1   **Supplementary information for**

2

3   **Prewhitening and Normalization Help Detect a Strong Cross-Correlation Between**

4   **Daily Wastewater SARS-CoV-2 RNA Abundance and COVID-19 Cases in a**

5   **Community**

6

7   Min Ki Jeon [1], Bo Li [1], Doris Yoong Wen Di [1], and Tao Yan [1]*

8

9   [1] Department of Civil and Environmental Engineering, University of Hawaii at Manoa,

10   Honolulu, HI 96822

11
12

13 **Table S1.** Primers and probes used for qPCR assays for detection of SARS-CoV-2 genes and fecal viral indicators from wastewater

Taget gene	Primer/probe	Sequence (5' → 3')	Amplicon size (bp)	Concentration (nM)	Reference
SARS-CoV-2 N1	2019-nCoV_N1-F	GAC CCC AAA ATC AGC GAA AT	72	150	(1)
	2019-nCoV_N1-R	TCT GGT TAC TGC CAG TTG AAT CTG		150	
	2019-nCoV_N1-P	FAM-ACC CCG CAT TAC GTT TGG TGG ACC-BHQ1		50	
SARS-CoV-2 N2	2019-nCoV_N1-F	TTA CAA ACA TTG GCC GCA AA	67	150	
	2019-nCoV_N1-R	GCG CGA CAT TCC GAA GAA		150	
	2019-nCoV_N1-P	FAM-ACA ATT TGC CCC CAG CGC TTC AG-BHQ1		50	
SARS-CoV-2 E	E_Sarbeco_F	ACA GGT ACG TTA ATA GTT AAT AGC GT	113	150	(2)
	E_Sarbeco_R	ATA TTG CAG CAG TAC GCA CAC A		150	
	E_Sarbeco_P	FAM-ACA CTA GCC ATC CTT ACT GCG CTT CG-BHQ1		50	
F+ RNA coliphages Group II (G2)	G2_F	TTA CTG TCG TTC CTG TTA GCA ATG	104	400	(3)
	G2_R	CRC CTG ACG CAC GAT AAC T		400	
	G2_P	FAM-ACG GCG TCG CTG AGT GGC TTT C-BHQ1		200	
F+ RNA coliphages Group III (G3)	G3_F	TAA ATC CCA CYA ACG GYG TTG C	108	400	
	G3_R	TTI CGA TTR CGI GAA GGC TG		400	
	G3_P1	FAM-TGG AGA AGC GTG TTA CCG TTT-BHQ1		200	
	G3_P2	FAM-TGG AGA AGC GTG TYA CAA TTT CTG TRT C-BHQ1		200	

Pepper mild mottle virus (PMMoV)	PMMoV-F	GAG TGG TTT GAC CTT AAC GTT GA	68	200	(4)
	PMMoV-R	TTG TCG GTT GCA ATG CAA GT		200	
	PMMoV-P	FAM-CCT ACC GAA GCA AAT G-IABkFQ		100	
Bovine coronavirus (BCoV)[a]	BCoV-F	CTG GAA GTT GGT GGA GTT	85	150	(5)
	BCoV-R	ATT ATC GGC CTA ACA TAC ATC		150	
	BCoV-P	FAM-CCT TCA TAT CTA TAC ACA TCA AGT TGT T-BHQ1		50	

14 [a] BCoV was directly spiked into the solid subsamples after collection and before viral RNA extraction; for the liquid subsamples,

15 BCoV was spiked into the supernatant after centrifugation and before PEG precipitation. The quantity initially seeded into each batch

16 of solid and liquid wastewater samples was ~800 genome copy/spike (Batch 1; 8/31/2020 to 9/2/2020, $n = 6$), and later increased to

17 higher levels to facilitate data analysis: 1.3×10^7 genome copy/spike (Batch 2; 9/8/2020 to 9/14/2020, $n = 12$), and 3.7×10^7 genome

18 copy/spike (Batch 3; 9/21/2020 to 9/28/2020, $n = 14$), and 2.5×10^7 genome copy/spike (Batch 4; 9/29/2020 to 10/4/2020, $n = 14$).

19 No significant difference in recovery was observed between the different spiking concentrations. Assessment of recovery based on

20 the spiked BCoV showed that the averaged recovery ratios were 1.7% ($\sigma = 1.7\%$) for the liquid subsamples ($n = 46$) and 0.25% ($\sigma =$

21 0.36%) for solid subsamples ($n = 46$).

