Title: SARS-CoV-2 titers in wastewater are higher than expected from clinically confirmed cases

Authors: Wu FQ(1); Xiao A(1); Zhang JB(1); Gu XQ(2); Lee WL(2); Kauffman K (3); Hanage WP(4); Matus M (5); Ghaeli N(5); Endo N(5); Duvallet C(5); Moniz K(1); Erickson TB(6); Chai PR (6); Thompson J(7); Alm EJ(1,2,5)

1: Center for Microbiome Informatics and Therapeutics, Departments of Biological Engineering and Civil & Environmental Engineering, Massachusetts Institute of Technology
2: Singapore-MIT Alliance for Research and Technology, National University of Singapore
3: University at Buffalo, The State University of New York
4: Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard T. H. Chan School of Public Health, Boston
5: Biobot Analytics, Cambridge MA
6: Division of Medical Toxicology, Department of Emergency Medicine, Brigham and Women's Hospital
7: Singapore Center for Environmental Life Sciences Engineering, Asian School of the Environment, Nanyang Technological University, Singapore

Abstract. Wastewater surveillance may represent a complementary approach to measure the presence and even prevalence of infectious diseases when the capacity for clinical testing is limited. Moreover, aggregate, population-wide data can help inform modeling efforts. We tested wastewater collected at a major urban treatment facility in Massachusetts and found the presence of SARS-CoV-2 at high titers in the period from March 18 - 25 using RT-qPCR. We then confirmed the identity of the PCR product by direct DNA sequencing. Viral titers observed were significantly higher than expected based on clinically confirmed cases in Massachusetts as of March 25. The reason for the discrepancy is not yet clear, and until further experiments are complete, these data do not necessarily indicate that clinical estimates are incorrect. Our approach is scalable and may be useful in modeling the SARS-CoV-2 pandemic and future outbreaks.

Introduction. Improved understanding of the presence and prevalence of SARS-CoV-2 at a population level can help government and hospital officials implement appropriate policies to mitigate the exponential spread of COVID-19, and diminish the future strain on healthcare facilities. Despite pandemic spread of SARS-CoV-2 worldwide, broad access to testing in the United States (US) has thus far been severely limited. While it is impractical to test every US resident for SARS-CoV-2, the virus has been found in the stool of confirmed COVID-19 patients (*1*), making it a promising candidate for wastewater-based epidemiology (WBE). WBE can help detect the presence of pathogens across municipalities and estimate population prevalence without individual testing, and inform public health officials of the efficacy of interventions. The closely related virus SARS-CoV was detected in wastewater from Chinese hospitals during the 2002-2003 SARS pandemic (*2*), and WBE has been used for the early detection and direct mitigation of disease outbreaks in Israel, Egypt and Sweden (*3–7*). We have previously used this technique to measure and map the use of pharmaceuticals across residential communities (*8*). Here, we describe an analytical technique to extract and detect genetic material from SARS-CoV-2 using wastewater collected at a treatment facility.

Results. We collected sewage samples at a major urban wastewater treatment facility in Massachusetts. Samples were transported to our laboratory where we conducted viral inactivation and enrichment, nucleic acid extraction, and RT-qPCR. As negative controls, we used biobanked wastewater samples from the same treatment facility taken before the first US case was documented.

Initial testing with PCR using primers specific for the SARS-CoV-2 *S* gene (*9*) indicated that both samples collected from the treatment facility on March 18 had a positive signal for SARS-CoV-2. We confirmed the signal by identifying a PCR product at ~150 bp. A no-template control for the PCR assays was negative. Sanger sequencing of the PCR products confirmed a 97-98% identity match to the SARS-CoV-2 *S* gene (Figure 1).

We next sought to quantify viral titer in sewage, and to establish our viral enrichment protocol using RT-qPCR. We used the US CDC primer/probe sets targeting the N1, N2, and N3 loci of the SARS-CoV-2 nucleocapsid gene to amplify cDNA reverse transcribed from viral RNA (Table 1). We tested different steps in our viral enrichment process. First, we examined raw (unfiltered) sewage precipitated with polyethylene glycol 8000 (PEG), which recovers both bacterial and viral nucleic acids. Next, we looked at samples taken from 0.2 um filtered sewage: we considered both the material collected on the filter and the filtrate. We found the strongest and most consistent results from the PEG-precipitated viral pellet from the 0.2 um filtrate, which was resuspended in Trizol for RNA extraction (Table 2).

