SUPPORTING INFORMATION

Pathogen and Indicator Trends in Southern Nevada Wastewater during and after the COVID-19 Pandemic

Katherine Crank[a], Katerina Papp[a], Casey Barber[a,b], Kai Chung[a], Emily Clements[a], Wilbur Frehner[a], Deena Hannoun[a], Travis Lane[a], Christina Morrison[a], Bonnie Mull[c], Edwin Oh[d], Phillip Wang[a], and Daniel Gerrity[a*]

[a] Southern Nevada Water Authority, P.O. Box 99954, Las Vegas, NV, 89193, USA

[b] School of Public Health, University of Nevada, Las Vegas, 4700 S. Maryland Parkway, Suite 335, Mail Stop 3063, Las Vegas, NV 89119, USA

[c] BCS Laboratories, Inc 4609 NW 6th St, STE A Gainesville, FL, 32609, USA

[d] Laboratory of Neurogenetics and Precision Medicine, University of Nevada Las Vegas, 4505 S. Maryland Parkway, Las Vegas, NV 89154

*Corresponding author. Email: daniel.gerrity@snwa.com

Table of Contents

Table S1. Number of samples collected from each facility.

	Surveillance Sampling Campaign (qPCR) March 2020 – December 2023			Sampling Sub-campaign January – December 2023	
Facility	**Archived samples March 2020-September 2022**	**Fresh samples October 2022-December 2023**	**Total samples (archived and fresh)**	**Protozoa**	**Virus culture**
1	149	38	187	13	7
2	131	38	169	12	7
3	86	38	124	12	7
4 [a, b]	74	37	111	12	7
4A [b]	97	38	135	0	7
4B	87	38	125	0	7
5	103	38	141	12	7
6	82	38	120	12	7
Total	**809**	**303**	**1112[c]**	**73**	**56**

[a] Facility 4 is the 24-hour composite of Facility 4A (west trunk line) and Facility 4B (east trunk line).
[b] Facility 4A (and by default Facility 4) receives solids and bypass flows from Facility 2.
[c] Fewer samples were analyzed for AdV due to insufficient quantities of archived nucleic acids/volumes of wastewater concentrates for certain samples. A total of 1107 samples (instead of 1112 samples) were assayed for AdV.

Table S2. Summary of sample starting volumes, equivalent sample volumes, and recoveries of non-excluded samples.

Concentration method	Starting volume (L)	BCoV recovery in archived samples	BCoV recovery in fresh samples	BCoV recovery in all samples	ESV for RNA targets (mL/reaction)	ESV for DNA targets[a] (mL/reaction)
CC[b] (n = 764)	0.25 ± 0.02	0.16 ± 0.18	0.13 ± 0.11	0.15 ±0 .17	0.41 ± 0.29	1.14 ± 0.69
HFUF[c] (n = 321)	10 ± 0	0.39 ± 0.28	0.43 ± 0.32	0.40 ± 0.29	0.24 ± 0.08	0.69 ± 0.16
HFUF + CC (n = 27)	9.7 ± 1.81	0.02 ± 0.01	N/A	0.02 ± 0.01	3.13 ± 1.19	8.87 ± 3.05
Average	**-**	**0.24 ± 0.25**	**0.21 ± 0.23**	**0.23 ± 0.24**	**0.41 ± 0.53**	**0.98 ± 1.17**

Values reported as mean ± standard deviation
[a]For samples where archived DNA volume was limited, DNA assays were run using cDNA, causing the average ESV to be lower than if the ESV was calculated with the averages of all other parameters.
[b]CC = Centricon centrifugal ultrafiltration
[c]HFUF = hollow fiber ultrafiltration

Table S3. Primers and probes for qPCR assays.

Target	Primer/Probe[a] Sequence (5'-3')	Final concentration per reaction (µM)	Cycling Conditions	Reference
Adenovirus[b]	F: AACTTTCTCTCTTAATAGACGCC	0.4	2 min at 95°C, 45x: 7s at 95°C, 45s at 56°C	[1, 2,] In house
	R: AGGGGGCTAGAAAACAAAA	0.4		
	P: CTGACACGGGCACTCTTCGC	0.2		
Bovine Coronavirus (spike)	F: CTGGAAGTTGG TGGAGTT	0.4	30 s at 95°C, 45x: 5s at 95°C, 30s at 60°C	[3]
	R: ATTATCGGCCTAACATACATC	0.4		
	P: CCTTCATATCTATACACATCAAGTTGTT	0.2		
CP56[c]	F: CAGAAGTACAAACTCCTAAAAAACGTAGAG	0.4	5 min at 95°C, 45x: 15s at 95°C, 60s at 60°C	[4]
	R: GATGACCAATAAACAAGCCATTAGC	0.4		
	P(HEX): AATAACGATTTACGTGATGTAAC	0.2		
Enterovirus	F: CCTCCGGCCCCTGAATG	0.3	5 min at 95°C, 45x: 15s at 95°C, 60s at 60°C	[2, 5]
	R: ACCGGATGGCCAATCCAA	0.9		
	P: CGGAACCGACTACTTTGGGTGTCCGT	0.1		
HF183[d]	F: ATCATGAGTTCACATGTCCG	0.6	5 min at 95°C, 45x: 15s at 95°C, 60s at 60°C	[6]
	R: CGTAGGAGTTTGGACCGTGT	0.6		
	P: CTGAGAGGAAGGTCCCCCACATTGGA	0.2		
Norovirus GIA[e]	F: CCATGTTCCGTTGGATGC	0.5	2 min at 95°C, 45x: 7s at 95°C, 45s at 56°C	[2, 7, 8,] In house
	R: TCCTTAGACGCCATCATCAT	0.5		
	P: AGATYGCGITCICCTGTCCA	0.1		

Target	Primer/Probe[a] Sequence (5'-3')	Final concentration per reaction (µM)	Cycling Conditions	Reference
Norovirus GIB[e]	F: CGCTGGATGCGNTTCCAT	0.4	5 min at 95°C, 45x: 15s at 95°C, 60s at 60°C	[2, 5, 9]
	R: CCTTAGACGCCATCATCATTTAC	0.4		
	P: TGGACAGGAGAYCGCRATCT	0.2		
Norovirus GII	F: ATGTTCAGRTGGATGAGRTTCTCWGA	0.4	5 min at 95°C, 45x: 15s at 95°C, 60s at 60°C	[2, 5]
	R: TCGACGCCATCTTCATTCACA	0.4		
	P: AGCACGTGGGAGGGCGATCG	0.2		
Pepper Mild Mottle Virus (probe-based)	F: GAGTGGTTTGACCTTAACGTTTGA	0.9	30 s at 95°C, 45x: 5s at 95°C, 30s at 60°C	[10]
	R: TTGTCGGTTGCAATGCAAGT	0.9		
	P: CCTACCGAAGCAAATG	0.25		

[a] Probes use the FAM fluorophore except for crAssphage 56 which uses the HEX fluorophore
[b] Targets serotypes 40/41
[c] crAssphage 56 or CPQ_56
[d] Human-specific *Bacteroides*
[e] Norovirus GIA and GIB target the same strain of NoV GI, accession number M87661 ([8, 9])

Table S4. Standard gBlock sequences and lengths.

