

1 [Applicability of Neighborhood and Building Scale Wastewater-Based Genomic Epidemiology to](#)

2 [Track the SARS-CoV-2 Pandemic and other Pathogens.](#)

3

4 Rachel R. Spurbeck^{1*}, Angela T. Minard-Smith¹, Lindsay A. Catlin²

5

6 Affiliations:

7 ¹Health Outcomes and Biotechnology Solutions, Battelle Memorial Institute, 505 King Ave,

8 Columbus, OH 43201; ²National Security Bioscience Center, Battelle Memorial Institute, 505

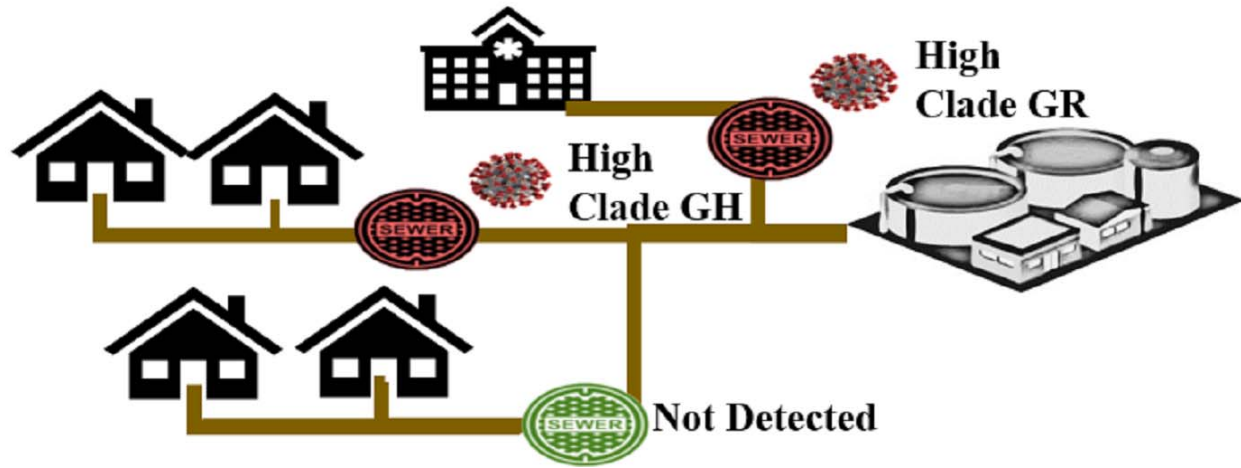
9 King Ave, Columbus, OH 43201

10 *Corresponding author: Rachel R. Spurbeck, electronic mailing address: spurbeck@battelle.org

11

12

13 Graphical Abstract:



15 Highlights

- 16
- 17
- 18
- 19
- 20
- 21
- Neighborhood or building level wastewater analysis accurately detects SARS-CoV-2
 - SARS-CoV-2 was detected in wastewater from one infected person out of 60 residents
 - Total RNAseq did not accurately detect SARS-CoV-2 in wastewater samples.
 - Targeted whole genome sequencing of wastewater samples identified Spike mutations.

22

23 **Abstract**

24 The benefits of wastewater-based epidemiology (WBE) for tracking the viral load of SARS-
25 CoV-2, the causative agent of COVID-19, have become apparent since the start of the pandemic.
26 However, most sampling occurs at the wastewater treatment plant influent and therefore can only
27 monitor SARS-CoV-2 concentration and spread within the entire catchment, which can
28 encompass multiple municipalities. Furthermore, most WBE only quantifies the virus, and
29 therefore miss crucial information that can be gained by sequencing SARS-CoV-2. Here we
30 demonstrate feasibility of sampling at the neighborhood or building complex level using a mix of
31 quantitative polymerase chain reaction (qPCR) and targeted sequencing to provide a more
32 refined understanding of the local dynamics of SARS-CoV-2 strains. When coupled with the
33 higher-level treatment plant samples, this creates an opportunity for health officials to monitor
34 the spread of the virus at different spatial and temporal scales to inform policy decisions.

35 Here we demonstrate the feasibility of tracking SARS-CoV-2 at the neighborhood, hospital,
36 and nursing home level with the ability to detect one COVID-19 positive out of 60 nursing home
37 residents. The viral load obtained was correlative with the number of COVID-19 patients being
38 treated in the hospital. Sequencing of the samples over time demonstrated that nonsynonymous
39 mutations fluctuate in the viral population, and wastewater-based sequencing could be an
40 efficient approach to monitor for vaccine or convalescent plasma escape mutants, as well as
41 mutations that could reduce the efficacy of diagnostics. Furthermore, while SARS-CoV-2 was
42 detected by untargeted RNA sequencing, qPCR and targeted whole genome amplicon
43 sequencing were more reliable methods for tracking the pandemic. From our sequencing data,
44 clades and shifts in mutation profiles within the community were traceable and could be used to
45 determine if vaccine or diagnostics need to be adapted to ensure continued efficacy.

46 **Keywords:**

47 Wastewater based epidemiology, SARS-CoV-2, targeted genome sequencing, mutation analysis,
48 RNAseq, viral load

49 **1.1 Introduction**

50 Wastewater based epidemiology (WBE) is a rapidly growing field for the surveillance of
51 pathogen load within human populations. Since it was discovered that SARS-CoV-2, the
52 causative agent of COVID-19, can be quantified from feces (Wu et al., 2020b), WBE has been
53 adapted to tracking the pandemic. Indeed, quantitative polymerase chain reaction (qPCR)-based
54 wastewater testing has been implemented by several countries and states to track the COVID-19
55 pandemic by monitoring the viral load in wastewater (Achak et al., 2020; Ahmed et al., 2020;
56 Bivins et al., 2020; Crits-Christoph et al., 2021; Foladori et al., 2020; Gonzalez et al., 2020;
57 Gormley et al., 2020; Hamouda et al., 2020; Haramoto et al., 2020; Hart and Halden, 2020;
58 Kitajima et al., 2020; Medema et al., 2020; Murakami et al., 2020; Weidhaas et al., 2020;
59 Westhaus et al., 2021; Wu et al., 2020a). To maximize the information gathered from
60 wastewater-based testing, we propose that targeted sequencing should be used in tandem with
61 qPCR to determine the viral load and identify the dominant viral strain in circulation, including
62 mutations that could affect diagnostic or therapeutic strategies. Furthermore, WBE implemented
63 within catchments at the neighborhood or building level will enable a more refined map of
64 COVID-19 hotspots.