These results suggest a fairly simple viral enrichment protocol is sufficient to achieve viral identification. Importantly, we included pasteurization (90 min at 60C) as a first step performed before sample containers were opened, increasing the safety of the protocol. Prior work on SARS-CoV-1 indicates that a 30 min heat inactivation at 60C is sufficient to inactivate the virus by over 6 log units (*10*).

Having shown that it is possible to detect the presence of SARS-CoV-2 in wastewater treatment facility samples, we next looked at whether the viral titer in

sewage could be quantified. Standard curves for all three primer sets showed linear

behavior over 6 log units (r^2 ranging from 0.981 to 0.998) using DNA standards (SARS-

CoV-2 nucleocapsid gene). We sought to validate our results against other samples,

including biobanked samples collected before the first known US case on January 20,

which could serve as a relevant negative control. Figure 2 shows estimated viral titer

per mL of sewage at the treatment facility over seven different sample dates, with

samples from two catchment areas (Southern and Northern) for each date. All four

samples from before the first known US SARS-CoV-2 case were negative for virus. In

contrast, all 10 samples taken in the range of March 18 - March 25 tested positive

(seven out of ten samples hit all the three primers with an average Ct for all samples

below 40) for virus by qPCR, using the DNA probes provided by the CDC (*11*). For each

positive sample, we confirmed the presence of a ~150 bp band after PCR using the S

gene primers (data not shown). Thus, our approach is sensitive enough to detect viral

titers similar to those currently seen in the treatment facility catchments. Spike-in

experiments with purified virus will be needed to establish rigorous limits of detection

and are currently underway.

Finally, we looked at how quickly viral signal degraded during storage at 4C. All

samples were received between March 20-25, but samples were stored and

reprocessed on April 4. All samples were pasteurized upon initial receipt. Figure 3

shows the variation between measured samples processed on both dates. There was

variation between sample runs, but no clear trend toward lower signals at the later

timepoint, suggesting samples can be stored for more than a week without significant

degradation of signal.

Discussion. These data suggest an order of magnitude estimate of approximately 100 viral particles per mL of sewage, but what conclusions can be drawn about disease prevalence? First, we note that any rigorous conclusions depend on a number of factors that are unknown, and thus additional experiments will be required to calibrate these numbers. Nonetheless, we can estimate an abundance based on the lowest observed values across these samples of ~10 copies/mL. If we assume typical stool sizes of 200g, diluted into an average volume of $1.36*10^9$ L, and a population of $2.3*10^6$ individuals each producing one stool per day, and we further assume that there is no loss of viral RNA in sewer lines and that excreted viruses are fully suspended in sewage, then we expect the viral titer in feces to be about 3000 times higher than that in sampled raw sewage, or about 30,000 particles per mL.

Estimates of viral load in stool from positive patients are still a matter of uncertainty, but at least one recent publication suggests levels as high as 600,000 viral genomes per mL of fecal material (*12*). This number would suggest roughly 5% of all fecal samples in the treatment facility catchment were positive for SARS-CoV-2 in the March 18 – 25 period, a number much higher than the 0.026% confirmed for the state of Massachusetts (a similar prevalence is obtained using individual counties represented by the wastewater treatment facility's catchment or state-wide estimates) on March 25. Another paper reported a maximum observed value close to 30,000,000 viral particles per mL in a single fecal sample (*13*). If we use this number instead, we would estimate a prevalence of 0.1%, closer to, but still much higher than the number of confirmed

clinical cases. Additional data on viral shedding in stool over the course of disease is

required to fully interpret these findings.

This discrepancy between confirmed cases and observed viral titers could arise

from a number of factors. First, our estimates are based on numbers and assumptions

that are currently subject to significant uncertainty, as discussed above, and therefore

could be incorrect. However, we note that our calculations are conservative because we

assume no loss in viral titer due to degradation, sample processing, or RNA extraction.