Target	Sequence (5'-3')	Length
Adenovirus (40/41)	CACTGTTAATGCAAACAACGAACTTTCTCTCTTAATAGACGCCCCACTT AATGCTGACACGGGCACTCTTCGCCTTCAAAGTGCTGCACCTCTTGGAC TAGTGGACAAAACACTAAAAGTTTTGTTTTCTAGCCCCCTCTATCTAGA TAATAACTTTC	158bp
Bovine Coronavirus	GAATAGTATCAGGTTGTTTATTAGAACTGGAAGTTGGTGGAGTTTCAAC CCAGAAACAAACAACTTGATGTGTATAGATATGAAGGGAAGGATGTAT GTTAGGCCGATAATTGAGGACTACCATACCCTTACGGTCACAATA	142bp
crAssphage[a]	CAGAAGTACAAACTCCTAAAAAACGTAGAGGTAGAGGTATTAATAACG ATTTACGTGATGTAACTCGTAAAAAGTTTGATGAACGTACTGATTGTAA TAAAGCTAATGGCTTGTTTATTGGTCATC	126bp
Enterovirus	TCTATTGAGCTAGTTAGTAGTCCTCCGGCCCCTGAATGCGGCTAATCCT AACTGCGGAGCACATGCCCTCAACCCAGAGGGTAGTGTGTCGTAACGG GCAACTCTGCAGCGGAACCGACTACTTTGGGTGTCCGTGTTTCCTTTTAT TCTTACATTGGCTGCCTATGGTGACAATCGCAGAATTGTTACCATATAG CTATTGGATTGGCCATCCGGTGTGCAATAGAGCTATTATAT	237bp
HF183[a]	GAAGATTAATCCAGGATGGGATCATGAGTTCACATGTCCGCATGATTAA AGGTATTTTCCGGTAGACGATGGGGATGCGTTCCATTAGATAGTAGGCG GGGTAACGGCCCACCTAGTCAACGATGGATAGGGGTTCTGAGAGGAAG GTCCCCCACATTGGAACTGAGACACGGTCCAAACTCCTACGGGAGGCA GCAGTGAGGAATA	207bp
Norovirus GIA[b]	TCAAGAAGTCAAAGAGGGGGGGCTTGAAATCTACATTCCTGGCTGGCA GGCCATGTTCCGTTGGATGCGAGTCCATGACCTAAGTTTGAGATCGCGC TCACCTGTCCATCTCCTGCCCGATTATGTAAATGATGATGGCGTCTAAG GACGCCCCAACAAACATGGATGGCACCAGTGGTGCCGGTCAGCTGGTT CCAGAGGCCAGTACAGCTG TCAAGAAGTCAAAGAGGGGGGGCTTGAAATCTACATTCCTGGCTGGCA GGCCATGTTCCGTTGGATGCGAGTCCATGACCTAAGTTTGAGATTGCGA TCCCCTGTCCATCTCCTGCCCGATTATGTAAATGATGATGGCGTCTAAG GACGCCCCAACAAACATGGATGGCACCAGTGGTGCCGGTCAGCTGGTT CCAGAGGCCAGTACAGCTG TCAAGAAGTCAAAGAGGGGGGGCTTGAAATCTACATTCCTGGCTGGCA GGCCATGTTCCGTTGGATGCGAGTCCATGACCTAAGTTTGAGATTGCGA TCACCTGTCCATCTCCTGCCCGATTATGTAAATGATGATGGCGTCTAAG GACGCCCCAACAAACATGGATGGCACCAGTGGTGCCGGTCAGCTGGTT CCAGAGGCCAGTACAGCTG	213 bp

Target	Sequence (5'-3')	Length
Norovirus GIB[b]	TCAAGAAGTCAAAGAGGGGGGGCTTGAAATCTACATTCCTGGCTGGCA GGCCATGTTCCGCTGGATGCGATTCCATGACCTAAGTTTGTGGACAGGA GACCGCAATCTCCTGCCCGATTATGTAAATGATGATGGCGTCTAAGGAC GCCCCAACAAACATGGATGGCACCAGTGGTGCCGGTCAGCTGGTTCCA GAGGCCAGTACAGCTG	210 bp
	TCAAGAAGTCAAAGAGGGGGGGCTTGAAATCTACATTCCTGGCTGGCA GGCCATGTTCCGCTGGATGCGGTTCCATGACCTAAGTTTGTGGACAGGA GATCGCGATCTCCTGCCCGATTATGTAAATGATGATGGCGTCTAAGGAC GCCCCAACAAACATGGATGGCACCAGTGGTGCCGGTCAGCTGGTTCCA GAGGCCAGTACAGCTG	
	TCAAGAAGTCAAAGAGGGGGGGCTTGAAATCTACATTCCTGGCTGGCA GGCCATGTTCCGCTGGATGCGGTTCCATGACCTAAGTTTGTGGACAGGA GACCGCGATCTCCTGCCCGATTATGTAAATGATGATGGCGTCTAAGGAC GCCCCAACAAACATGGATGGCACCAGTGGTGCCGGTCAGCTGGTTCCA GAGGCCAGTACAGCTG	
Norovirus GII[b]	CGTACCCAGACAAGAGCCAATGTTCAGATGGATGAGATTCTCAGATCTG AGCACGTGGGAGGGCGATCGCAATCTGGCTCCCAGTTTTGTGAATGAA GATGGCGTCGAGTGACGCCAACCCATCTGCG	128 bp
	CGTACCCAGACAAGAGCCAATGTTCAGGTGGATGAGGTTCTCTGATCTG AGCACGTGGGAGGGCGATCGCAATCTGGCTCCCAGTTTTGTGAATGAA GATGGCGTCGAGTGACGCCAACCCATCTGCG	
	CGTACCCAGACAAGAGCCAATGTTCAGATGGATGAGGTTCTCTGATCTG AGCACGTGGGAGGGCGATCGCAATCTGGCTCCCAGTTTTGTGAATGAA GATGGCGTCGAGTGACGCCAACCCATCTGCG	
	CGTACCCAGACAAGAGCCAATGTTCAGGTGGATGAGGTTCTCAGATCTG AGCACGTGGGAGGGCGATCGCAATCTGGCTCCCAGTTTTGTGAATGAA GATGGCGTCGAGTGACGCCAACCCATCTGCG	
Pepper Mild Mottle Virus (probe-based)	GAGTGGTTTGACCTTAACGTTTGAGAGGCCTACCGAAGCAAATGTCGCA CTTGCATTGCAACCGACAA	68bp
Pepper Mild Mottle Virus (SYBR-based)	ATTAGGCGTAGATCCATTGGTGGCAGCAAAGGTAATGGTAGCTGTGGTT TCAAATGAGAGTGGTTTGACCTTAACGTTTGAGAGGCCTACCGAAGCAA ATGTCGCACTTGCATTGCAACCGACAATTACATCAAAGGAGG	140 bp

*Red font indicates single polymorphism difference

Text S1. Master standard curve.

The gblocks were prepared following manufacturer's instructions (IDT), and the standard stocks were quantified using digital PCR. For sample analysis, each qPCR run included 3 replicate standard curves. Standard curves were created fresh from the stock every three runs, or more frequently if curves were visually identifiable as poorly performing (e.g., poor efficiency or lack of amplification in the lower dilutions). For final sample quantification, standard curves were combined across all qPCR runs. All Cqs associated with a specific dilution were used to fit the linear regression; replicates were not averaged for each run before being added to the master curve. Advanced methods for combining multiple standard curves (such as using Bayesian analyses as suggested by Sivaganesan et al. 2010[11]) were not employed as the standard errors associated with the slope and intercept values from each curve were acceptably small due to the large number of replicates included (data not shown). Individual run standard curves were omitted from the master curve if they contained any visually-identified outliers. Master standard curves for each assay are shown in Figure S1. Norovirus GIA, GIB, and GII standards were pooled to account for polymorphisms.

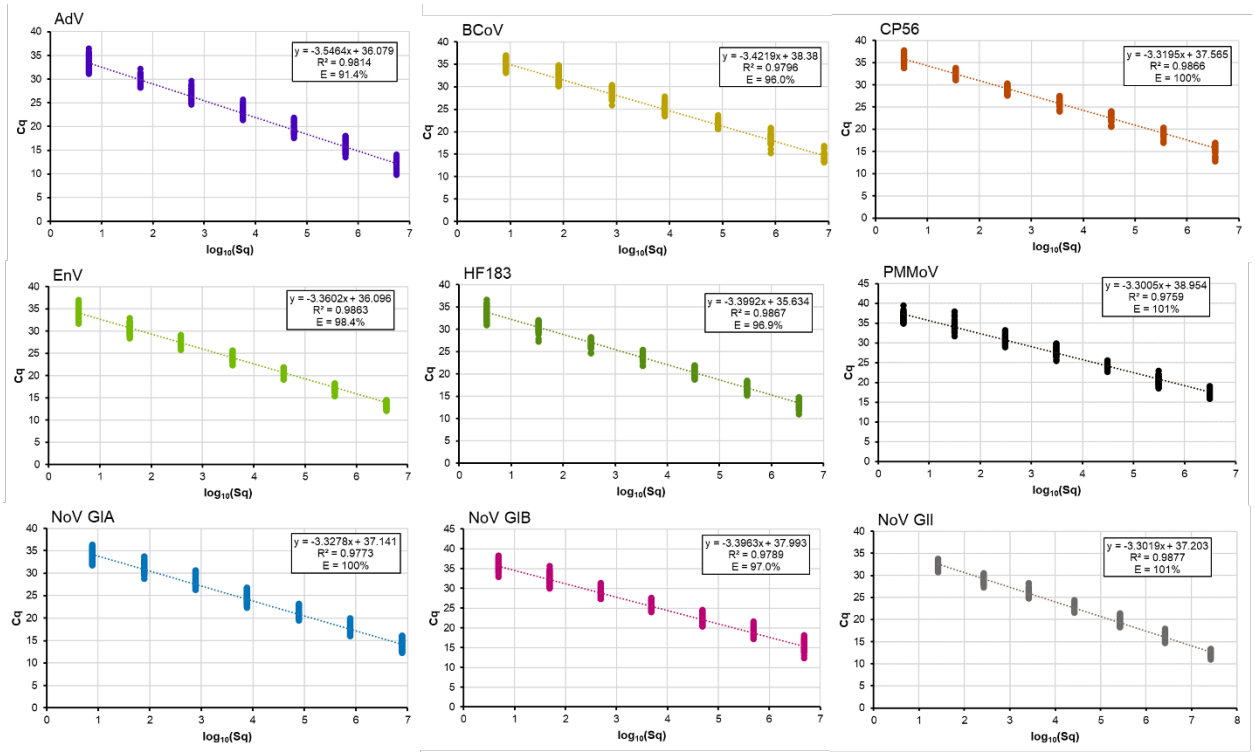

Figure S1. Master standard curves for qPCR assays.

