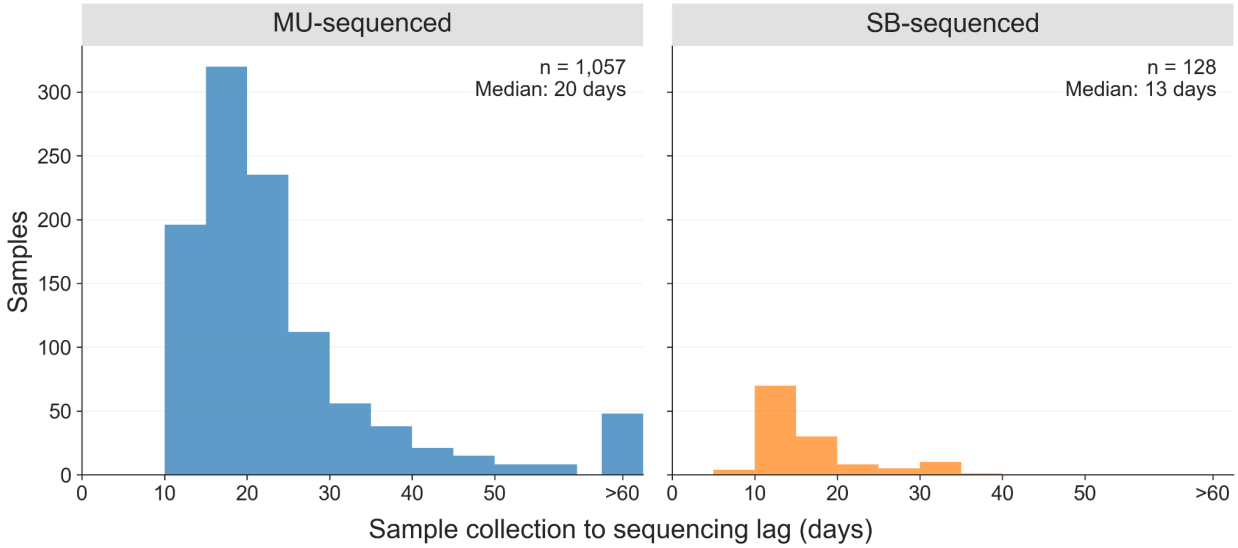
Figures

Tables

Supplementary Figure S1. Sequencing depth (total read pairs) over time by site and sequencing partner.

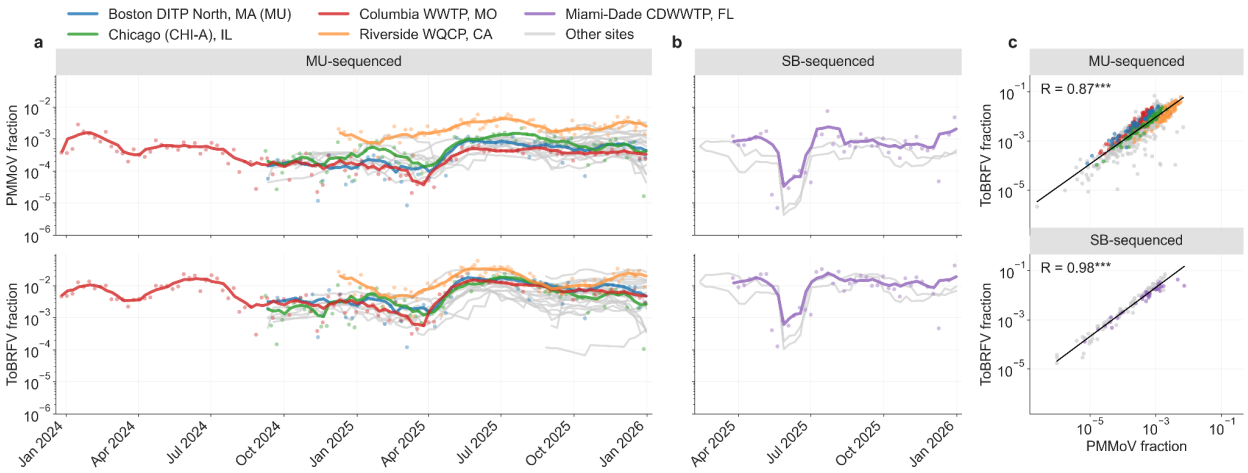


Supplementary Figure S2. CASPER contribution to untargeted wastewater metagenomic sequencing (WW-MGS) data on the NCBI Sequence Read Archive (SRA). (a) CASPER data from this release (PRJNA1247874; 357 Tb) and (Grimm et al. 2025) (PRJNA1198001; 13 Tb) as a fraction of the cumulative untargeted WW-MGS bases deposited on the NCBI SRA over time. (b) Total untargeted WW-MGS bases by study for publicly available WW-MGS datasets identified in our literature survey (Supplementary Table S1).