Second, the estimates of viral titer per positive stool may be too low if a small number of

individuals shed very high levels of virus. While this can be tested, it may require

extensive testing of individual stools in order to arrive at an accurate estimate. Given the

importance of assessing the fraction of SARS-CoV-2 infections that present with

symptoms, we note that our results are consistent with the idea that a significant fraction

of cases that are not detected with current testing algorithms, and that this fraction may

include a large number of patients without symptoms. (*14*, *15*).

To discriminate between these possibilities, additional catchments of different

sizes should be tested. By integrating data on the presence or absence of viral particles

across catchments of varying size within the same geographic region, it may be

possible to estimate disease prevalence independent from knowing the average viral

titer in infected stool. For example, if disease prevalence is 1 in 10,000, then

approximately 50% of catchment areas representing ~7,000 individuals would be

positive (although cases may cluster within households, so additional experiments or

modeling would be needed to derive precise numbers). These experiments require

sampling upstream areas in the sewage system, using specialized equipment to capture

time-integrated samples, and are currently underway. Once these experiments are

completed in a small number of regions, they could be used to estimate disease

prevalence in other regions using wastewater treatment facility data alone.

Conclusions. These data demonstrate the feasibility of measuring SARS-CoV-2 in

wastewater. The implications of this research are that WBE can be leveraged to detect

population level prevalence of SARS-CoV-2 in cities across the world. In a setting

where in-person testing may not be available, longitudinal analysis of wastewater can

provide population-level estimates of the burden of SARS-CoV-2.

These data may help inform decisions surrounding the advancement or scale-

back of social distancing and quarantine efforts based on wastewater catchment-level

estimations of prevalence. Additionally, wastewater collection at the municipal or

community level may allow for more granular detection of SARS-CoV-2 in cities with

lower COVID-19 disease burden, thereby functioning as an early warning system to

help preemptively enact public health measures prior to the widespread onset of

disease.

Methods. 24 hour composite samples of raw sewage were taken from the wastewater

treatment facility influent, and pasteurized at 60C for 90 min to inactivate virus. Raw

sewage was then filtered through a 0.2 um membrane (Millipore Sigma) to remove

bacterial cells and debris. Filters were discarded, as initial tests revealed little to no viral

RNA on filters. 4 g of Polyethylene glycol 8000 (8% w/v, Millipore Sigma) and 0.9 g

NaCl (0.3 M, Millipore Sigma) was then added to 40 mL filtrate, and centrifuged at

12,000 X g for 2 hours or until a pellet was visible. The viral pellet was then resuspended in Trizol (Thermofisher) for RNA extraction, followed by reverse transcription (reverse transcriptase, NEB) and qPCR (TaqMan fast advanced master mix, Thermofisher) with CDC primers (IDT) for the nucleocapsid *N* gene. Positive control (IDT) is a plasmid containing complete nucleocapsid gene from SARS-CoV-2, and used to create the standard curves for N1, N2, and N3 primers.

REFERENCES

1. Y. Wu, C. Guo, L. Tang, Z. Hong, J. Zhou, X. Dong, H. Yin, Q. Xiao, Y. Tang, X. Qu, L. Kuang, X. Fang, N. Mishra, J. Lu, H. Shan, G. Jiang, X. Huang, Prolonged presence of SARS-CoV-2 viral RNA in faecal samples. *The Lancet Gastroenterology & Hepatology*. **0** (2020), doi:10.1016/S2468-1253(20)30083-2.

2. X. W. Wang, J. Li, T. Guo, B. Zhen, Q. Kong, B. Yi, Z. Li, N. Song, M. Jin, W. Xiao, X. Zhu, C. Gu, J. Yin, W. Wei, W. Yao, C. Liu, J. Li, G. Ou, M. Wang, T. Fang, G. Wang, Y. Qiu, H. Wu, F. Chao, J. Li, Concentration and detection of SARS coronavirus in sewage from Xiao Tang Shan Hospital and the 309th Hospital of the Chinese People's Liberation Army. *Water Sci. Technol.* **52**, 213–221 (2005).