65 WBE has gained popularity during the COVID-19 pandemic due to its ability to track the
66 amount of the virus in a population using non-invasive sampling techniques (Murakami et al.,
67 2020). Most studies sample influent to the wastewater treatment plant, which provides the ability
68 to monitor the entire catchment, but does not have the resolution to gain a more detailed

69 understanding of virus hotspots within that catchment. Hotspot information would be helpful for
70 contact tracers to focus on neighborhoods, or for COVID-19 conscious individuals to determine
71 what part of town is safest to go grocery shopping. The objectives of the current study were to
72 demonstrate the application of WBE to track the spread of SARS-CoV-2 within a city and
73 develop analytical pipelines to maximize information collected from these samples. In this work,
74 wastewater was sampled from manholes located at mapped sewer junction points in the same
75 city to provide a refined picture of COVID-19 prevalence in those communities.

76 WBE primarily focuses on viral load as the measure used to track SARS-CoV-2 prevalence
77 and spread, but the technology exists for obtaining a more nuanced view of the viral strains and
78 mutations that are in the samples. In this work, we explored not only qPCR to determine the viral
79 load, but also targeted whole genome sequencing of SARS-CoV-2, and untargeted RNA
80 sequencing to determine if SARS-CoV-2 can be detected without enrichment along with other
81 pathogens of public health concern. Ideally, a wastewater-based tracking method should be
82 optimized to understand all pathogen dynamics over time in a population. This can be achieved
83 by non-targeted sequencing methods.

84 The sampling locations in this study were chosen specifically to enable comparison of viral
85 load in wastewater to known case load or resident population numbers mapped to the sewer
86 junction points. Two manholes used for sampling were located directly outside of major
87 hospitals known to be treating COVID-19 patients, one was located outside of a nursing home,
88 and a fourth manhole was located downstream of a residential neighborhood. The neighborhood
89 also has a known number of residents, but an unknown number of COVID-19 cases. Therefore,
90 the data collected could inform an estimated viral load in the community which may be
91 correlated to the number of cases. Furthermore, two different RNA extraction methods and two

92 sequencing approaches were used and evaluated to determine the most viable method to provide
93 as much information as possible from each sample.

94 1.2 Materials and Methods

95 *1.2.1 Wastewater sample collection.*

96 One-liter sample collections were conducted at four specific locations in Toledo, OH from
97 manhole access once a week for three weeks. The first week of sampling, at three locations, the
98 sample collections were grab samples due to low flow. The fourth location, Promedica Toledo
99 Hospital, was collected as a 24-hour composite using an autosampler (ISCO 3700C). The
100 subsequent collections at all locations were 24-hour composite samples. In low flow situations, a
101 bucket was used to collect wastewater prior to use of the autosampler to ensure enough volume
102 was present for the composite collection. All samples were collected between 10 AM and 12 PM
103 EST and placed on ice in coolers during transportation to the analytical laboratory. As back up
104 during analysis, 250 mL of each wastewater sample was stored. The following sample metadata
105 was gathered each week during each sample collection: time of collection, method of collection,
106 pH of samples, ambient temperature, and water temperature at time of collection.

107 *1.2.2 Viral particle concentration.*

108 To encourage detachment of virions from organic material in the wastewater, 107 mL of
109 glycine buffer (0.05M glycine, 3% beef extract, pH 9.6) was added to 750 mL of wastewater and
110 mixed for one minute (Vlok et al., 2019). The samples were then centrifuged at 8,000 x g for 30
111 minutes. The supernatant was filtered through 0.22 micron filters to remove bacteria, fungi, or
112 eukaryotic cells present in the samples. The virions were precipitated from the filtrate using a
113 PEG-8000/NaCl solution ((PEG/NaCl) 320g/L PEG-8000, 70 g/L (w/v) NaCl) at a ratio of one-

114 part PEG/NaCl to three parts filtrate (Wu et al., 2020a). Mixtures were gently stirred for 12 hours
115 at 4°C. Precipitated virions were collected by centrifugation for 90 minutes at 13,000 x g at 4°C.
116 Pellets were resuspended in 600 µL 1x PBS. The volume changes throughout processing were
117 captured to enable back calculation of total viral load.

118 *1.2.3 Viral RNA extraction.*

119 Two methods were used for RNA extraction. First, RNA was extracted from the collected
120 viral preparations using TRIzol Reagent (ThermoFisher, Cat# 15596026) following the
121 manufacturer's instructions. RNA quantity and quality were measured by NanoDrop. Second,
122 viral RNA was extracted by QIAamp Viral RNA Mini Kit (Qiagen, Cat# 52904) following the
123 manufacturer's protocol.