Text S2. Equivalent sample volumes, limits of quantification, and duplex optimization.

Equivalent Sample Volume (ESV)

Equivalent sample volume (ESV) is the amount of the original sample that is analyzed in each qPCR reaction, cell culture well/flask, microscope slide, etc. For the qPCR assays, the ESV calculation differs for each concentration method (hollow fiber ultrafiltration (HFUF) vs. Centricon (CC)) and assay type (RNA vs DNA). The qPCR ESV equations are shown below in Equations S1-7, and a summary of ESVs and sample-specific volumes are shown in Table S4. Similar information for culturable viruses and *Giardia*/*Cryptosporidium* is described later.

(Eq. S1)

$$\text{ESV (HFUF+CC, RNA)} = \frac{\text{vol. cDNA template } \left(\frac{\mu L}{rxn}\right)}{\text{cDNA synthesis vol. } \left(\frac{\mu L}{rxn}\right)} \times \frac{\text{vol. RNA used for cDNA synthesis } (\mu L)}{\text{RNA extract vol.}(\mu L)} \times \frac{\text{CC conc. extracted (mL)}}{\text{tot. CC conc. vol. (mL)}} \times \frac{\text{vol. through CC (mL)}}{\text{HFUF conc. vol. (mL)}} \times 10^4 \text{ mL wastewater}$$

(Eq. S2)

$$\text{ESV (HFUF+CC, DNA)} = \frac{\text{vol. DNA template } \left(\frac{\mu L}{rxn}\right)}{\text{DNA extract vol.}(\mu L)} \times \frac{\text{CC conc. extracted (mL)}}{\text{tot. CC conc. vol. (mL)}} \times \frac{\text{vol. through CC (mL)}}{\text{HFUF conc. vol. (mL)}} \times 10^4 \text{ mL wastewater}$$

(Eq. S3)

$$\text{ESV (CC, RNA)} = \frac{\text{vol. cDNA template } \left(\frac{\mu L}{rxn}\right)}{\text{cDNA synthesis vol } \left(\frac{\mu L}{rxn}\right)} \times \frac{\text{vol. RNA used for cDNA synthesis } (\mu L)}{\text{RNA extract vol.}(\mu L)} \times \frac{\text{CC conc. extracted (mL)}}{\text{tot. CC conc. vol (mL)}} \times \text{vol. through CC (mL)}$$

(Eq. S4)

$$\text{ESV (CC, DNA)} = \frac{\text{vol. DNA template } \left(\frac{\mu L}{rxn}\right)}{\text{DNA extract vol.}(\mu L)} \times \frac{\text{CC conc. extracted (mL)}}{\text{tot. CC conc. vol. (mL)}} \times \text{vol. through CC (mL)}$$

(Eq. S5)

$$\text{ESV (HFUF, RNA)} = \frac{\text{vol. cDNA template } \left(\frac{\mu L}{rxn}\right)}{\text{cDNA synthesis vol } \left(\frac{\mu L}{rxn}\right)} \times \frac{\text{vol. RNA used for cDNA synthesis } (\mu L)}{\text{RNA extract vol.}(\mu L)} \times \frac{\text{conc. extracted (mL)}}{\text{tot. HFUF conc. vol. -pellet vol.}} \times 10^4 \text{ mL wastewater}$$

(Eq. S6)

$$\text{ESV (HFUF, DNA)} = \frac{\text{vol. DNA template } \left(\frac{\mu L}{rxn}\right)}{\text{DNA extract vol.}(\mu L)} \times \frac{\text{conc. extracted (mL)}}{\text{tot. HFUF conc. vol. -pellet vol.}} \times 10^4 \text{ mL wastewater}$$

(Eq. S7)

$$\log_{10}\text{gc per liter} = \frac{\frac{Sq}{ESV} \times 1000}{\text{recovery}}$$, where Sq = starting quantity from the qPCR standard curve.

qPCR Limit of Quantification (LoQ)

The limit of quantification (LoQ) was determined for each assay as previously described by Gerrity et al. (2021)[12]. Briefly, known "test" dilutions of standards ranging from 1000 to 0.1 gc per reaction for non-NoV standards, 3000 to 0.3 gc per reaction for NoV GIA and NoV GIB, and 4000 to 0.4 gc per reaction for NoV GII were quantified in six to nine replicate reactions for each dilution. The lowest test dilution demonstrating consistent amplification was used in a one-sided t-test to generate the LoQ with 99% confidence, following the U.S. Environmental Protection Agency procedure for calculating the method detection limit (Method Detection Limit Procedure, Revision 2, 2016 – based upon 40 CFR 136, Appendix B). The Cq-based and concentration-based LoQs (calculated with an average ESV and master standard curve) are reported in Table S5. Concentration-based LoQs are reported to facilitate literature comparison; however, categorizations for data analysis were performed using sample-specific average Cqs. Molecular assays were run in triplicate and concentrations were averaged across replicates; if there was one non-detect (ND), it was excluded from the average. Single amplifications (i.e., amplification in only one of three replicates), either above or below the LoQ, were considered NDs. If one, two, or three replicates amplified later than LoQ, the instrument provided Cqs were included in the averaging calculation. A "detection" was defined in this study as amplification above the baseline fluorescence threshold and displaying classic sigmoidal curvature. For statistical analysis and reporting purposes, the detection limit was defined as the highest observed Cq for a given assay.

Table S5. Cq-based and concentration-based limits of quantification.

Target	LoQ (Cq)	LoD (Cq)	Average LoQ (log_{10} gc/L) for HFUF	Average LoQ (log_{10} gc/L) for CC	Average LoQ (log_{10} gc/L) for HFUF + CC	Average LoD (log_{10} gc/L)	Average LoQ (log_{10} gc/L)
Adenovirus	33.60	37.97	4.72	4.92	3.92	3.28	4.52
Bovine Coronavirus	32.85	44.55	6.04	6.24	5.13	2.38	5.80
crAssphage	31.16	37.68	5.95	6.14	5.15	3.78	5.75
Enterovirus	32.09	41.96	5.61	5.81	4.70	2.44	5.37
HF183[a]	30.88	N/A	5.42	5.61	4.62	3.82	5.22
Norovirus GIA	34.17	38.13	5.31	5.51	4.40	3.89	5.08
Norovirus GIB	32.85	42.42	5.94	6.14	5.02	2.88	5.70
Norovirus GII	34.36	42.87	5.28	5.48	4.37	2.47	5.04
PMMoV (probe)	33.49	39.41	6.08	6.28	5.16	4.04	5.84

LoD and LoQ concentration values are not recovery corrected. N/A = not applicable as they were not determined.
[a]HF183 LoD was set to 1 copy per well.

HF183 & CP56 Duplex Assay

The slope, y-intercept, R^2, efficiency (E), and LoQ values based on the corresponding standard curves are summarized in Table S6. The duplex and monoplex assays both demonstrated acceptable linearity and dynamic ranges. Standard curves for monoplex and duplex assay comparison are shown in Figure S2.

Table S6. Comparison of standards in monoplex and duplex assays.