3. S. Blomqvist, L. E. Bassioni, E. M. E. M. Nasr, A. Paananen, S. Kaijalainen, H. Asghar, E. de Gourville, M. Roivainen, Detection of Imported Wild Polioviruses and of Vaccine-Derived Polioviruses by Environmental Surveillance in Egypt. *Applied and Environmental Microbiology*. **78**, 5406 (2012).

4. M. Hellmér, N. Paxéus, L. Magnius, L. Enache, B. Arnholm, A. Johansson, T. Bergström, H. Norder, Detection of pathogenic viruses in sewage provided early warnings of hepatitis A virus and norovirus outbreaks. *Appl. Environ. Microbiol.* **80**, 6771–6781 (2014).

5. E. Kopel, E. Kaliner, I. Grotto, Lessons from a Public Health Emergency — Importation of Wild Poliovirus to Israel. *New England Journal of Medicine*. **371**, 981–983 (2014).

6. Y. Berchenko, Y. Manor, L. S. Freedman, E. Kaliner, I. Grotto, E. Mendelson, A. Huppert, Estimation of polio infection prevalence from environmental surveillance data. *Sci Transl Med*. **9** (2017), doi:10.1126/scitranslmed.aaf6786.

7. R. J. Duintjer Tebbens, M. Zimmermann, M. A. Pallansch, K. M. Thompson, Insights from a Systematic Search for Information on Designs, Costs, and Effectiveness of Poliovirus Environmental Surveillance Systems. *Food Environ Virol*. **9**, 361–382 (2017).

8. N. Endo, N. Ghaeli, C. Duvallet, K. Foppe, T. B. Erickson, M. Matus, P. R. Chai, Rapid Assessment of Opioid Exposure and Treatment in Cities Through Robotic Collection and Chemical Analysis of Wastewater. *J Med Toxicol*. **16**, 195–203 (2020).

9. F. Zhang, O. Abudayyeh, J. Gootenberg, A protocol for detection of COVID-19 using CRISPR diagnostics. *v.20200321*.

10. H. F. Rabenau, J. Cinatl, B. Morgenstern, G. Bauer, W. Preiser, H. W. Doerr, Stability and inactivation of SARS coronavirus. *Med. Microbiol. Immunol.* **194**, 1–6 (2005).

11. CDC, Coronavirus Disease 2019 (COVID-19). *Centers for Disease Control and Prevention* (2020), (available at https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html).

12. N. Zhang, Y. Gong, F. Meng, Y. Bi, P. Yang, F. Wang, Virus shedding patterns in nasopharyngeal and fecal specimens of COVID-19 patients. *medRxiv*, in press, doi:10.1101/2020.03.28.20043059.

13. R. Wölfel, V. M. Corman, W. Guggemos, M. Seilmaier, S. Zange, M. A. Müller, D. Niemeyer, T. C. Jones, P. Vollmar, C. Rothe, M. Hoelscher, T. Bleicker, S. Brünink, J. Schneider, R.

Ehmann, K. Zwirglmaier, C. Drosten, C. Wendtner, Virological assessment of hospitalized patients with COVID-2019. *Nature*, 1–10 (2020).

14. R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, J. Shaman, Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2). *Science* (2020), doi:10.1126/science.abb3221.

15. Y. Bai, L. Yao, T. Wei, F. Tian, D.-Y. Jin, L. Chen, M. Wang, Presumed Asymptomatic Carrier Transmission of COVID-19. *JAMA* (2020), doi:10.1001/jama.2020.2565.

Table 1 Primers and probes used in this study.

Primer	Primer/probe sequence (5'-3')	Amplicon size	Target gene	Reference
S-F	GAAATTAATACGACTCACTATAGGGAGGT TTCAAACTTTACTTGCTTTACATAGA	137 bp	Glycoprotein (S)	9
S-R	TCCTAGGTTGAAGATAACCCACATAATAAG			
N1-F	GACCCCAAAATCAGCGAAAT			
N1-R	TCTGGTTACTGCCAGTTGAATCTG	72 bp		
N1-P	FAM-ACCCCGCATTACGTTTGGTGGACC-BHQ1			
N2-F	TTACAAACATTGGCCGCAAA			
N2-R	GCGCGACATTCCGAAGAA	67 bp	Nucleocapsid (N)	11
N2-P	FAM-ACAATTTGCCCCCAGCGCTTCAG-BHQ1			
N3-F	GGGAGCCTTGAATACACCAAAA			
N3-R	TGTAGCACGATTGCAGCATTG	72 bp		
N3-P	FAM-AYCACATTGGCACCCGCAATCCTG-BHQ1			

Note: highlighted primers/probes are used for qPCR. F: forward; R: reverse; P: probe.