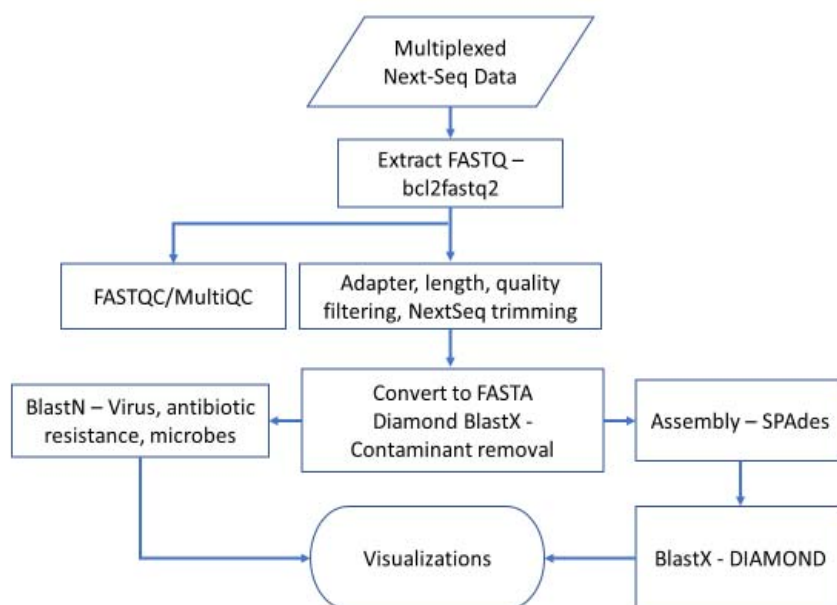
124 *1.2.4 qPCR.*

125 Each RNA sample was analyzed by reverse transcription quantitative polymerase chain
126 reaction (RT-qPCR) targeting the nucleocapsid gene segment (N1) to determine the SARS-CoV-
127 2 viral load. Briefly, 5 µL of 4X TaqMan Fast Viral One-Step Master Mix (ThermoFisher, Cat#
128 4444432) 0.33 µL 60X custom TaqMan Gene Expression Assay primer/probe mix (Forward
129 primer 5'-GAC CCC AAA ATC AGC GAA AT-3'; Reverse primer 5'-TCT GGT TAC TGC
130 CAG TTG AAT CTG-3'; Probe 5'-FAM-ACC CCG CAT TAC GTT TGG TGG ACC-NFQ-
131 MGB-3') (ThermoFisher, Cat# 4331348), 9.67 µL molecular biology grade water, and 5 µL
132 sample were combined for a 20 µL reaction volume. Each plate contained an 8 standard (SARS-
133 CoV-2 synthetic RNA from Biosynthesis, Inc.) serial dilution (1E7 to 1E0) run in triplicate, no
134 template control in triplicate, and samples were in duplicate. Plates were run according to the

135 following thermal cycling parameters: 50°C for 5 min, 95°C for 20 sec, and 40 cycles of 95°C
136 for 3 sec followed by 60°C for 30 sec in an ABI 7500 fast quantitative thermocycler.

137 *1.2.5 RNA Virus metagenomics.*

138 Total RNA was sequenced from each sample. RNA libraries were prepared using the Rapid
139 RNAseq kit (Swift, Cat# R2096) from Swift Biosciences and sequenced on an Illumina NextSeq
140 using two NextSeq 500/550 High Output Kits v2.5 (150 Cycles, 75x 75 bp). No template control
141 libraries were prepared alongside RNAseq samples, but not sequenced. Data were analyzed
142 following the pipeline outlined in the flowchart depicted in Figure 1.



143
144 **Figure 1.** Processing flowchart for Total RNA-Seq data. RNA-Seq data was obtained from runs
145 on an Illumina NextSeq instrument. Files will demultiplexed, quality checked, and trimmed.
146 Sequences were assembled using SPAdes into contigs and then compared to a known reference

147 sequences in the refseq database using BlastN or BlastX- DIAMOND to identify pathogens in
148 the wastewater samples.

149 *1.2.6 COVID-19 amplicon sequencing.*

150 Samples were tested specifically for SARS-CoV-2 by amplicon sequencing using the SNAP
151 SARS-CoV-2 Kit (Swift, Cat# SN-5X296 and COVG1-96) from Swift Biosciences. Briefly, first
152 strand cDNA was synthesized from 11 μ L of total RNA from each sample using the Superscript
153 IV First Strand Synthesis System with random hexamers (Invitrogen, Cat# 18091050) following
154 the manufacturer's instructions. The resulting cDNA (10 μ L) was used as template for the SARS-
155 CoV-2 Kit following the manufacturer's instructions. Prior to sequencing, the samples were
156 quantified by Qubit HS DNA kit (Invitrogen, Cat# Q32851), pooled, and sequenced on an
157 Illumina MiSeq 300 cycle V3 flow cell (Illumina, Cat# MS-102-2002). Samples were batched by
158 week to enable high depth of coverage for single nucleotide polymorphism analysis. The
159 amplicon data were aligned to the NCBI Reference Sequence NC_045512.2 (Severe acute
160 respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1, complete genome) and analyzed using
161 the pipeline provided by Swift Biosciences. A positive control and a no template control were
162 processed and quantified. The no template control was not sequenced, but the positive control
163 was sequenced demonstrating full genome coverage of reference sequence NC_045512.2.

164 *1.4 Results*

165 *1.4.1 Sampling site selection.*

166 Unlike most wastewater epidemiology studies which sample at the influent going into the
167 wastewater treatment plants servicing a particular region of a city, our project studied the
168 biological components in raw sewage derived from manholes in four specific locations in the city

169 of Toledo, Ohio. The sites were carefully chosen through discussion with the city wastewater
170 management team and the Toledo-Lucas County Public Health Department to ensure the sites
171 were in locations with a high probability of SARS-CoV-2 circulation. Two sites were the direct
172 effluent from hospitals known to be actively treating COVID-19 patients at the time of the study,
173 Promedica Toledo Hospital (P, Figure 2A), and Mercy Health Heart and Vascular Institute at St.
174 Vincent Medical Center (V, Figure 2B). A third site was outside a nursing home, Continuing
175 Healthcare Solutions (N, Figure 2C) and a fourth was outside a residential neighborhood,
176 Cresthaven (Ragan Woods, in the western portion of the Southwyck neighborhood, C, Figure
177 2D).