	Monoplex assay (Mean Cq)		Duplex Assay (Mean Cq)	
Standard copy/reaction	**CP56**	**HF183**	**CP56**	**HF183**
1.00E+07	16.40	11.81	15.14	12.54
1.00E+06	19.64	15.08	18.24	15.38
1.00E+05	23.17	18.62	22.23	19.78
1.00E+04	26.81	22.28	25.30	22.77
1.00E+03	29.81	25.50	28.66	26.31
1.00E+02	32.72	28.34	31.71	29.70
1.00E+01	35.11	31.00	34.65	33.13
Slope	-3.175	-3.264	-3.282	-3.467
Intercept	38.938	34.901	38.263	36.663
R^2	0.9942	0.9927	0.9954	0.9962
E (%)	107%	102%	102%	94%
LOQ	N/A[a]	N/A	31.16117	30.88161

[a]N/A = not applicable since these were not performed.

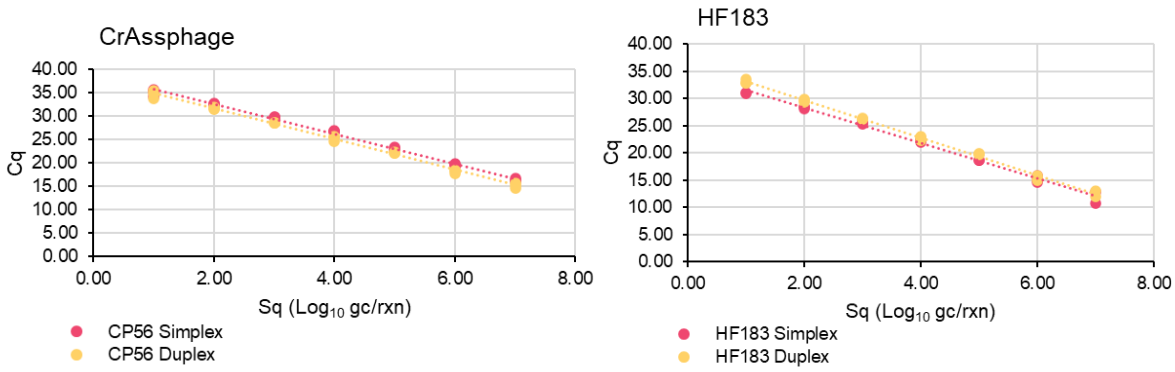

Figure S2. Comparison of standard curves of monoplex and duplex assays.

Text S3. Additional methods information and summary of recovery for culturable AdV, culturable EnV, *Cryptosporidium,* and *Giardia*.

Recovery efficiency for culturable AdV and EnV was determined as the average of recoveries of spiked MS2 and phiX174, consistent with Pecson et al. (2022)[2]. Several MS2 recoveries (n = 4) were greater than 100%, and these recoveries were set to 100% for averaging. Sample recovery across 56 samples averaged 34% ± 21% and ranged between 8% and 93% (after averaging MS2 and phiX174). Concentrations were calculated according to Eqs. S8 and S9.

(**Eq. S8)**

$$\text{Concentration } \left(\frac{\text{MPN}}{\text{L}}\right) = \frac{\text{MPN}}{\text{ESV}_{\text{sample}}\ (\text{L}) \times \text{Recovery}}$$

(**Eq. S9)**

$$\text{ESV (L)} = \frac{\text{FCSV}_{analyzed}\ (\text{mL})}{\text{FSCV}_{total}\ (\text{mL})} \times 1\ \text{L wastewater}$$

where, $\text{FCSV}_{\text{total}}$ is the final concentrated sample volume, or the volume of resuspended pellet after polyethylene glycol precipitation and chloroform extraction (4.96 mL ± 0.57 mL across all samples), and $\text{FCSV}_{\text{analyzed}}$ is the volume of resuspended pellet used in each cell culture assay (constant at 1.11 mL).

Concentrations of *Cryptosporidium* and *Giardia* were calculated according to Eq. S10, and recovery was calculated according to Eq. S11. For *Giardia,* recovery across 73 samples averaged 55% ± 21% and ranged between 3% and 90%. For *Cryptosporidium*, recovery across 73 samples averaged 31% ± 19% and ranged between 3% and 91%. ESVs for all samples were constant at 0.1 L, since the entire sample volume was analyzed.

(**Eq. S10)**

$$\text{Concentration}\left(\frac{\text{(oo)cysts}}{\text{L}}\right)=\frac{\text{Native (oo)cysts counted}}{\text{ESV}_{\text{sample}}\text{ (L)} \times \text{Recovery}}$$

(**Eq. S11)**

$$\text{Recovery (\%)} = \frac{\text{Number of spiked (oo)cysts counted}}{\text{Total number of spiked (oo)cysts}} \times 100\%$$

where, "spiked (oo)cysts" represent a ColorSeed^{TM} matrix spike[13] consisting of 100 *Cryptosporidium* oocysts and 100 *Giardia* cysts.

Table S7. Methods of recovery estimation.

Description of Method	Abbreviation
Re-quantified facility-specific recovery average	R1
Original facility-specific recovery average	R2
Original recovery value	R3
Overall re-quantified recovery average (all facilities pooled)	R4
Overall original recovery average (all facilities pooled)	R5
PMMoV degradation term, with PMMoV concentrations calculated via different assays; SYBR-based concentrations were converted to probe-based concentrations using linear regression (Figure S3). Values over 100% recovery were set to 100%.	PM1
PMMoV degradation term, with PMMoV concentrations calculated via different assays; SYBR-based concentrations were converted to probe-based concentrations by subtracting the overall average difference between the two methods (0.26 $\log_{10}$ gc/L)	PM2
PMMoV degradation term, no adjustment for differing assays	PM3
Combinations of the above	R* × P*
Estimation with supervised machine learning (SML) using a form of decision tree modeling (*cubist*)	SML

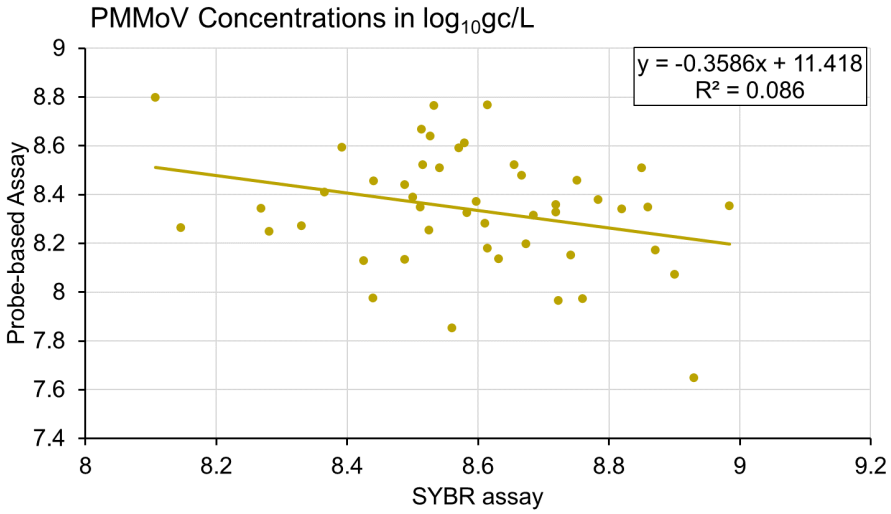

Figure S3. Linear regression of the two PMMoV assays (probe vs. SYBR) run on 48 paired samples. Negative correlation indicates that the SYBR assay may have had non-specific amplification compared to the probe-based assay.

Text S4. Supervised machine learning.

For the SML approach, non-detects were imputed for all markers except BCoV as follows. The $\log_{10}$-transformed non-recovery-corrected data were fit to a censored normal distribution using the *fitdistcens* function in R with maximum likelihood estimation (MLE). Then, for each non-detect, a uniform distribution was created between 0 and the cumulative distribution function (CDF) probability for the sample-specific LoD. The CDF of the LoD is the probability that a value in the distribution is less than or equal to the LoD. The uniform distribution was sampled, and the resultant probability was transformed back into a concentration using the corresponding quantiles of the fitted distribution. Detection status of each target was preserved as a discrete variable.

Cubist is an algorithm that generates rule-based predictive models to capture the relationship between the input variables and the target. It divides the data into subsets via decision trees, which are then used with a regression model to generate rules. The rules are then pruned to simplify the model and improve the generalizability. *Cubist* models use committees, which are similar to boosting, where iterative models are created to correct for errors in previous trees. The final prediction is an average of the predictions of each model, sometimes combined with a nearest neighbor algorithm. The number of committees and number of neighbors are the two main hyperparameters that can be adjusted. The interpretability of *cubist* rules is a strength, allowing more transparency than other machine learning algorithms, so it was selected.