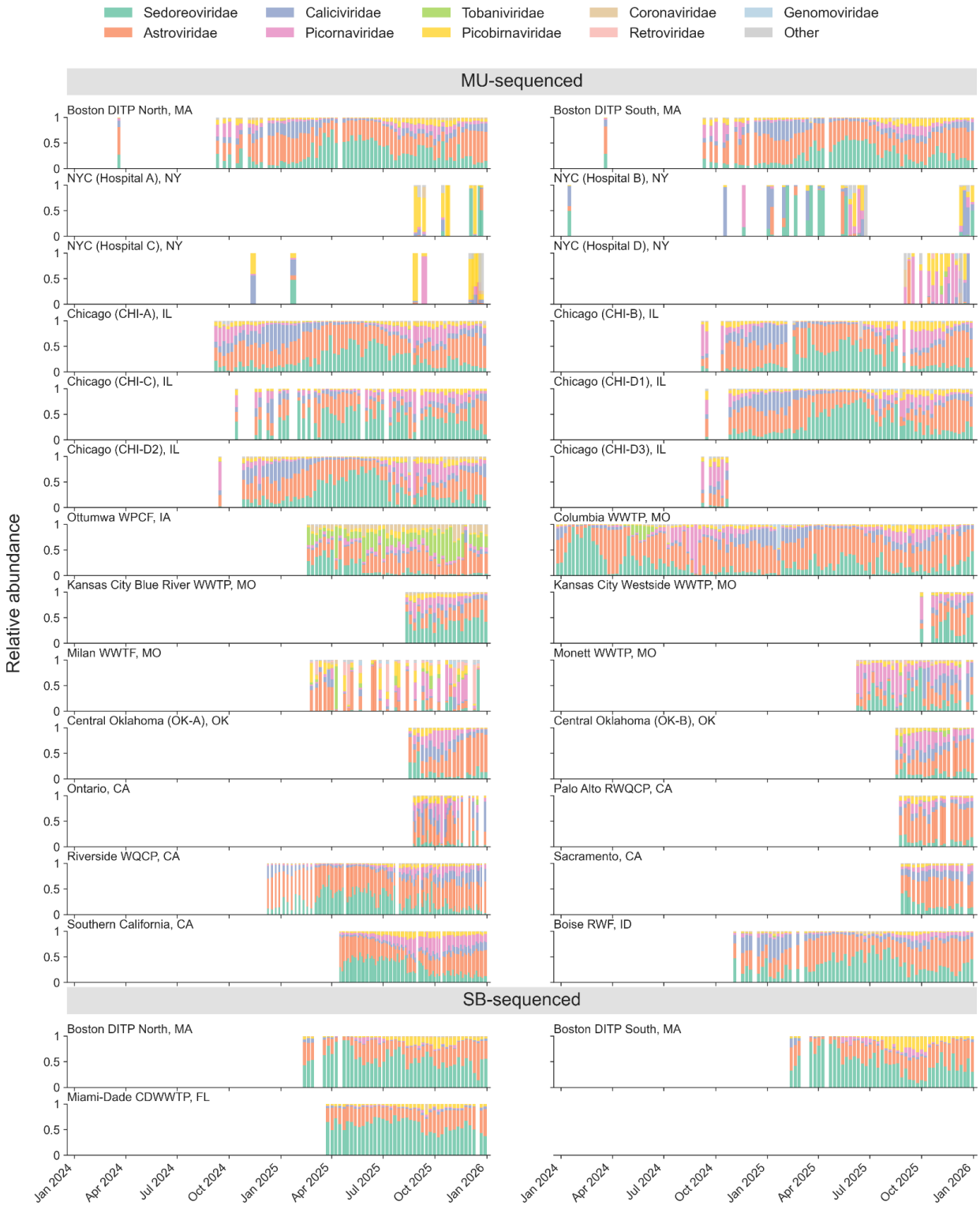


Supplementary Figure S3. Sample turnaround time from collection to sequencing data delivery. Distribution of turnaround time (days) by site and sequencing laboratory. The upper range reflects NYC hospital grab samples that were stored at 4°C for several months prior to extraction during a pilot phase.

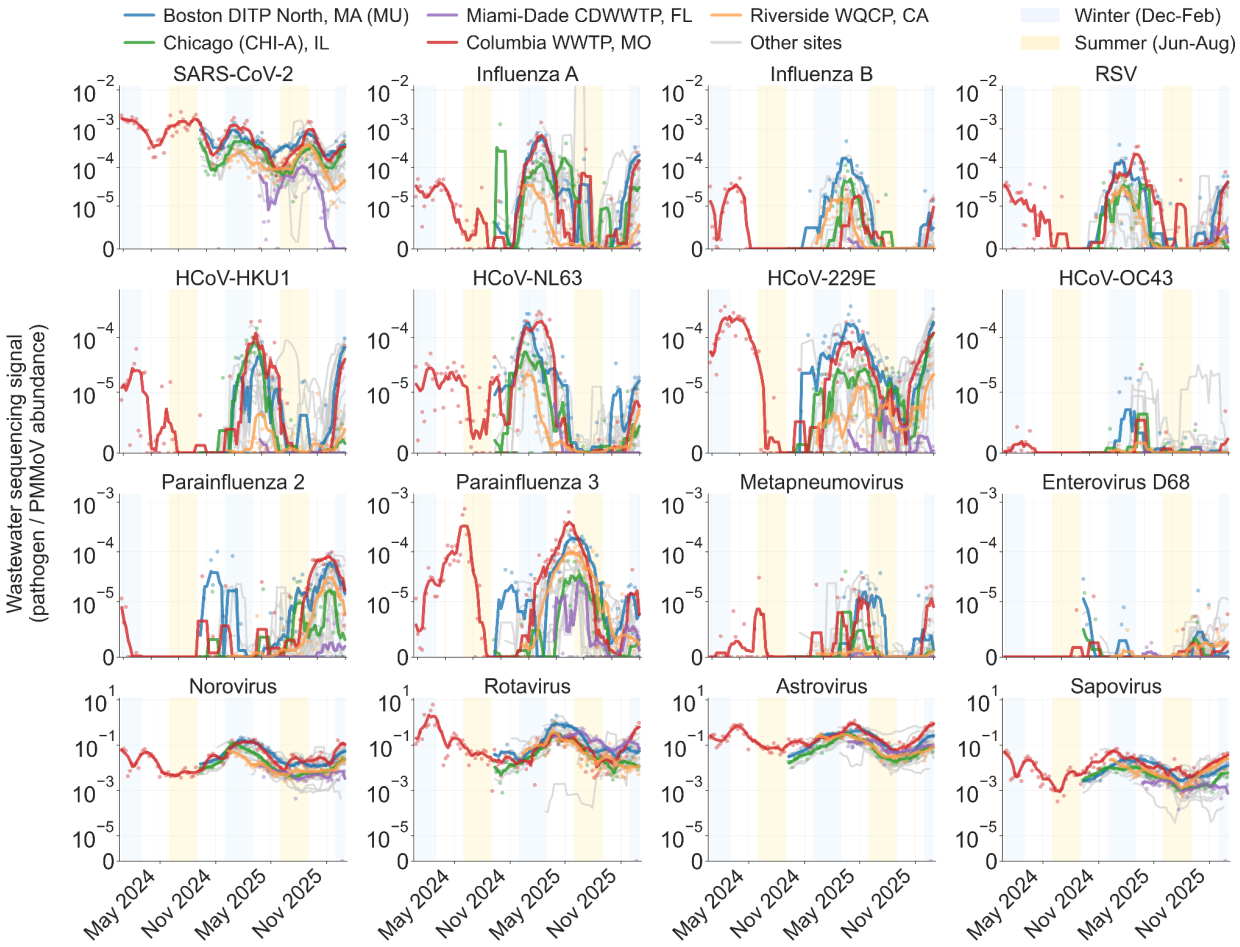
Supplementary Figure S4. Taxonomic composition over time by site. Proportion of reads assigned to major taxonomic categories for each sampling location, grouped by sequencing partner.