179 **Figure 2.** Sampling sites within Toledo, Ohio. Orange circles indicate manholes sampled. Brown
180 lines indicate sewer lines. A. Promedica Toledo Hospital. B. St. Vincent's Heart Center. C.
181 Cresthaven Neighborhood. D. Continuing Healthcare of Toledo Nursing home.

182 *1.4.2 Known case load at sampling sites by week and qPCR detection of SARS-CoV-2.*

183 The RNA extracted by the TRIzol reagent did not show evidence of any SARS-CoV-2 in the
184 samples, however, the QIAamp Viral/Pathogen RNA preparations produced detectable levels of
185 SARS-CoV-2 by qPCR in 9 of the 12 samples (Table 1). For comparison, the percentage of
186 known COVID-19 cases in the hospitals and nursing home were determined by dividing the
187 number of known COVID-19 positive patients by the total number of inpatients, and a weak
188 positive correlation is observed between the number of infected and viral load in the wastewater.

189 **Table 1.** Percent of positive cases compared to the viral load detected in wastewater.

190

	<i>7/14/20</i>		<i>7/22/20</i>		<i>7/28/20</i>	
<i>Site</i>	<i>Percent Positive (%)</i>	<i>Viral Load (gc/mL)</i>	<i>Percent Positive (%)</i>	<i>Viral Load (gc/mL)</i>	<i>Percent Positive (%)</i>	<i>Viral Load (gc/mL)</i>
<i>C</i>	<i>Unknown</i>	39.1	<i>Unknown</i>	4.5	<i>Unknown</i>	2.9
<i>N</i>	0	ND	0	ND	2	0.6
<i>P</i>	<i>DNP</i>	2.3	<i>DNP</i>	1.0	<i>DNP</i>	ND
<i>V</i>	8	17.4	12	9.0	15	110.6

191

192

193

194

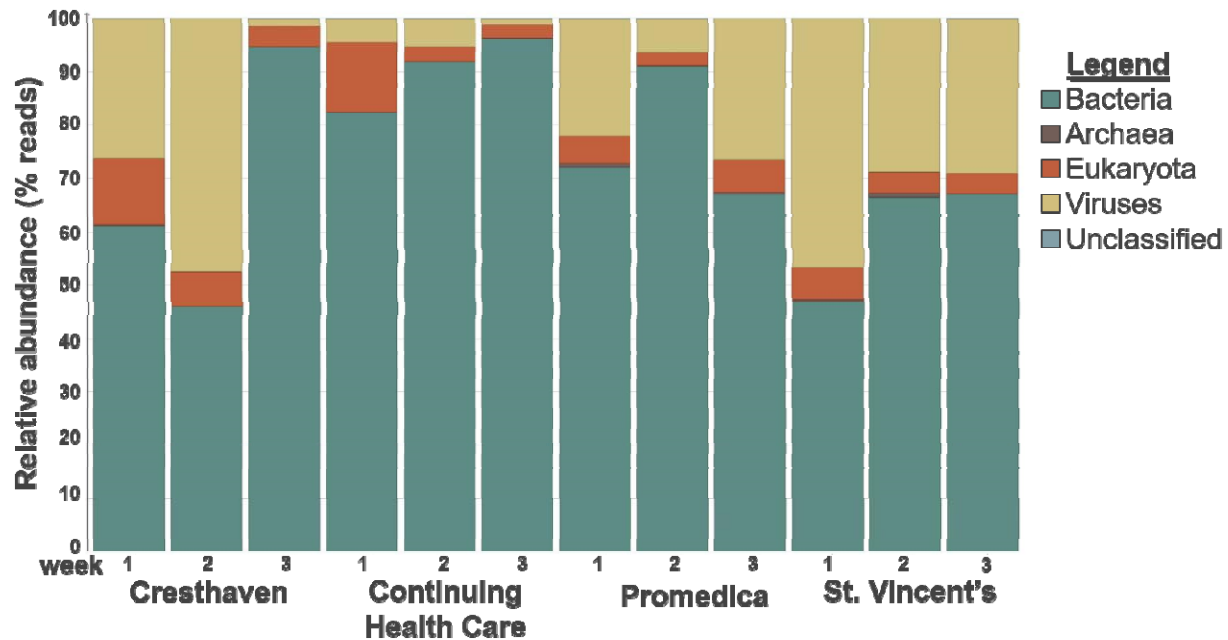
195

196 DNP = Data not provided, ND = not detected

197 *1.4.4 Non-targeted RNAseq detection of SARS-CoV-2*

198 In theory, RNAseq should afford direct SARS-CoV-2 signature detection in wastewater without
199 requiring enrichment. However, despite following viral concentration protocols, a large amount
200 of the sequence data was found to map to other organisms such as bacteria, bacteriophage, and
201 eukaryotic organisms (Figure 3). Reads mapping to SARS-CoV-2 were detected in only five out
202 of the twelve samples, with at most 3 reads mapping to the viral genome, despite there being an
203 average of 11M raw reads per sample and 3M filtered reads per sample. SARS-CoV-2 was
204 detected in Cresthaven samples collected week 1 (3 reads) and week 2 (3 reads), and St.
205 Vincent's samples collected in week 1 (1 read), week 2 (2 reads), and week 3 (2 reads). This low

206 coverage is not sufficient for confidence in reporting presence of the pathogen at the local level.
207 Due to the relatively small genome and other larger genomes present in the wastewater, the
208 SARS-CoV-2 reads were likely obscured. We therefore recommend tracking the pandemic using
209 a local surveillance level (residential neighborhoods or buildings) combined with a targeted
210 enrichment method, such as amplicon-based sequencing.