The following variables were included. Markers appended with "_nc" refer to non-corrected $\log_{10}$-transformed concentrations, with detected data included. Markers appended with ".esv" refer to the equivalent sample volume (ESV) for each assay. These differ between the DNA and RNA assays, although some DNA assays have the same ESV as the RNA assays when

archived cDNA was used. Markers appended with ".qc" refer to the detection status code referring to that sample's status as ND, <LoQ, or quantifiable detect. The following continuous variables were included in the initial screening of algorithms: original recovery value, storage time (days), CP56_nc, HF183_nc, HF183.esv, AdV.esv, EnV.esv, NoV.GIA.esv, NoV.GIB.esv, NoV.GII.esv, PMMoV.esv, AdV_nc, EnV_nc, NoV.GIA_nc, NoV.GIB_nc, NoV.GII_nc, PMMoV_nc, PMMoV.nc.Hamza (SYBR-based concentration) and average temperature on the sampling day (°F). The following discrete values were included: HF183.qc, AdV.qc, EnV.qc, NoV.GIA.qc, NoV.GIB.qc, NoV.GII.qc, facility, and concentration method. QC values for CP56, HF183, and PMMoV were not included as they were nearly constant.

Optimization for the cubist algorithm included identifying the least important variables (using the *varImp* function) and omitting them to determine whether greater accuracy could be obtained with less computation time. Then, a wider range of hyperparameters was tested. With all the variables, the RMSE for cubist was 0.159. When NoV.GII.qc, AdV.qc, NoV.GIA.esv, NoV.GIB.esv, NoV.GII.esv, PMMoV.esv, and HF183.qc were omitted, it decreased slightly, to 0.158, suggesting that those parameters do not add value to the method of estimating recovery, or that they are unlikely to be quantitatively related to recovery. A range of hyperparameters was tested (1-100 committees and 0-9 neighbors), but the default hyperparameters, with 20 committees and 0 neighbors, resulted in the lowest RMSE. The R^2 was 0.662 and the MAE was 0.0959. The ranked importance of the included variables is as follows: pre-archived recovery > storage time > PMMoV_nc > EnV_nc > PMMoV.nc.Hamza > EnV.esv > concentration method > AdV.esv > NoV.GII_nc > NoV.GIA_nc > AdV_nc > NoV.GIB_nc > EnV.qc > NoV.GIB.qc > NoV.GIA.qc > temperature > facility > HF183.esv > CP56_nc > HF183_nc. This is consistent with expectations, as pre-archived recovery, storage time, and PMMoV concentration changes

could be expected to affect overall recovery and degradation. The code for the SML model is deposited on GitHub: https://github.com/kcrank1/Southern-Nevada-wastewater.

Text S5. Examining the impact of recovery estimation method.

When available, direct quantification of BCoV (or re-quantification of BCoV in archived samples) was used to determine sample-specific recoveries. For archived samples that were non-detect or <LoQ for BCoV (19% of the total dataset), 24 different recovery estimation methods were evaluated, and three were chosen to correct concentrations for all samples to assess the impact of the estimation method on final pathogen distributions. These three consisted of the top two ranked methods as measured by root mean square error (RMSE) and also the most intuitive method—substituting original recovery—although the latter was not ranked highly by RMSE. The top two methods were (1) supervised machine learning (SML) and (2) facility-specific averages of original recovery multiplied by PMMoV degradation after correcting SYBR-based concentrations with linear regression (R2×PM1). There were no significant differences between the distributions that were recovery corrected with SML or R2×PM1 (Table S8). R2×PM1 and SML were also compared to original sample-specific recovery values, and all comparisons and qPCR markers yielded p-values less than 0.05 (data not shown). This indicates that simply substituting the original recovery value causes a significant underestimate of marker concentrations in the original sample due to degradation during storage, and thus warranting an additional adjustment. For example, assume the original measured concentration of a hypothetical sample is 4,000 gc/L. With a 20% recovery adjustment, the actual concentration in the original sample would be 20,000 gc/L. However, if that sample experiences 10% nucleic acid degradation during storage, the measured concentration upon reanalysis would be 3,600 gc/L, and 18,000 gc/L after recovery adjustment. Thus, the estimated concentration of the original sample upon reanalysis would be lower than the "real" concentration unless adjusted with a degradation correction factor.

Table S8. Kolmogrov-Smirnov p-values for the comparison of recovery-adjusted marker concentrations (detected data only) using SML and R2×PM1. Original recovery (R3)-corrected concentrations included for comparison. All p-values for R3 comparisons were < 0.05 (data not shown).

	w/ R2×PM1 est. recovery		w/ SML est. recovery		w/ R3 est. recovery		D statistic	P-value
Marker	Mean (log_{10} gc/L)	Median (log_{10} gc/L)	Mean (log_{10} gc/L)	Median (log_{10} gc/L)	Mean (log_{10} gc/L)	Median (log_{10} gc/L)	(R2×PM1 vs SML)	(R2×PM1 vs SML)
AdV	6.53	6.64	6.54	6.64	6.37	6.50	0.01290323	0.9999989
NoV GIA	6.71	6.73	6.71	6.71	6.60	6.66	0.01520913	0.9999888
NoV GIB	6.65	6.65	6.65	6.62	6.52	6.51	0.01515152	0.9999710
NoV GII	6.41	6.46	6.42	6.46	6.24	6.26	0.02191235	0.9694376
EnV	5.92	5.92	5.93	5.92	5.80	5.83	0.02399128	0.9544882
CP56	9.19	9.21	9.21	9.22	9.01	9.05	0.01618705	0.9986208
HF183	8.04	8.38	8.06	8.40	7.85	8.18	0.01271571	0.9999920

Table S9. Summary statistics of all targets (non-recovery-corrected).

Parameter	*Crypto*	*Giardia*	AdV Cult.	EnV Cult.	AdV Mol.	EnV Mol.	NoV GIA Mol.	NoV GIB Mol.	NoV GI Sum	NoV GII Mol.	CP56 Mol.	HF183 Mol.	PMMoV (Probe)	PMMoV (SYBR)
Number of samples (#)	73	73	56	56	1107	1112	1112	1112	1112	1112	1112	1112	807	1108
Detection Frequency	81%	100%	96%	96%	84%	82%	71%	77%	80%	90%	100%	99%	100%	100%
Mean[a] (log_{10} target/L)	1.62	3.41	2.74	3.22	5.66	5.10	5.91	5.84	6.08	5.54	8.29	7.15	8.40	8.37
St. Dev.[a] (log_{10} target/L)	0.51	0.46	0.68	0.80	1.10	0.63	0.78	0.87	0.91	0.86	0.53	1.34	0.63	0.48
Min[b] (log_{10} target/L)	0.70	1.89	1.53	1.04	2.80	3.91	3.47	4.52	3.55	3.25	5.42	3.65	5.10	5.30
Max (log_{10} target/L)	2.92	4.53	4.35	4.93	8.04	7.25	7.74	8.05	8.21	7.43	9.87	9.49	10.29	9.52
Fitted distribution [c]	μ=1.43	μ=3.41	μ=2.69	μ=3.15	μ=5.35	μ=4.55	μ=5.35	μ=5.08	μ=5.55	μ=5.25	μ=8.29	μ=7.1	μ=8.39	μ=8.33
	σ=0.61	σ=0.45	σ=0.72	σ=0.86	σ=1.26	σ=1.2	σ=1.14	σ=1.55	σ=1.31	σ=1.19	σ=0.53	σ=1.4	σ=0.67	σ=0.51

[a] Mean and standard deviation are of samples with detected target only.
[b] Minimum is the lowest measured concentration above the LoQ.
[c] Distribution fit to censored data using 'fitdistcens' with MLE or non-censored data with 'fitdist'. Non-detect values were considered left-censored, and <LoQ values were considered interval-censored between the LoD and LoQ. Distributions are normal distributions of log_{10}-transformed data, with mean and standard deviation reported in log_{10} target/L.