22

23 **Table S2.** Thermal cycling conditions of qPCR assays and calibration curve information

Target gene	Temperature (°C)	Time	Cycle	Calibration equation	qPCR efficiency	R^2	Limit of quantification (GC/rxn)	Limit of detection (GC/rxn)
SARS-CoV-2 N1	95	2 min	1	y = -3.484x + 40.429	93.8%	0.991	10	2.5
SARS-CoV-2 N2	95	3 sec	45	y = -3.420x + 38.228	96.1%	0.999	10	2.5
	55	30 sec						
SARS-CoV-2 E	95	3 min	1	y = -3.403x + 37.882	96.8%	0.999	10	2.5
	95	15 sec	45					
	58	30 sec						
G2	95	2 min	1	y = -3.552x + 38.841	91.4%	0.994	10	2.5
G3	95	15 sec	45	y = -3.571x + 39.277	90.6%	0.998	10	2.5
	56	30 sec						
PMMoV	95	2 min	1	y = -3.238x + 38.075	106.6%	0.999	10	2.5
	95	15 sec	45					
	53	30 sec						
	72	30 sec						
BCoV	95	2 min	1	y = -3.286x + 38.993	102.0%	0.998	10	5
	95	15 sec	45					
	56	28 sec						

25 **Table S3.** Slope, trend, and normality test results of before and after prewhitening of SARS-CoV-2 abundance (both raw and

26 normalized) from wastewater and daily clinical case numbers

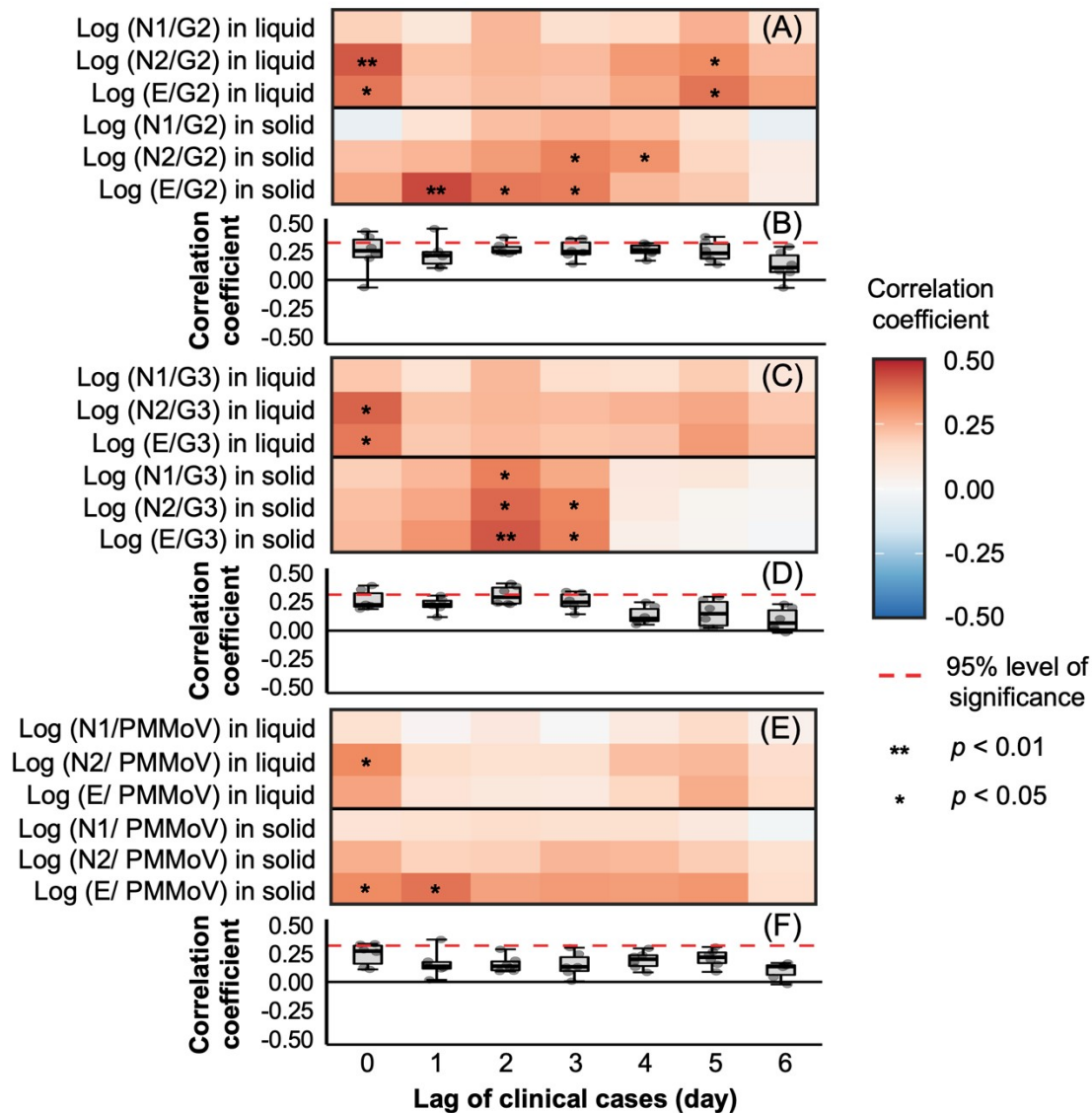| Sample	Analysis	Prewhitening	Average	Standard Deviation	Maximum value	Minimum value
Honouliuli WWTP (n = 24)	Slope	Before	-0.056	0.024	-0.021	-0.1
		After	0.002	0.005	0.013	-0.01
	Mann-Kendall trend test p-value[a]	Before	0.013	0.034	0.137	0
		After	0.809	0.159	1	0.435
	Shapiro-Wilk test p-value[b]	Before	0.101	0.104	0.445	0.002
		After	0.443	0.252	0.963	0.093
Sand Island WWTP (n = 24)	Slope	Before	-0.055	0.02	-0.014	-0.1
		After	0.003	0.005	0.011	-0.007
	Mann-Kendall trend test p-value	Before	0	0.001	0.007	0
		After	0.799	0.169	1	0.443
	Shapiro-Wilk test p-value	Before	0.069	0.1	0.449	0.002
		After	0.344	0.236	0.831	0.055
Daily clinical case numbers (n = 1)	Slope	Before	-	-	-3.5081	
		After	-	-	0.2655	
	Mann-Kendall trend test p-value	Before	-	-	0	
		After	-	-	0.308	
	Shapiro-Wilk test p-value	Before	-	-	0.002	
		After	-	-	0.436	