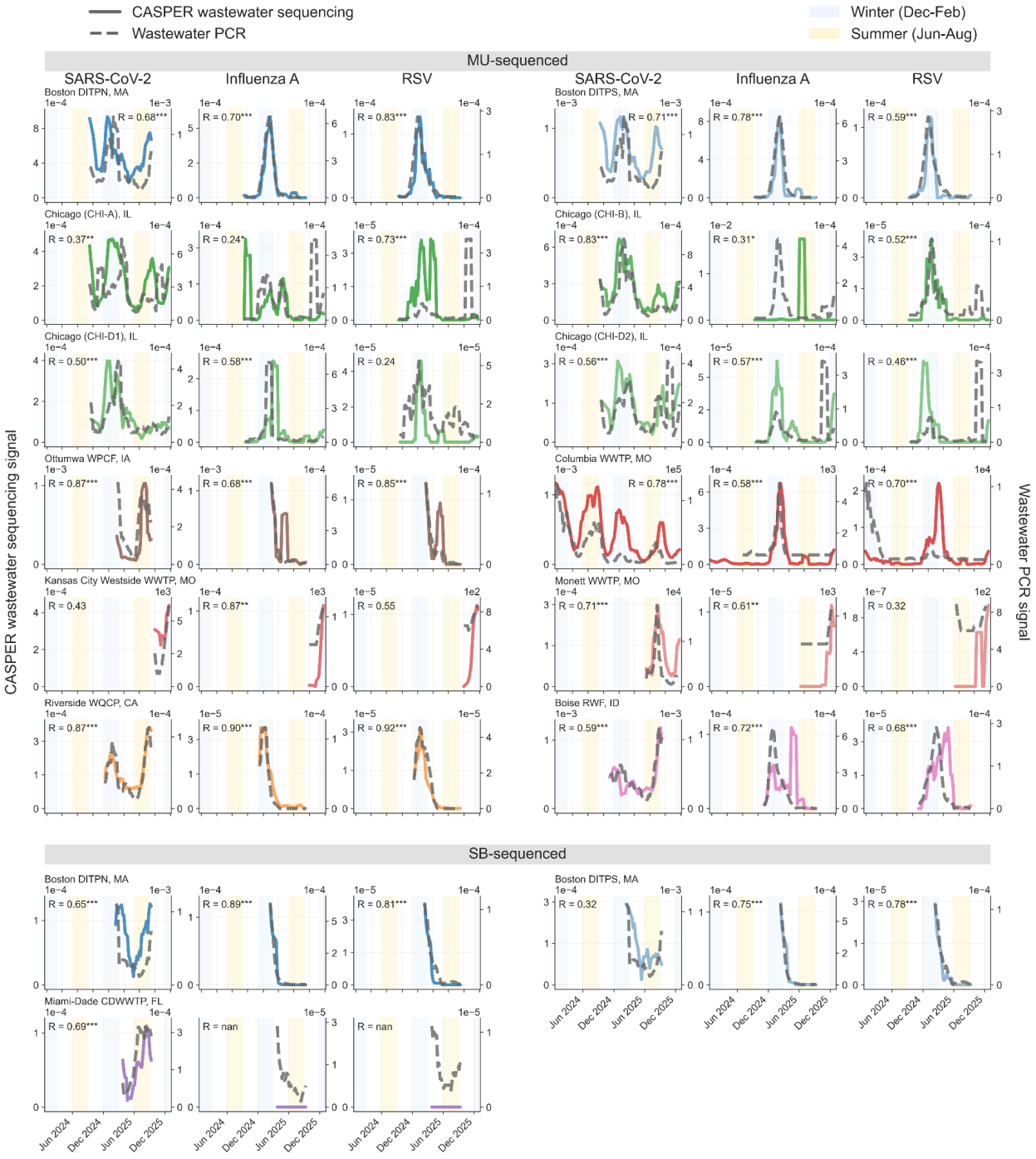
Supplementary Figure S5. Pepper mild mottle virus (PMMoV) and tomato brown rugose fruit virus (ToBRFV) abundance over time and correlation across sites. (a) Fraction of total reads classified as PMMoV (top row) and ToBRFV (bottom row) over time for samples sequenced at the University of Missouri (MU) (a) and at SecureBio (SB) (b) sites. (c) Correlation between PMMoV and ToBRFV read fractions for MU-sequenced (top; n=1,078, Pearson R=0.87, p<0.001) and SB-sequenced (bottom; n=128, Pearson R=0.98, p<0.001) samples. *p < 0.05, **p < 0.01, ***p < 0.001.