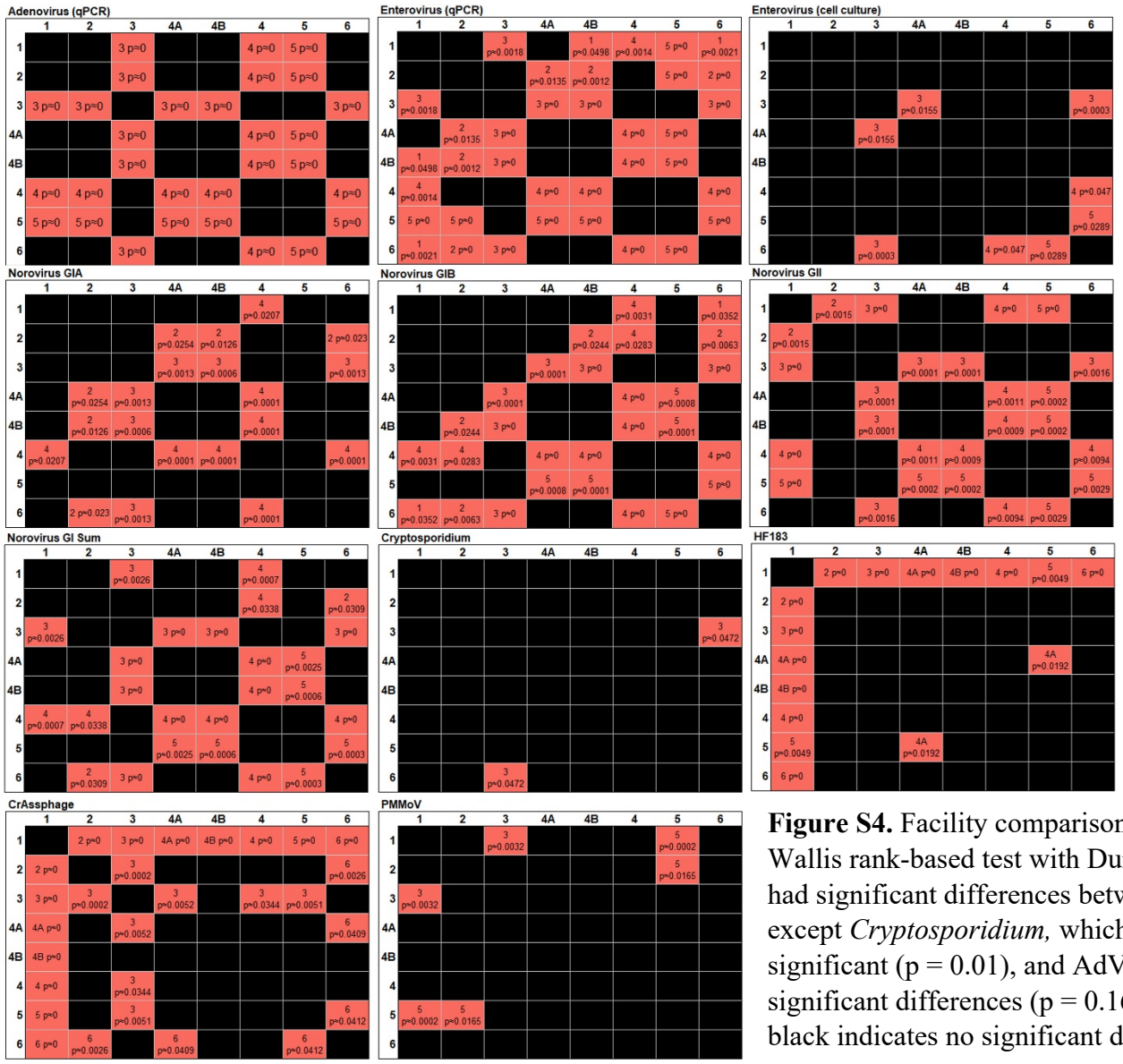

Figure S4. Facility comparison p-values from the Kruskal-Wallis rank-based test with Dunn post-hoc. All markers had significant differences between facilities ($p << 0.05$), except *Cryptosporidium,* which was only slightly significant ($p = 0.01$), and AdV (culture), which had no significant differences ($p = 0.16$). Red indicates $p < 0.05$, black indicates no significant difference.

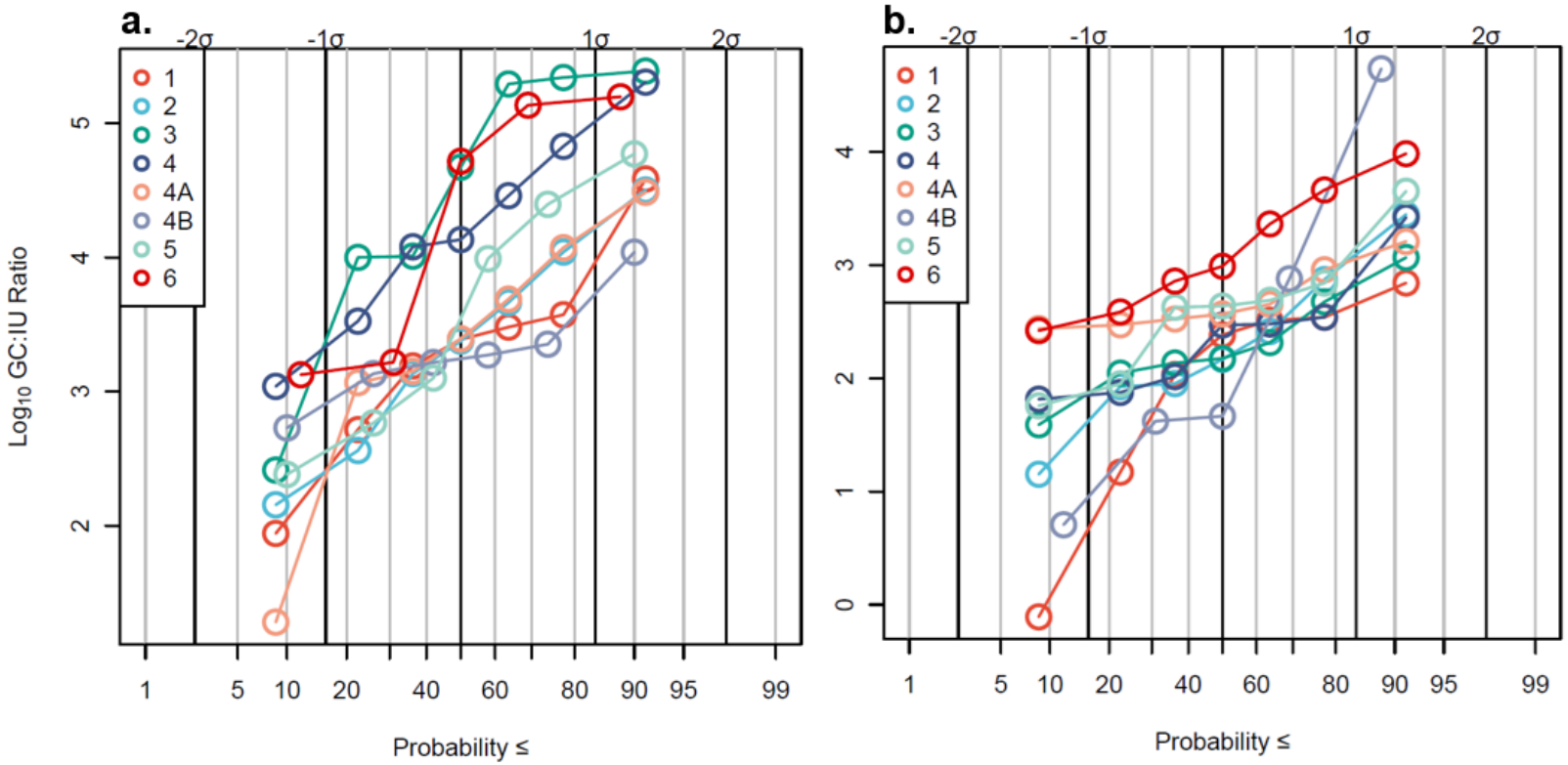

Figure S5. Probability plots for (a) adenovirus and (b) enterovirus of the ratio of the recovery-corrected concentrations measured using qPCR to concentrations measured using cell culture (i.e., $\log_{10}$-transformed GC:IU ratio). Only detected values are plotted.