27 [a] The null hypothesis of Mann-Kendall trend test is that the data has no trend or serial correlation structure throughout the observed

28 time points.

29 [b] The null hypothesis of Shapiro-Wilk test is that the data is normally distributed.
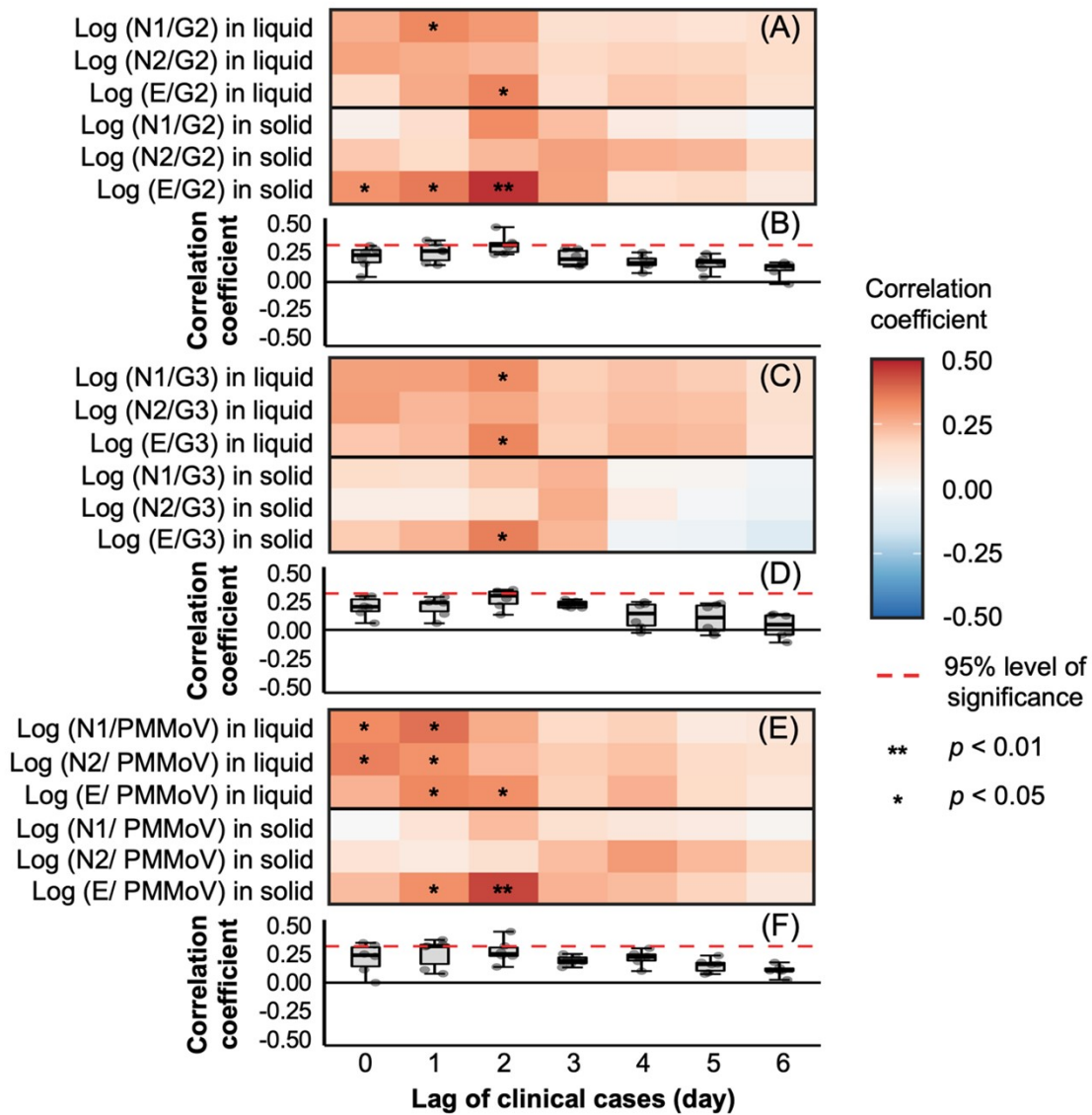
Figure S1. Cross-correlation between the non-prewhitened COVID-19 new case numbers and the non-prewhitened SARS-CoV-2 RNA normalized abundance in wastewater samples from the SI WWTP. The normalized abundance was calculated by dividing SARS-CoV-2 RNA abundance by F+ RNA coliphage Group II (A, B), Group III (C, D), and PMMoV (E, F). All normalized abundances were transformed into log forms. Red dashed lines represent a 95% level of significance and the *p*-value of the correlation less than 0.05 are displayed as asterisks. The middle, upper, and lower lines in the box of the boxplot represent the median, 25th, and 75th percentiles, respectively, and the whiskers represent the largest and smallest values outside of the interquartile range.

40

Figure S2. Cross-correlation between the non-prewhitened COVID-19 new case numbers and the non-prewhitened SARS-CoV-2 RNA normalized abundance in wastewater samples from the HO WWTP. The normalized abundance was calculated by dividing SARS-CoV-2 RNA abundance by F+ RNA coliphage Group II (A, B), Group III (C, D), and PMMoV (E, F). All normalized abundances were transformed into log