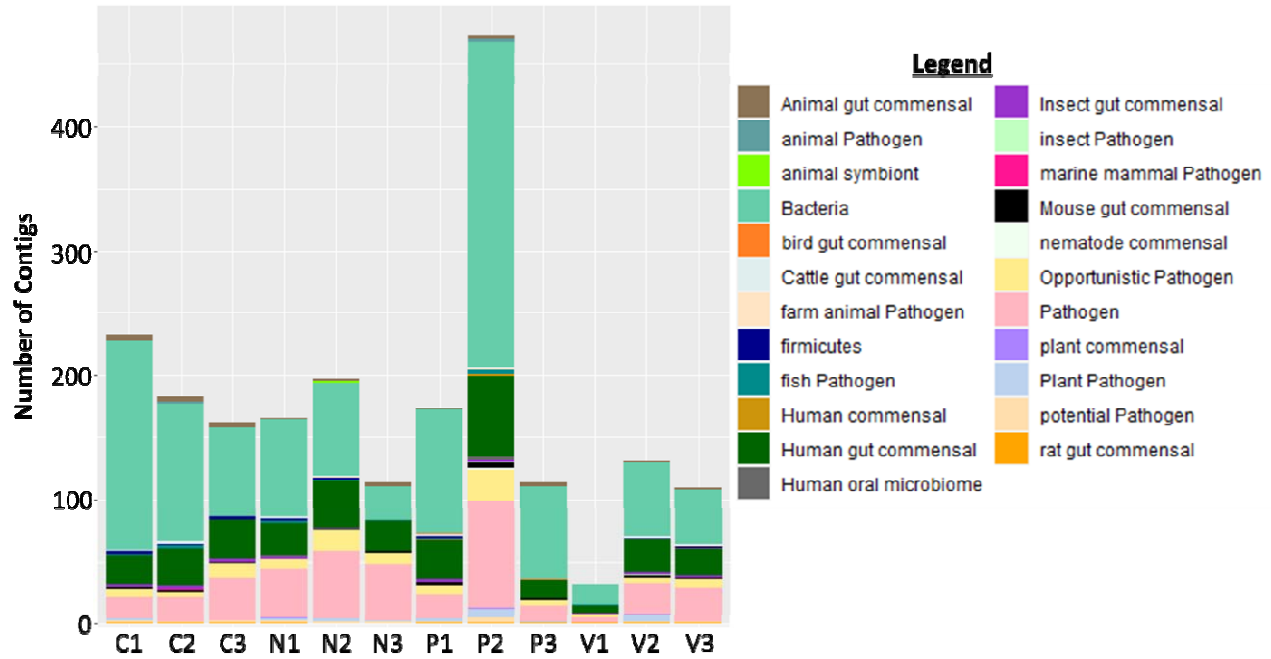


211

212 **Figure 3.** Relative abundance of viruses, bacteria, archaea, and eukaryotes identified by total
213 RNA sequencing of wastewater. Collection dates are indicated as 1 (7/14/20), 2 (7/21/20), or 3
214 (7/28/20).

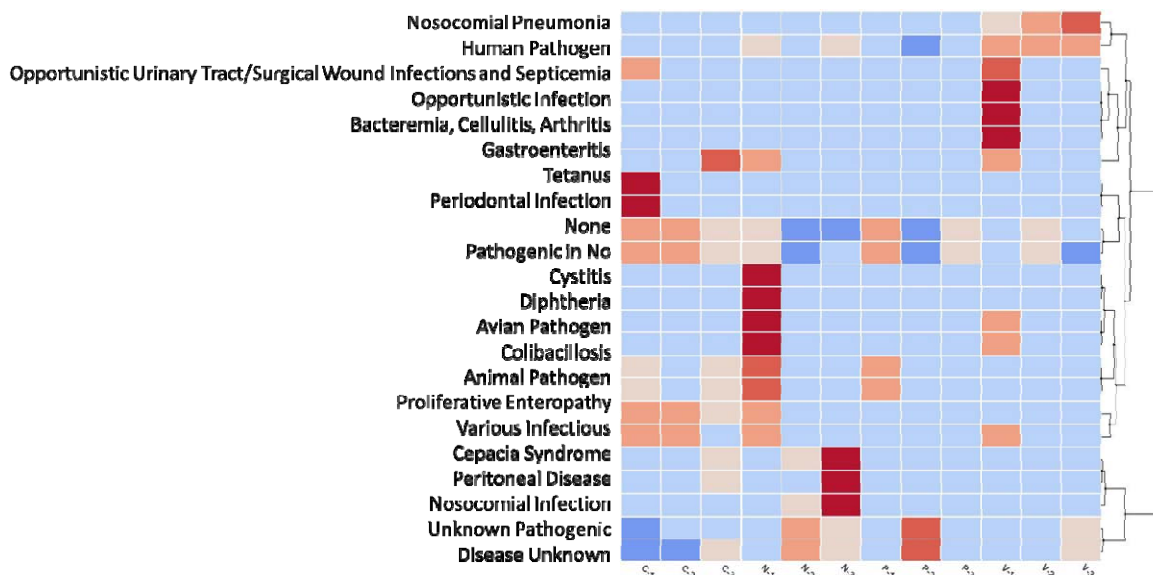
215 Despite removal of microorganisms and enrichment for viral particles, a large portion of the
216 sequencing data (51-90%) mapped to bacteria, and a fraction of the data (1-19%) mapped to
217 macroorganisms, which demonstrates our viral preparation also included RNA from lysed cells
218 (Figure 3). Although not the target of this study, the sequenced microbial RNA identified several
219 commensal species that are nonpathogenic to their hosts, as well as infectious and opportunistic