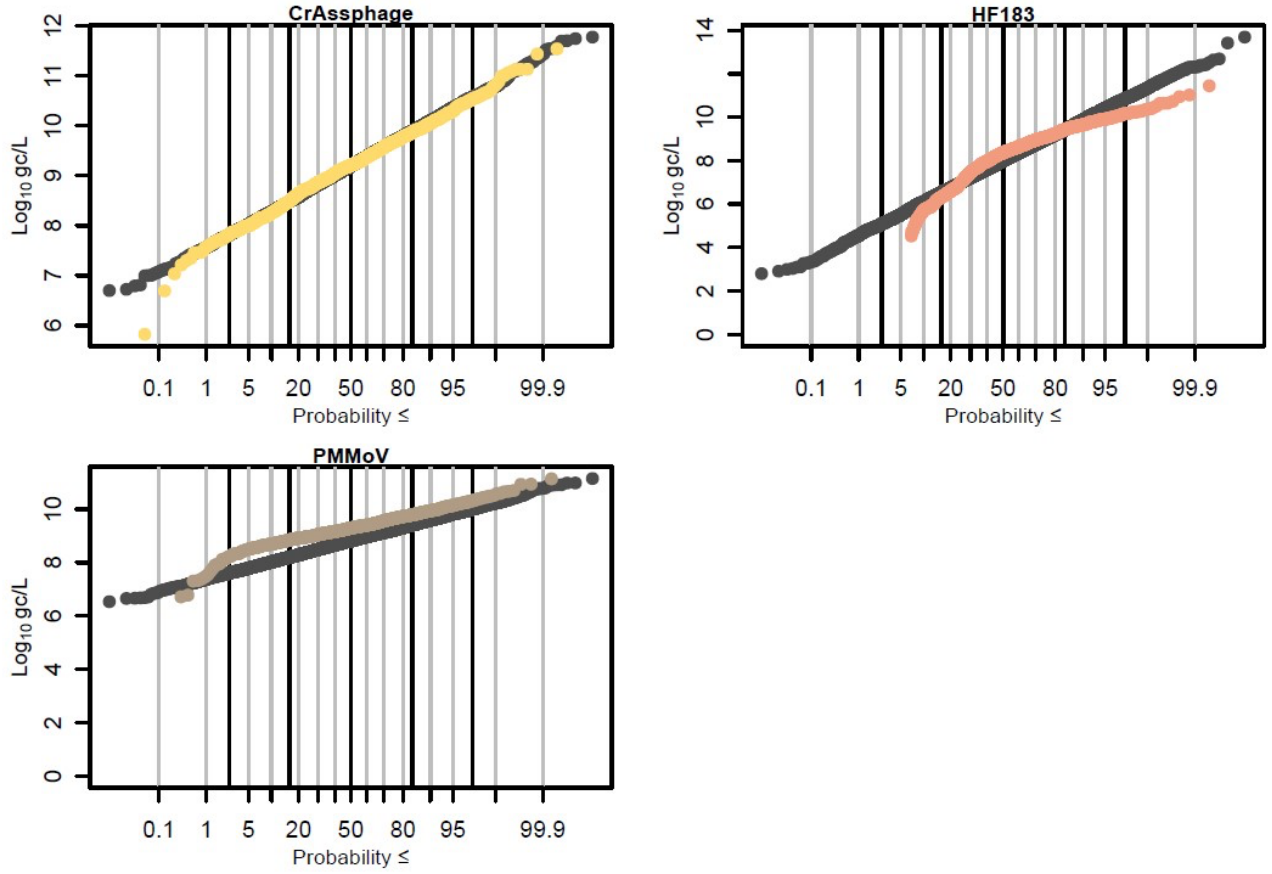

Figure S6. Probability plots for fitted normal distributions (red) and observed data (blue) for log_{10}-transformed human fecal indicator concentrations. <LoD and ND data not plotted but included in the calculation of the percentiles (exceedance probabilities). Solid black vertical lines represent standard deviations from the 50th percentile.

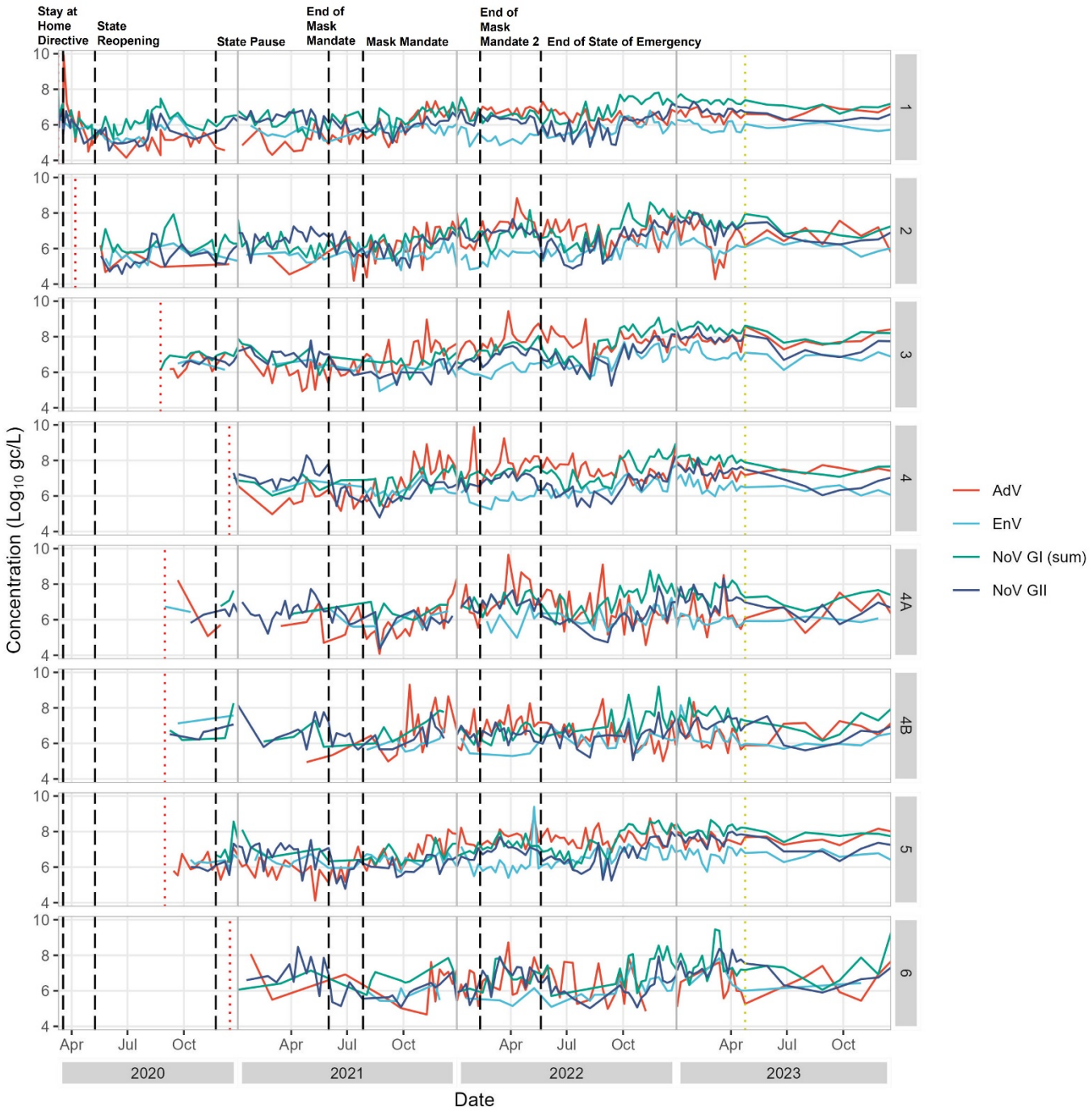

Figure S7. Quantifiable concentrations of viral molecular markers over time. Dashed lines represent the start of different phases of the COVID-19 pandemic in Nevada. Red dotted lines represent sampling start dates for each facility. Weekly sampling switched to monthly sampling on April 24, 2023, represented by yellow dotted line.