forms. Red dashed lines represent a 95% level of significance and the p-value of the correlation less than 0.05 are displayed as asterisks. The middle, upper, and lower lines in the box of the boxplot represent the median, 25th, and 75th percentiles, respectively, and the whiskers represent the largest and smallest values outside of the interquartile range.

50

References

1.   Lu X, Wang L, Sakthivel SK, Whitaker B, Murray J, Kamili S, et al. US CDC real-time reverse transcription PCR panel for detection of severe acute respiratory syndrome coronavirus 2. Emerging infectious diseases. 2020;26(8):1654.

2.   Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DK, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. Eurosurveillance. 2020;25(3):2000045.

3.   Friedman SD, Cooper EM, Calci KR, Genthner FJ. Design and assessment of a real time reverse transcription-PCR method to genotype single-stranded RNA male-specific coliphages (Family Leviviridae). Journal of virological methods. 2011;173(2):196-202.

4.   Haramoto E, Kitajima M, Kishida N, Konno Y, Katayama H, Asami M, et al. Occurrence of pepper mild mottle virus in drinking water sources in Japan. Applied and environmental microbiology. 2013;79(23):7413-8.

5.   Decaro N, Elia G, Campolo M, Desario C, Mari V, Radogna A, et al. Detection of bovine coronavirus using a TaqMan-based real-time RT-PCR assay. Journal of virological methods. 2008;151(2):167-71.