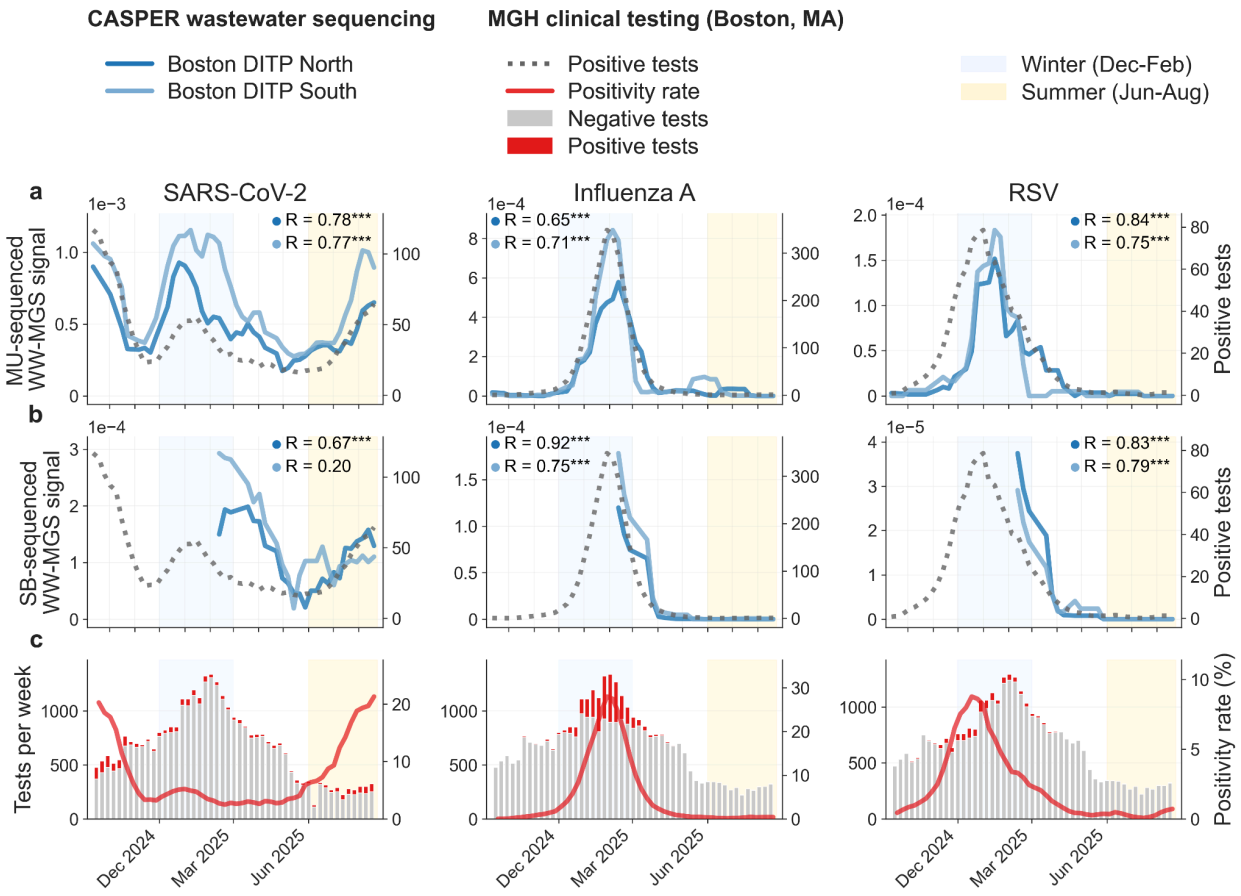
Supplementary Figure S6. Vertebrate-infecting virus family composition over time by site. Relative abundance of vertebrate-infecting virus (VV) reads by viral family for each sampling location, grouped by sequencing partner.



Supplementary Figure S7. Temporal abundance of select viruses across CASPER sites (log scale). PMMoV-normalized abundance of human-infecting respiratory viruses (top rows) and gastrointestinal viruses (bottom row) over time.

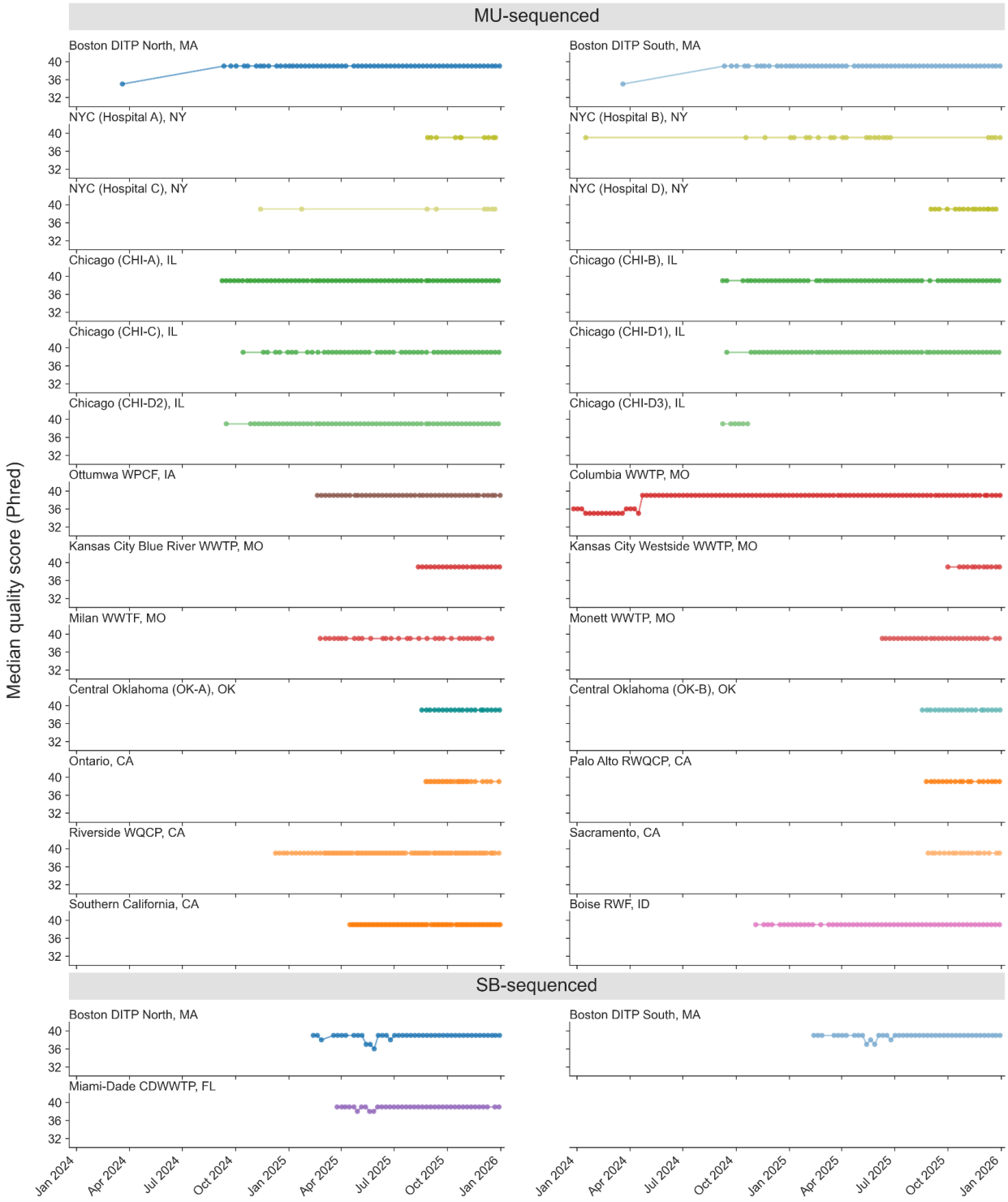


Supplementary Figure S8. Agreement between wastewater metagenomic sequencing and wastewater PCR measurements across all paired sites. Pepper mild mottle virus (PMMoV)-normalized wastewater metagenomic sequencing abundance compared with wastewater PCR concentrations from the CDC National Wastewater Surveillance System data portal. Missouri sites (Columbia, Monett, Kansas City) are not PMMoV-normalized as PMMoV-normalized PCR data were unavailable. Time series data are aggregated to MMWR epidemiological weeks and smoothed with a 5-week centered moving average. Spearman correlation coefficients (R) calculated on week-aligned smoothed values. *p < 0.05, **p < 0.01, ***p < 0.001. MU: University of Missouri; SB: SecureBio.