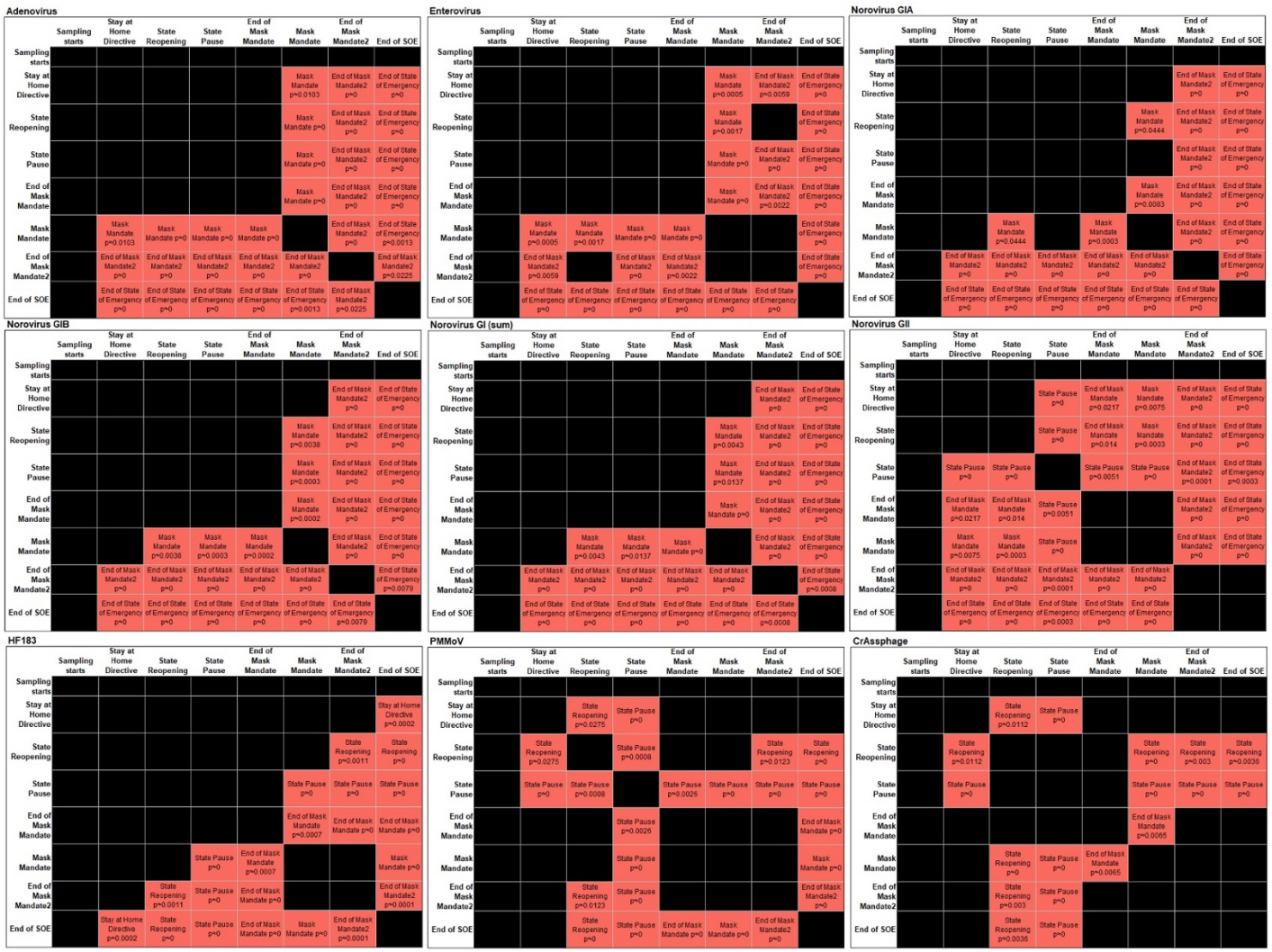

Figure S8. Pandemic phase comparison p-values from the Kruskal-Wallis rank-based test with Dunn post-hoc. The phases with the higher concentrations are listed in the cells with p-values. Only significant ($p < 0.05$) values are displayed.

Table S10. Distribution fittings for samples collected under 'normal' vs. COVID-19 'pandemic' conditions.

Marker	Condition	Mean (log_{10} gc/L)	SD (log_{10} gc/L)	ΔMean (log_{10} gc/L)
AdV	Normal	6.81	0.90	1.09
AdV	Pandemic	5.72	1.28	
EnT	Normal	6.13	0.71	1.18
EnT	Pandemic	4.94	1.15	
HF183	Normal	7.14	1.58	-1.43
HF183	Pandemic	8.58	1.08	
NoV.GI.Sum	Normal	7.26	0.93	1.31
NoV.GI.Sum	Pandemic	5.95	0.96	
NoV.GIA	Normal	6.94	0.89	1.12
NoV.GIA	Pandemic	5.82	0.81	
NoV.GIB	Normal	6.89	1.06	1.55
NoV.GIB	Pandemic	5.35	1.35	
NoV.GII	Normal	6.57	0.98	0.72
NoV.GII	Pandemic	5.85	1.12	
PMMoV	Normal	9.13	0.54	0.40
PMMoV	Pandemic	8.73	1.00	

Distributions fit to a normal distribution of log_{10}-transformed, recovery-corrected, interval- and left-censored concentrations in gc/L. Normal conditions were defined as 05/20/2022 to 03/18/2023, and pandemic conditions were defined as 03/18/2020 to 5/20/2022. crAssphage not included as there was no significant difference between pandemic and non-pandemic concentrations ($p = 0.709$; Mann-Whitney's U test).

Text S6. Data availability.

All concentration data are included in an Excel file that can be accessed at the following location: https://github.com/kcrank1/Southern-Nevada-wastewater.

References

1. Ko, G.; Jothikumar, N.; Hill, V. R.; Sobsey, M. D., Rapid detection of infectious adenoviruses by mRNA real-time RT-PCR. *Journal of virological methods* **2005,** *127*, (2), 148-153.
2. Pecson, B. M.; Darby, E.; Danielson, R.; Dearborn, Y.; Di Giovanni, G.; Jakubowski, W.; Leddy, M.; Lukasik, G.; Mull, B.; Nelson, K. L., Distributions of waterborne pathogens in raw wastewater based on a 14-month, multi-site monitoring campaign. *Water Research* **2022,** *213*, 118170.
3. Decaro, N.; Elia, G.; Campolo, M.; Desario, C.; Mari, V.; Radogna, A.; Colaianni, M. L.; Cirone, F.; Tempesta, M.; Buonavoglia, C., Detection of bovine coronavirus using a TaqMan-based real-time RT-PCR assay. *Journal of virological methods* **2008,** *151*, (2), 167-171.
4. Stachler, E.; Kelty, C.; Sivaganesan, M.; Li, X.; Bibby, K.; Shanks, O. C., Quantitative CrAssphage PCR assays for human fecal pollution measurement. *Environmental science & technology* **2017,** *51*, (16), 9146-9154.
5. USEPA;, Method 1615: Measurement of enterovirus and norovirus occurrence in water by culture and RT-qPCR. **2012**, 1–91.
6. Haugland, R. A.; Varma, M.; Sivaganesan, M.; Kelty, C.; Peed, L.; Shanks, O. C., Evaluation of genetic markers from the 16S rRNA gene V2 region for use in quantitative detection of selected Bacteroidales species and human fecal waste by qPCR. *Systematic and applied microbiology* **2010,** *33*, (6), 348-357.
7. Hill, V. R.; Mull, B.; Jothikumar, N.; Ferdinand, K.; Vinjé, J., Detection of GI and GII noroviruses in ground water using ultrafiltration and TaqMan real-time RT-PCR. *Food and Environmental Virology* **2010,** *2*, 218-224.
8. Jothikumar, N.; Lowther, J. A.; Henshilwood, K.; Lees, D. N.; Hill, V. R.; Vinjé, J., Rapid and sensitive detection of noroviruses by using TaqMan-based one-step reverse transcription-PCR assays and application to naturally contaminated shellfish samples. *Applied and environmental microbiology* **2005,** *71*, (4), 1870-1875.
9. Svraka, S.; Duizer, E.; Vennema, H.; de Bruin, E.; van der Veer, B.; Dorresteijn, B.; Koopmans, M., Etiological role of viruses in outbreaks of acute gastroenteritis in The Netherlands from 1994 through 2005. *Journal of clinical microbiology* **2007,** *45*, (5), 1389-1394.
10. Haramoto, E.; Kitajima, M.; Kishida, N.; Konno, Y.; Katayama, H.; Asami, M.; Akiba, M., Occurrence of pepper mild mottle virus in drinking water sources in Japan. *Applied and environmental microbiology* **2013,** *79*, (23), 7413-7418.
11. Sivaganesan, M.; Haugland, R. A.; Chern, E. C.; Shanks, O. C., Improved strategies and optimization of calibration models for real-time PCR absolute quantification. *Water Research* **2010,** *44*, (16), 4726-4735.

12. Gerrity, D.; Papp, K.; Stoker, M.; Sims, A.; Frehner, W., Early-pandemic wastewater surveillance of SARS-CoV-2 in Southern Nevada: Methodology, occurrence, and incidence/prevalence considerations. *Water research X* **2021,** *10*, 100086.
13. Francy, D. S.; Simmons III, O. D.; Ware, M. W.; Granger, E. J.; Sobsey, M. D.; Schaefer III, F. W., Effects of seeding procedures and water quality on recovery of Cryptosporidium oocysts from stream water by using US Environmental Protection Agency Method 1623. *Applied and Environmental Microbiology* **2004,** *70*, (7), 4118-4128.