220 pathogens (Figure 4). While not pertaining to the COVID-19 pandemic, this finding
221 demonstrates that other pathogenic organisms can be identified and tracked using wastewater
222 sequencing, and therefore non-targeted wastewater sequencing could be continued after the
223 pandemic wanes to improve our understanding of pathogen distribution within local
224 communities and healthcare facilities. Several pathogenic categories of interest were detected in
225 the hospitals, nursing home, and community, including microbes associated with nosocomial
226 infections, gastroenteritis, cystitis, pneumonia, septicemia, and tetanus (Figure 5) from
227 assembled contigs. The etiological agents identified for nosocomial infections included the
228 *Acinetobacter* species: *A. baumannii* A. *junii* SH205, and *A. Iwoffii* SH145. *Campylobacter jejuni*
229 *subsp. jejuni* BH-01-0142, a known cause of gastroenteritis, was identified in the Cresthaven
230 neighborhood, the Continuing Healthcare nursing home, and St. Vincent's Hospital. The
231 causative agent of cystitis identified in the first week of testing at Continuing Healthcare was
232 *Escherichia coli* UTI89. This data suggests that WBE can be utilized to track contagions outside
233 of a pandemic and could be a beneficial tool for future epidemiological and biosurveillance
234 studies.



235

236 **Figure 4.** Ecological niches associated with the microorganisms identified by total RNA
237 sequencing identified several human, animal and plant pathogens, as well as potential and
238 opportunistic pathogens. Sample sites are indicated as Nursing home (N), St. Vincent's Heart
239 Center (V), Promedica Toledo Hospital (P), and the Cresthaven Community (C). Collection
240 dates are indicated as 1 (7/14/20), 2 (7/21/20), or 3 (7/28/20).



241
 242 **Figure 5.** The abundance of pathogens identified by total RNA sequencing varied by sampling
 243 site and over time. Notably, pathogens associated with nosocomial infections were found in the
 244 effluent from the nursing home (N) and St. Vincent’s Heart Center (V), but not in Promedica
 245 Toledo Hospital (P) or the Cresthaven Community (C). weeks of collection are indicated as 1
 246 (7/14/20), 2 (7/21/20), or 3 (7/28/20).

247 *1.4.5 Amplicon sequencing*

248 The whole genome of SARS-CoV-2 was enriched from the wastewater samples by targeted
 249 amplicon sequencing, enabling single nucleotide polymorphisms (SNP) and insertion/deletion
 250 mutations (indels) to be identified when aligned to the reference strain SARS-CoV-2 Genome
 251 Wuhan HU-1. Table 2 presents non-synonymous mutations identified in the Toledo samples that
 252 are unique (previously unobserved) or existing mutations (previously observed). Of the twelve
 253 samples analyzed in this study, nine were COVID-19 positive based on qPCR. Sequencing the
 254 positive samples, only five produced full genomes with enough coverage for confident calls:
 255 weeks 1 and 2 from the Cresthaven neighborhood and weeks 1, 2, and 3 from St. Vincent’s
 256 hospital (Table 2 and Figure 6). The mutations identified enabled Clade classification.

257 Community samples taken from the Cresthaven neighborhood fell into Clade GH, which carries
 258 the Spike D614G mutation. St. Vincent’s data showed several mutations at variable penetrance
 259 indicating multiple strains are present in these samples (Figure 6), therefore the Clade was
 260 described as Other. Despite having multiple strains present, all three weeks had SARS-CoV-2
 261 sequences carrying the Spike D614G mutation.

262 **Table 2.** Consensus sequences identified nonsynonymous mutations and categorized the strains
 263 by clade.

Sample	Genome coverage (%)	Length (AA)	#Mutations (%)	Unique Mutations	Existing Mutations	Clade
C-1	97.43%	9688	13 (0.13%)	NSP14_V115I	NSP2_T85I, NSP3_M1788I , NSP5_L89F, NSP12_P323L, NSP14_N129D, NSP16_R216C, Spike_D614G, NS3_Q57H, NS3_G172V, NS8_S24L, N_P67S, N_P199L,	GH
C-2	86.09%	8262	13 (0.16%)	NSP13_G170R NSP14_S418C Spike_N370K	NSP2_T85I, NSP5_L89F, NSP14_N129D, NSP16_R216C, Spike_D614G, NS3_Q57H, NS3_G172V, NS8_S24L, N_P67S, N_P199L	GH
C-3	1.88%	61	--	--	--	--
N-3	8.87%	752	1 (0.13%)			--
P-1	30.66%	2459	4 (0.16%)			GR
P-2	26.31%	2164	7 (0.32%)			GH
V-1	92.78%	9642	14 (0.15%)	NSP3_A416V NSP6_K151R	NSP2_T85I, NSP3_M1788I, NSP5_L89F, NSP12_P323L, NSP14_N129D, NSP16_R216C, Spike_D614G, NS3_Q57H, NS3_G172V , NS8_S24L, N_P67S, N_P199L	Other
V-2	91.66%	8826	15 (0.17%)	NSP3_A1262S NSP3_N1263T NSP14_S418C M_V10I	NSP1_H110Y , NSP2_T85I, NSP3_M1788I, NSP5_L89F, NSP12_P323L, NSP14_N129D, NSP16_R216C, Spike_D614G, NS3_Q57H, NS8_S24L, N_P67S	Other
V-3	99.57%	9710	9 (0.09%)	NSP3_F1107L NSP14_N129B	NSP2_T85I, NSP3_M1788I, NSP12_P323L, NSP16_R216C, Spike_D614G, NS3_G172V , M_A2V	Other

264 *Gray rows indicate samples with partial genome coverage, which may not reliably call the clade. Bold mutations
 265 are ones that were not present in all weeks at the same site.