Supplementary Figure S9. Clinical respiratory virus testing from Massachusetts General Hospital. (a) University of Missouri (MU)-sequenced pepper mild mottle virus (PMMoV)-normalized wastewater metagenomic sequencing abundance from Boston Deer Island Treatment Plant (DITP) North and South systems compared with weekly positive test counts from Massachusetts General Hospital (MGH) testing. (b) SecureBio (SB)-sequenced PMMoV-normalized wastewater metagenomic sequencing abundance compared with the same clinical data. (c) Weekly testing volume with negative and positive results, along with positivity rate. Clinical testing followed a seasonal algorithm with expanded multiplex molecular panels during fall and winter months (see Methods). Spearman correlation coefficients (R) shown with significance: *p < 0.05, **p < 0.01, ***p < 0.001.

Supplementary Figure S10. GC content over time by site. Mean GC content (%) for each sampling location, grouped by sequencing partner. Values were calculated from subsampled reads after adapter trimming and quality filtering (see Methods). The transient decrease for Columbia, MO in mid-2024 coincides with the transition from NovaSeq 6000 to NovaSeq X.



Supplementary Figure S11. Sequencing quality scores over time by site. Median raw read Phred quality score for each sampling location, grouped by sequencing partner. The increase for Columbia, MO and Boston, MA in mid-2024 coincides with the transition from NovaSeq 6000 to NovaSeq X.

Supplementary Figure S12. Read length over time by site. Mean raw read length (bp) for each sampling location, grouped by sequencing laboratory. Most libraries were sequenced at 2x150 bp; 17 MU-processed libraries were sequenced at 2x100 bp due to diagnostic runs or sequencing configuration errors (see Methods).

Supplementary Figure S13. Quality control (QC) pass rate over time by site. Fraction of reads passing adapter trimming and quality filtering by FASTP (see Methods).

Supplementary Table S1. Summary of sequencing depth across published wastewater metagenomic sequencing (WW-MGS) studies. Median and total gigabases (Gb) are reported for untargeted WW-MGS samples only, excluding amplicon or hybrid-capture samples where applicable. Base pair statistics were obtained either directly from the NCBI Sequence Read Archive (SRA) or estimated from values reported in the manuscript text, tables, or supplementary materials. Studies where per-sample sequencing depth could not be determined from public data or manuscript reporting are noted. CASPER data comprise two BioProjects: the Los Angeles pilot dataset (PRJNA1198001; Grimm et al., 2025) and the multi-site release described in this paper (PRJNA1247874; Justen et al., 2026).

Study	Bioproject	Median Gb (WW-MGS samples)	Total Gb (WW-MGS samples)	Base stats method
Cantalupo et al. 2011	PRJNA70623	0.4	1.1	SRA
Ng et al. 2012	PRJNA169010	0.2	0.8	SRA
Bibby and Peccia 2013		3.2	38.7	estimated
Hjelmsø et al. 2017	PRJEB15242	0.9	40.9	estimated
Fernandez-Cassi et al. 2018		0.3	1.2	estimated
Adriaenssens et al. 2018	PRJNA421889	0.1	0.7	SRA
Maritz et al. 2019	PRJEB28033	5.1	86.4	SRA
Martínez-Puchol et al. 2020	–	Could not determine	Could not determine	–
Gulino et al. 2020	PRJEB28033	5.1	86.4	SRA
Nieuwenhuijse et al. 2020	PRJEB23496	1.4	146.4	SRA
Guajardo-Leiva et al. 2020	PRJNA648644	6.8	6.8	SRA
Brinch et al. 2020	PRJEB34633	4.1	27.6	SRA
Wang et al. 2020	–	1.8	23.6	estimated
Crits-Christoph et al. 2021	PRJNA661613	6.3	45.2	SRA
Rothman et al. 2021	PRJNA729801	3.9	1189.4	SRA
Spurbeck et al. 2023	PRJNA924011	Could not determine	Could not determine	–
Stockdale et al. 2023	PRJNA842541	5.6	771.8	SRA
McCall et al. 2023	PRJNA943189	10.8	21.6	SRA
Child et al. 2023	PRJEB62830	Could not determine	Could not determine	–
Fernandez-Cassi and Kohn 2024	PRJNA948647	0.8	12.4	SRA
Wyler et al. 2024	PRJNA948850	2.5	307.5	estimated
Tierney et al. 2024	PRJNA946141	6.0	8235.1	SRA
Grimm et al. 2025 (CASPER)	PRJNA1198001	327.3	13132.6	SRA
Guajardo-Leiva et al. 2025	PRJNA1196808	10.1	136.3	SRA
Worp et al. 2025	PRJEB87273	0.1	28.7	estimated
Bellekom et al. 2026	PRJNA1261831	Could not determine	Could not determine	–
Justen et al., 2026 (CASPER)	PRJNA1247874	289.7	357010	SRA

Supplementary Table S2. Historical sequencing cost of a typical CASPER sample. Cost per megabase (Mb) is derived from the NHGRI Genome Sequencing Costs dataset through February 2022, with subsequent values estimated based on observed sequencing output and pricing (Wetterstrand 2019; Kaufman 2025). Per-sample cost is calculated using the median CASPER untargeted sequencing depth in this release of 290 Gb per sample.