Table 2 Mean Ct values for Southern and Northern samples (unfiltered vs filters vs filtrate).

	N1 primer	N2 primer	N3 primer
Southern-Filtrate	33.87	38.39	37.12
Southern-Filter	NA	NA	NA
Southern-Unfiltered	NA	NA	NA
Northern-Filtrate	34.91	NA	37.85
Northern-Filter	37.86	NA	38.14
Northern-Unfiltered	35.04	NA	38.09

```
>Northern-Forward   ----------------------------------------------------
>COVID-19-S gene     GAAATTAATACGACTCACTATAGGGAGGTTTCAAACTTTACTTGCTTTAC    50
>Northern-Reverse    GAAATTAATACGACTCACTATAGGGAGGTTTCAAACTTTACTTGCTTTAC

>Northern-Forward   -----NNNNNNNNNNACTNCTGNNGNNTCTTCTTNNGGTTGGACAGCTGGT
>COVID-19-S gene     ATAGAAGTTATTTGACTCCTGGTGATTCTTCTTCAGGTTGGACAGCTGGT   100
>Northern-Reverse    ATAGAAGTTATTTGACTCCTGGTG-NTCTTCTTCAGGTNGTGNNNNNNNNN

>Northern-Forward   GCTGCAGCTTATTATGTGGGTTATCTTCAACCTAGGA
>COVID-19-S gene     GCTGCAGCTTATTATGTGGGTTATCTTCAACCTAGGA                137
>Northern-Reverse    NNNNNNN------------------------------
```

Figure 1 Sanger Sequencing results and alignment to COVID-19 *S* gene.

Highlighted grey: forward and reverse primers; highlighted yellow: aligned sequence;

highlighted purple: mismatched sequence from either primer. The first 24 bases in the

COVID-19 *S* gene is the T7 promoter added by PCR with S-F and S-R primers in Table