266

267 A week-to-week comparison of Cresthaven samples yielded highly similar existing
 268 nonsynonymous mutation loads with ten mutations observed in both weeks. In addition, different
 269 unique nonsynonymous mutations were observed, and one existing nonsynonymous mutation
 270 was lost from week 1 to week 2. All three weeks from St. Vincent's had five existing
 271 nonsynonymous mutations in common (NSP2_T85I, NSP3_M1788I, NSP12_P323L,
 272 NSP16_R216C, and Spike_D614G). However, St. Vincent's weeks 1 and 2 had an additional
 273 five existing nonsynonymous mutations in common (NSP5_L89F, NSP14_N129D, NS3_Q57H,
 274 NS8_S24L, and N_P67S), demonstrating that nonsynonymous mutation load has changed over
 275 time.

276 Similarly, the number of SNPs and indels in the whole genome sequencing data changed
 277 over time at each site (Table 3). The three genes most targeted by diagnostics were prone to both
 278 SNPs and indels, demonstrating the importance of continuous monitoring by sequencing as the
 279 pandemic continues. Over the three weeks of this study, the viral population present in the
 280 effluent from St. Vincent's underwent the most striking changes, with the total number of SNPs
 281 and indels increasing each week. Importantly, two deletions found in the S protein in the third
 282 week of sampling (7/28/20) at St. Vincent's was represented in 34% of the data; a deletion of a
 283 full codon, Δ ACC23493-23496, found in 9% of the data, and one was found at 25%: Δ G24696.

284 **Table 3.** SNP frequency at genes targeted for diagnostics or therapeutics and number of indels.

	<i>Cresthaven</i>		<i>St. Vincent's</i>		
	Week 1	Week 2	Week 1	Week 2	Week 3
<i>orf1ab</i>	0.056%	0.038%	0.042%	0.103%	0.211%
<i>S</i>	0.078%	0.078%	0.026%	0.052%	0.235%
<i>N</i>	0.159%	0.159%	0.317%	0.635%	0.238%
<i>all other</i>	0.170%	0.085%	0.142%	0.227%	0.340%
<i>#indels</i>	0	2	3	9	17

285

286 1.5 Discussion

287 The objective of this work was to first demonstrate the utility of wastewater surveillance at
288 the neighborhood or hospital level and then to identify the best method for tracking SARS-CoV-
289 2 and other pathogens to enable a population-based approach to pandemic biosurveillance. Our
290 data have demonstrated that while total RNAseq, targeted whole genome sequencing, and qPCR
291 methodologies can be used to detect SARS-CoV-2 in wastewater, a combination of qPCR to
292 identify viral load and targeted whole genome sequencing to detect mutations provides deeper
293 insights to support public health decisions during pandemic surveillance than qPCR alone.
294 RNAseq, while identifying other pathogens present in the samples such as microbes that cause
295 nosocomial infections, periodontitis, gastroenteritis, pneumonia, or cystitis, did not recover
296 enough reads mapping to SARS-CoV-2 to detect the viral pathogen with any confidence.
297 However, the non-targeted surveillance technique would effectively alert hospitals and long-term
298 care facilities to potential issues with nosocomial pathogens or residents that may have infections
299 that need to be treated and can otherwise go undiagnosed such as cystitis due to asymptomatic or
300 recurrent infections. Use in community settings would alert neighborhoods to potential norovirus
301 or *Salmonella* outbreaks and help public health officials pinpoint restaurants or grocery stores
302 selling contaminated products.

303 Utilizing amplicon sequencing to target and sequence the full SARS-CoV-2 genome enabled
304 mutation analysis of the viral population identified in the wastewater samples. As demonstrated
305 recently, new strains of SARS-CoV-2 are emerging which are hyper transmissible, and
306 potentially could carry mutations in the Spike protein that could be vaccine or convalescent
307 plasma therapy escape mutants (England, 2020; Kemp et al., 2020; Tu et al., 2021). Also, the
308 Spike, N, or Orf1ab genes are diagnostic targets so any SNP or indel identified in these genes

309 that could affect primer or probe binding of a diagnostic test are of high interest and need to be
310 reported to the scientific community to enable improved diagnostics. Indeed, several mutations
311 have been identified that affect different COVID-19 diagnostic tests (Wang et al., 2020).
312 Furthermore, nonsynonymous mutations in the Spike protein could indicate the potential for
313 vaccine escape and will need to be characterized to understand impact on vaccine efficacy.
314 Sequencing from wastewater samples within a city or at hospital effluent can aid in surveillance
315 of these mutations and reduces the burden on supplies, as instead of sampling, quantifying and
316 sequencing SARS-CoV-2 in individuals, wastewater enables population sampling, reducing the
317 number of assays and reagents used during surveillance.

318 With both convalescent plasma and the vaccines targeting the spike protein, it is imperative
319 that nonsynonymous mutations be monitored in this gene. Indeed, it has been demonstrated that
320 spike protein mutations increase in SARS-CoV-2 as a patient is treated with convalescent
321 plasma, indicating the antibodies are placing evolutionary pressure on the virus that selects for
322 mutations in this therapeutic target (Kemp et al., 2020). All Toledo wastewater samples
323 sequenced carried the Spike G614D mutation, which may confer higher transmissibility to the
324 virus, and is characteristic of clade G. Indeed, three of the sequenced samples were categorized
325 as GH and one as GR. In a study in Asia, clades GH and GR were found to contain high levels of
326 SNPs, indicating a lot of diversity in these strains (Sengupta et al., 2020). Mutation analysis of
327 wastewater samples collected in Toledo OH during July 2020 did not find signature for the two
328 SARS-CoV-2 clade 20C/G variants that were later identified in Columbus, OH (Tu et al.,
329 2021). Likewise, the strain UK-B.1.1.7 (clade 20I/501Y.V1)(England, 2020; Galloway et al.,
330 2021), also thought to be hyper transmissible and bearing mutations in the S protein, was not
331 present in July 2020 Toledo, OH wastewater samples. However, another spike mutation,

332 Spike_N370K, was observed in week 1 of the Cresthaven neighborhood, and two deletions
333 (Δ ACC23493-23496 and Δ G24696) were observed in week 3 at St. Vincent's. Therefore,
334 wastewater-based sequencing of SARS-CoV-2 can identify spike protein mutations which may
335 impact transmission and response to the vaccine or convalescent plasma therapy.