Date	Cost per Mb	Cost per median CASPER sample (290 Gb)	Source
Jan 2007	$522.71	$151.43M	NHGRI
Jan 2008	$102.13	$29.59M	NHGRI
Jan 2009	$2.59	$750.32k	NHGRI
Jan 2010	$0.52	$150.64k	NHGRI
Jan 2011	$0.23	$66.63k	NHGRI
Jan 2012	$0.09	$26.07k	NHGRI
Jan 2013	$0.06	$17.38k	NHGRI
Jan 2014	$0.04	$11.59k	NHGRI
Jan 2015	$0.04	$11.59k	NHGRI
May 2016	$0.013	$3.77k	NHGRI
Feb 2017	$0.011	$3.19k	NHGRI
Feb 2018	$0.014	$4.06k	NHGRI
Feb 2019	$0.011	$3.19k	NHGRI
Feb 2020	$0.007	$2.03k	NHGRI
Feb 2021	$0.009	$2.61k	NHGRI
Feb 2022	$0.006	$1.74k	NHGRI
Jul 2023	$0.005	$1.45k	estimated
Jul 2024	$0.003	$869.10	estimated
Jan 2025	$0.002	$579.40	estimated
Jan 2026	$0.002	$579.40	estimated

Supplementary Table S3. National Wastewater Surveillance System (NWSS) sewershed identifiers for CASPER sites. Identifiers enable linkage to publicly available NWSS PCR data. Sites marked "n/a" lack site-specific NWSS data; sites marked "–" either have no NWSS match or declined to share this information.

Sewershed	NWSS sewershed_id
Boston DITP North, MA	752, 742
Boston DITP South, MA	743, 742
New York (Hospital A), NY	n/a
New York (Hospital B), NY	n/a
New York (Hospital C), NY	n/a
New York (Hospital D), NY	n/a
Chicago (CHI-A), IL	419
Chicago (CHI-B), IL	413
Chicago (CHI-C), IL	n/a
Chicago (CHI-D1), IL	423
Chicago (CHI-D2), IL	429
Chicago (CHI-D3), IL	n/a
Ottumwa WPCF, IA	541
Columbia WWTP, MO	1045
Kansas City Blue River WWTP, MO	1074
Kansas City Westside WWTP, MO	1075
Milan WWTF, MO	n/a
Monett WWTP, MO	1041
Miami-Dade CDWWTP, FL	309
Central Oklahoma (OK-A), OK	–
Central Oklahoma (OK-B), OK	–
Ontario, CA	143
Palo Alto RWQCP, CA	175
Riverside WQCP, CA	138
Sacramento, CA	140
Southern California, CA	–
Boise RWF, ID	372, 374

Supplementary Table S4. Correlation between wastewater metagenomic sequencing and wastewater PCR data. Spearman correlation coefficients (R) comparing wastewater metagenomic sequencing (WW-MGS) abundance with wastewater PCR measurements from the CDC National Wastewater Surveillance System data portal. PCR pathogen concentrations are normalized by pepper mild mottle virus (PMMoV) for all sites except Columbia, Monett, and Kansas City, MO, where PMMoV-normalized PCR data were unavailable and raw concentrations were used instead. For WW-MGS data, we report correlations using two normalization approaches: PMMoV-normalized relative abundance and ToBRFV (tomato brown rugose fruit virus)-normalized relative abundance. Correlations were calculated using MMWR week-aggregated data with centered 5-week smoothing. *p < 0.05, **p < 0.01, ***p < 0.001. MU: University of Missouri; SB: SecureBio.

		Spearman correlation coefficient (R)					
		PMMoV-normalized sequencing			ToBRFV-normalized sequencing		
Sampling site	**N (weeks)**	**SARS-CoV-2**	**Influenza A**	**RSV**	**SARS-CoV-2**	**Influenza A**	**RSV**
MU-sequenced							
Boston DITP North, MA	50	0.68***	0.70***	0.83***	0.75***	0.70***	0.82***
Boston DITP South, MA	50	0.71***	0.78***	0.59***	0.72***	0.77***	0.57***
Chicago (CHI-A), IL	69	0.37**	0.24*	0.73***	0.37**	0.24	0.73***
Chicago (CHI-B), IL	63	0.83***	0.31*	0.52***	0.85***	0.31*	0.52***
Chicago (CHI-D1), IL	63	0.50***	0.58***	0.24	0.66***	0.60***	0.21
Chicago (CHI-D2), IL	63	0.56***	0.57***	0.46***	0.58***	0.56***	0.41***
Ottumwa WPCF, IA	31	0.87***	0.68***	0.85***	0.88***	0.63***	0.91***
Columbia WWTP, MO	101	0.78***	0.58***	0.70***	0.76***	0.59***	0.76***
Kansas City Westside WWTP, MO	9	0.43	0.87**	0.55	0.45	0.87**	0.60
Monett WWTP, MO	24	0.71***	0.61**	0.32	0.53**	0.61**	0.32
Riverside WQCP, CA	41	0.87***	0.90***	0.92***	0.78***	0.90***	0.90***
Boise RWF, ID	42	0.59***	0.72***	0.68***	0.80***	0.73***	0.76***
SB-sequenced							
Boston DITP North, MA	29	0.65***	0.89***	0.81***	0.66***	0.89***	0.81***
Boston DITP South, MA	29	0.32	0.75***	0.78***	0.31	0.75***	0.78***
Miami-Dade CDWWTP, FL	26	0.69***	nan	nan	0.59**	nan	nan

Supplementary Table S5. Wastewater PCR assay details and data source designations for CASPER comparison sites. Data were obtained from the CDC National Wastewater Surveillance System (NWSS) data portal; source designations indicate whether data were generated by state and local health departments or by WastewaterSCAN (Boehm et al., 2024, 2026).