1. Results shown only for sequence of template form Northern influent, results from

Southern influent were similar.
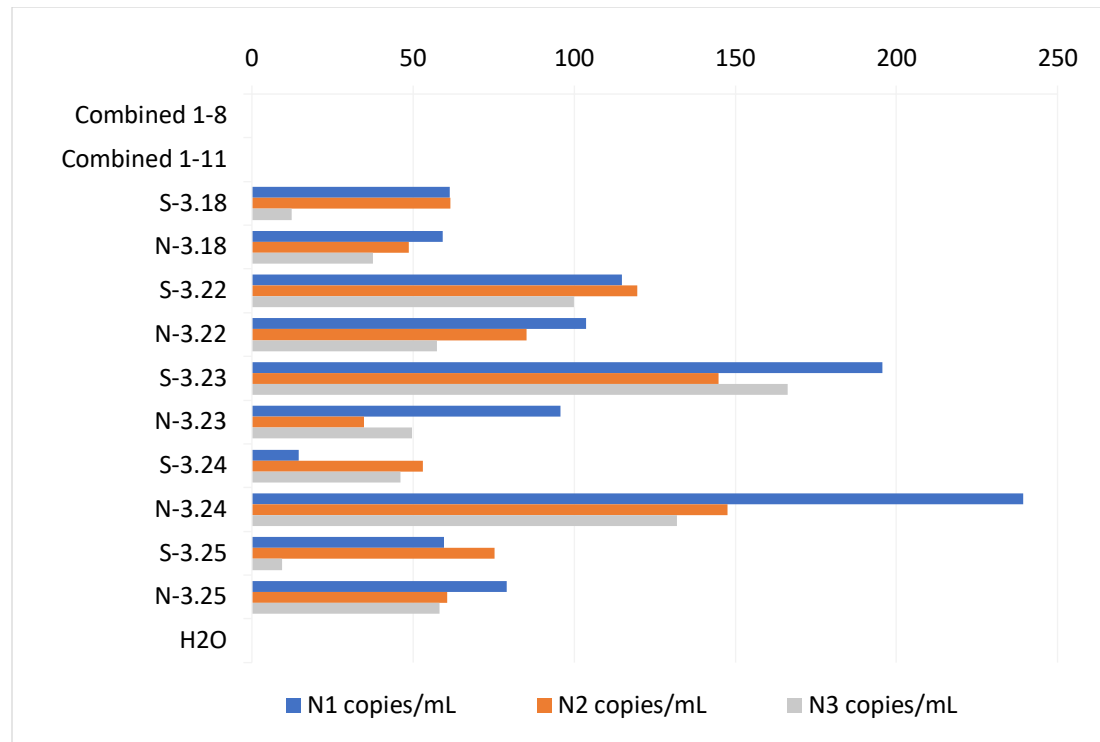
Figure 2. Estimated viral titer per mL of sewage. Estimated gene copies for each of the three CDC primer sets are shown for all dates. Northern (N) and Southern (S) influents are shown separately, except for biobanked January samples, which are combined. H2O represents a PCR negative control.
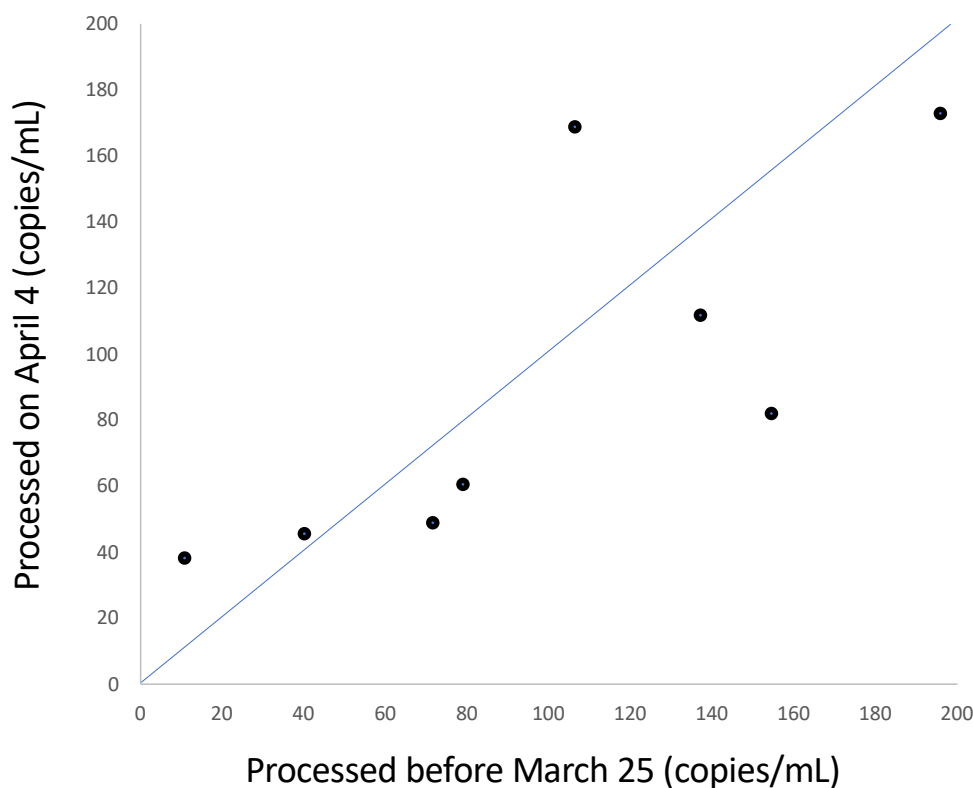
Figure 3. Sample stability over time. Estimated copies per mL sewage estimated from the same samples filtered and processed on different dates. The x-axis shows samples processed between March 20-25, 2020, and are plotted against samples re-processed on April 4, 2020 (samples were pasteurized upon receipt and stored at 4C until processing). The blue line corresponds to x=y. Each point represents an average over all three primer sets for each sample.