Sampling site	Source	PCR type	PCR targets	Normalization
Boston DITP North, MA (742)				
SARS-CoV-2	WastewaterSCAN	ddPCR	N	PMMoV
Influenza A	WastewaterSCAN	ddPCR	INFA1, INFA1+INFA2	PMMoV
RSV	WastewaterSCAN	ddPCR	RSV-A+RSV-B	PMMoV
Boston DITP South, MA (742)				
SARS-CoV-2	WastewaterSCAN	ddPCR	N	PMMoV
Influenza A	WastewaterSCAN	ddPCR	INFA1, INFA1+INFA2	PMMoV
RSV	WastewaterSCAN	ddPCR	RSV-A+RSV-B	PMMoV
Chicago (CHI-A), IL (419)				
SARS-CoV-2	State/local health dept.	dPCR	N1	PMMoV
Influenza A	State/local health dept.	dPCR	INFA1+INFA2	PMMoV
RSV	State/local health dept.	dPCR	RSV-A+RSV-B	PMMoV
Chicago (CHI-B), IL (413)				
SARS-CoV-2	State/local health dept.	dPCR	N1	PMMoV
Influenza A	State/local health dept.	dPCR	INFA1+INFA2	PMMoV
RSV	State/local health dept.	dPCR	RSV-A+RSV-B	PMMoV
Chicago (CHI-D1), IL (423)				
SARS-CoV-2	State/local health dept.	dPCR	N1	PMMoV
Influenza A	State/local health dept.	dPCR	INFA1+INFA2	PMMoV
RSV	State/local health dept.	dPCR	RSV-A+RSV-B	PMMoV
Chicago (CHI-D2), IL (429)				
SARS-CoV-2	State/local health dept.	dPCR	N1	PMMoV
Influenza A	State/local health dept.	dPCR	INFA1+INFA2	PMMoV
RSV	State/local health dept.	dPCR	RSV-A+RSV-B	PMMoV
Ottumwa WPCF, IA (541)				
SARS-CoV-2	WastewaterSCAN	ddPCR	N	PMMoV
Influenza A	WastewaterSCAN	ddPCR	INFA1+INFA2	PMMoV
RSV	WastewaterSCAN	ddPCR	RSV-A+RSV-B	PMMoV
Columbia WWTP, MO (1045)				
SARS-CoV-2	State/local health dept.	dPCR	N2+N1	None
Influenza A	State/local health dept.	dPCR	M	None
RSV	State/local health dept.	dPCR	M+N2	None
Kansas City Westside WWTP, MO (1075)				
SARS-CoV-2	State/local health dept.	dPCR	N2+N1	None
Influenza A	State/local health dept.	dPCR	M	None
RSV	State/local health dept.	dPCR	M+N2	None
Monett WWTP, MO (1041)				
SARS-CoV-2	State/local health dept.	dPCR	N2+N1	None
Influenza A	State/local health dept.	dPCR	M	None
RSV	State/local health dept.	dPCR	M+N2	None
Miami-Dade CDWWTP, FL (309)				
SARS-CoV-2	WastewaterSCAN	ddPCR	N	PMMoV
Influenza A	WastewaterSCAN	ddPCR	INFA1, INFA1+INFA2	PMMoV
RSV	WastewaterSCAN	ddPCR	RSV-A+RSV-B	PMMoV
Riverside WQCP, CA (138)				
SARS-CoV-2	WastewaterSCAN	ddPCR	N	PMMoV
Influenza A	WastewaterSCAN	ddPCR	INFA1+INFA2	PMMoV
RSV	WastewaterSCAN	ddPCR	RSV-A+RSV-B	PMMoV
Boise RWF, ID (372)				
SARS-CoV-2	WastewaterSCAN	ddPCR	N	PMMoV
Influenza A	WastewaterSCAN	ddPCR	INFA1, INFA1+INFA2	PMMoV
RSV	WastewaterSCAN	ddPCR	RSV-A+RSV-B	PMMoV

Supplementary Table S6. Correlation between wastewater metagenomic sequencing and clinical testing data. Spearman correlation coefficients (R) comparing wastewater metagenomic sequencing (WW-MGS) abundance from Boston Deer Island Treatment Plant (DITP) with positive test counts from Massachusetts General Hospital (MGH) clinical diagnostic testing. Clinical data represent aggregated weekly positive test counts from multiplex nucleic acid amplification tests. For WW-MGS data, we report correlations using two normalization approaches: PMMoV (pepper mild mottle virus)-normalized relative abundance and ToBRFV (tomato brown rugose fruit virus)-normalized relative abundance. Correlations were calculated using MMWR week-aggregated data with centered 5-week smoothing. *p < 0.05, **p < 0.01, ***p < 0.001. MU, University of Missouri; SB, SecureBio.

		Spearman correlation coefficient (R)					
		PMMoV-normalized sequencing			ToBRFV-normalized sequencing		
Sampling site	**N (weeks)**	**SARS-CoV-2**	**Influenza A**	**RSV**	**SARS-CoV-2**	**Influenza A**	**RSV**
MU-sequenced							
Boston DITP North, MA	47	0.78***	0.65***	0.84***	0.70***	0.65***	0.83***
Boston DITP South, MA	47	0.78***	0.71***	0.75***	0.77***	0.69***	0.75***
SB-sequenced							
Boston DITP North, MA	26	0.67***	0.92***	0.83***	0.61***	0.92***	0.83***
Boston DITP South, MA	26	0.20	0.75***	0.79***	0.02	0.75***	0.79***

References

Adriaenssens, Evelien M., Kata Farkas, Christian Harrison, David L. Jones, Heather E. Allison, and Alan J. McCarthy. 2018. “Viromic Analysis of Wastewater Input to a River Catchment Reveals a Diverse Assemblage of RNA Viruses.” *mSystems* 3 (3). https://doi.org/10.1128/mSystems.00025-18.

Bellekom, Ben, Catherine Troman, Shannon Fitz, Joyce Odeke Akello, Nicholas C. Grassly, and Alexander G. Shaw. 2026. “Comparison of the Sensitivity of Targeted and Untargeted (metagenomic) Methods for the Detection of Viral Pathogens in Wastewater.” *The Science of the Total Environment* 1013 (181333): 181333.

Bibby, Kyle, and Jordan Peccia. 2013. “Identification of Viral Pathogen Diversity in Sewage Sludge by Metagenome Analysis.” *Environmental Science & Technology* 47 (4): 1945–1951.

Boehm, Alexandria B., Marlene K. Wolfe, Amanda L. Bidwell, et al. 2024. “Human Pathogen Nucleic Acids in Wastewater Solids from 191 Wastewater Treatment Plants in the United States.” *Scientific Data* 11 (1): 1141.

Boehm, Alexandria B., Marlene K. Wolfe, Amanda L. Bidwell, et al. 2026. “Pathogen Nucleic Acids Data in Wastewater Solids from 147 Treatment Plants in the United States: 2024-2025.” *Data in Brief* 65 (112503): 112503.

Brinch, Christian, Pimlapas Leekitcharoenphon, Ana S. R. Duarte, Christina A. Svendsen, Jacob D. Jensen, and Frank M. Aarestrup. 2020. “Long-Term Temporal Stability of the Resistome in Sewage from Copenhagen.” *mSystems* 5 (5). https://doi.org/10.1128/mSystems.00841-20.

Cantalupo, Paul G., Byron Calgua, Guoyan Zhao, et al. 2011. “Raw Sewage Harbors Diverse Viral Populations.” *mBio* 2 (5). https://doi.org/10.1128/mBio.00180-11.

Child, Harry T., George Airey, Daniel M. Maloney, et al. 2023. “Comparison of Metagenomic and Targeted Methods for Sequencing Human Pathogenic Viruses from Wastewater.” *mBio* 14 (6): e0146823.

Crits-Christoph, Alexander, Rose S. Kantor, Matthew R. Olm, et al. 2021. “Genome Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-2 Variants.” *mBio* 12 (1). https://doi.org/10.1128/mBio.02703-20.

Fernandez-Cassi, Xavier, and Tamar Kohn. 2024. “Comparison of Three Viral Nucleic Acid Preamplification Pipelines for Sewage Viral Metagenomics.” *Food and Environmental Virology* 16 (3): 1–22.

Fernandez-Cassi, X., N. Timoneda, S. Martínez-Puchol, et al. 2018. “Metagenomics for the Study of Viruses in Urban Sewage as a Tool for Public Health Surveillance.” *The Science of the Total Environment* 618 (March): 870–880.

Grimm, Simon L., Jason A. Rothman, William J. Bradshaw, et al. 2025. “Deep Metatranscriptomic Sequencing Data of Wastewater from Los Angeles, USA, 2023-2024.” *Scientific Data* 13 (1): 158.

Guajardo-Leiva, Sergio, Jonás Chnaiderman, Aldo Gaggero, and Beatriz Díez. 2020. “Metagenomic Insights into the Sewage RNA Virosphere of a Large City.” *Viruses* 12 (9):

1050.

Guajardo-Leiva, Sergio, Beatriz Díez, Cecilia Rojas-Fuentes, et al. 2025. “From Sewage to Genomes: Expanding Our Understanding of the Urban and Semi-Urban Wastewater RNA Virome.” *Environmental Research* 276 (121509): 121509.

Gulino, K., J. Rahman, M. Badri, J. Morton, R. Bonneau, and E. Ghedin. 2020. “Initial Mapping of the New York City Wastewater Virome.” *mSystems* 5 (3). https://doi.org/10.1128/mSystems.00876-19.

Hjelmsø, Mathis Hjort, Maria Hellmér, Xavier Fernandez-Cassi, et al. 2017. “Evaluation of Methods for the Concentration and Extraction of Viruses from Sewage in the Context of Metagenomic Sequencing.” *PloS One* 12 (1): e0170199.

Kaufman, Jeff. 2025. “How Much Data From a Sequencing Run?” SecureBio, June 25. https://data.securebio.org/jefftk-notebook/how-much-data-from-a-sequencing-run.

Maritz, Julia M., Theresa A. Ten Eyck, S. Elizabeth Alter, and Jane M. Carlton. 2019. “Patterns of Protist Diversity Associated with Raw Sewage in New York City.” *The ISME Journal* 13 (11): 2750–2763.

Martínez-Puchol, Sandra, Marta Rusiñol, Xavier Fernández-Cassi, et al. 2020. “Characterisation of the Sewage Virome: Comparison of NGS Tools and Occurrence of Significant Pathogens.” *The Science of the Total Environment* 713 (136604): 136604.

McCall, Camille, Ryan A. Leo Elworth, Kristine M. Wylie, et al. 2023. “Targeted Metagenomic Sequencing for Detection of Vertebrate Viruses in Wastewater for Public Health Surveillance.” *ACS ES&T Water*, ahead of print, August 9. https://doi.org/10.1021/acsestwater.3c00183.

Ng, Terry Fei Fan, Rachel Marine, Chunlin Wang, et al. 2012. “High Variety of Known and New RNA and DNA Viruses of Diverse Origins in Untreated Sewage.” *Journal of Virology* 86 (22): 12161–12175.

Nieuwenhuijse, David F., Bas B. Oude Munnink, My V. T. Phan, et al. 2020. “Setting a Baseline for Global Urban Virome Surveillance in Sewage.” *Scientific Reports* 10 (1): 13748.

Rothman, Jason A., Theresa B. Loveless, Joseph Kapcia 3rd, et al. 2021. “RNA Viromics of Southern California Wastewater and Detection of SARS-CoV-2 Single-Nucleotide Variants.” *Applied and Environmental Microbiology* 87 (23): e0144821.

Spurbeck, Rachel R., Lindsay A. Catlin, Chiranjit Mukherjee, Anthony K. Smith, and Angela Minard-Smith. 2023. “Analysis of Metatranscriptomic Methods to Enable Wastewater-Based Biosurveillance of All Infectious Diseases.” *Frontiers in Public Health* 11 (March). https://doi.org/10.3389/fpubh.2023.114527510.3389/fpubh.2023.1145275.s001.

Stockdale, Stephen R., Adam M. Blanchard, Amit Nayak, et al. 2023. “RNA-Seq of Untreated Wastewater to Assess COVID-19 and Emerging and Endemic Viruses for Public Health Surveillance.” *The Lancet Regional Health. Southeast Asia* 14 (100205): 100205.

Tierney, Braden T., Jonathan Foox, Krista A. Ryon, et al. 2024. “Towards Geospatially-Resolved Public-Health Surveillance via Wastewater Sequencing.” *Nature Communications* 15 (1): 8386.

Wang, Hao, Julianna Neyvaldt, Lucica Enache, et al. 2020. “Variations among Viruses in Influent

Water and Effluent Water at a Wastewater Plant over One Year as Assessed by Quantitative PCR and Metagenomics." *Applied and Environmental Microbiology* 86 (24). https://doi.org/10.1128/AEM.02073-20.

Wetterstrand, K. A. 2019. "DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)." Genome.gov, NHGRI, March 13. https://www.genome.gov/about-genomics/fact-sheets/DNA-Sequencing-Costs-Data.

Worp, Nathalie, David F. Nieuwenhuijse, Ray W. Izquierdo-Lara, et al. 2025. "Unveiling the Global Urban Virome through Wastewater Metagenomics." *Nature Communications* 16 (1): 10707.

Wyler, Emanuel, Chris Lauber, Artür Manukyan, et al. 2024. "Pathogen Dynamics and Discovery of Novel Viruses and Enzymes by Deep Nucleic Acid Sequencing of Wastewater." *Environment International* 190 (108875): 108875.