336 Most wastewater-based epidemiology takes place at the influent to a wastewater treatment
337 facility, providing a large catchment that encompasses multiple municipalities. In contrast, our
338 work focused on sampling effluent from a neighborhood, two hospitals, and a nursing home.
339 While these sites provided smaller population sizes and can have flow rate issues (some of our
340 sample locations could not provide 24-hour composites due to low flow), sampling at regions
341 within a city provides more refined information on where strains are circulating in the
342 community. For example, strains circulating in Cresthaven were all in the clade GH, while a
343 strain identified in Promedica Toledo Hospital belonged to clade GR, and strains at St. Vincent's
344 were categorized as Other. While all strains bore some mutations in common, each week's
345 sample had mutations unique to that sampling, demonstrating that mutations are common in
346 SARS-CoV-2, and several strains are present in one city.

347 While population size is a concern, qPCR from effluent of our smallest population, that of
348 the nursing home Continuing Healthcare, was able to detect the single case of COVID-19 in
349 week 3, from a population of 60 residents. Since Ohio has implemented weekly testing of
350 nursing home residents, this detection was confirmed by the public health department's data. The
351 case load data from St. Vincent's hospital demonstrated an increase in hospitalized COVID-19
352 cases over time, and our qPCR data correlates with this trend, indicating that with more
353 sampling, a predictive model could be developed and validated to identify an approximate

354 number of infected people present in the catchment over time. Using data from the ongoing
355 study, we are currently working to develop and validate candidate models.

356 Another highly discussed aspect of wastewater-based epidemiology is the degradation of
357 target viral RNA during its transit through the sewer system. While SARS-CoV-2 RNA is intact
358 enough at the inlet to the wastewater treatment plant for qPCR identification, sequencing
359 samples taken closer to the source, such as the effluent from a building or a neighborhood may
360 provide less degraded samples for whole genome sequencing, and therefore provide a better
361 representation of the viral mutations in circulation. In the future, we would like to compare RNA
362 quality of samples from different locations within a catchment to optimize viral sequencing.

363 1.5.1 Conclusions

364 This study has demonstrated that wastewater epidemiology at the local level can be used to
365 determine the viral load in a neighborhood or building and identify strain variants in circulation.
366 Total RNAseq from wastewater samples, while effective in identifying bacterial pathogens in
367 circulation, was not able to consistently detect SARS-CoV-2 for biosurveillance purposes, due to
368 the low abundance of the viral RNA in comparison with all other RNA in the sample. Therefore,
369 it is recommended that a combination of qPCR and targeted sequencing of SARS-CoV-2 be
370 utilized to track the pandemic and provide early warning of new strains circulating in the
371 population which could otherwise evade detection by diagnostics or indicate vaccine escape
372 mutants. The use of wastewater in to track mutations is limited by the methods to
373 bioinformatically deconvolute the viral sequences, and therefore in this work we presented the
374 mutations identified without separating the sequences into individual genomes. Future directions
375 will incorporate long read sequencing or linked read sequencing to enable deconvolution into
376 individual viral strains. Furthermore, a comparison of genomic sequences from samples at

377 wastewater treatment plants versus the local manholes utilized in this study would determine if
378 the time between viral shedding in waste and collection impacts viral RNA integrity limiting the
379 use of genomic sequencing to building or neighborhood wastewater samples for high quality full
380 genome sequences.

381 [Acknowledgements](#)

382 The authors would like to thank Dennis McIntyre, Great Lakes Environmental Center for sample
383 collection; Angela Tucker, City of Toledo, Division of Environmental Services for help with
384 sample site identification and access to sewers; and Dr. Jennifer Gottschalk, Toledo-Lucas
385 County Public Health Department, and St. Vincent's Heart Center for providing COVID-19 case
386 data. We would also like to thank Drs. Jared Schuetter and Trevor Petrel for assistance with
387 critique and editing the manuscript.

388 Funding: This work was supported by the National Science Foundation [grant number:
389 2033137].

390 [References](#)

- 391 Achak M, Alaoui Bakri S, Chhiti Y, M'Hamdi Alaoui FE, Barka N, Boumya W. SARS-CoV-2 in
392 hospital wastewater during outbreak of COVID-19: A review on detection, survival and
393 disinfection technologies. *Sci Total Environ* 2020: 143192.
- 394 Ahmed W, Angel N, Edson J, Bibby K, Bivins A, O'Brien JW, et al. First confirmed detection of
395 SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater
396 surveillance of COVID-19 in the community. *Sci Total Environ* 2020; 728: 138764.
- 397 Bivins A, North D, Ahmad A, Ahmed W, Alm E, Been F, et al. Wastewater-Based
398 Epidemiology: Global Collaborative to Maximize Contributions in the Fight Against
399 COVID-19. *Environmental Science & Technology* 2020; 54: 7754-7757.

400 Crits-Christoph A, Kantor RS, Olm MR, Whitney ON, Al-Shayeb B, Lou YC, et al. Genome
401 Sequencing of Sewage Detects Regionally Prevalent SARS-CoV-2 Variants. *mBio* 2021;
402 12: e02703-20.

403 England PH. Investigation of novel SARS-CoV-2 variant: variant of concern 202012/01,
404 technical briefing 3. Public Health England, London, United Kingdom, 2020.

405 Foladori P, Cutrupi F, Segata N, Manara S, Pinto F, Malpei F, et al. SARS-CoV-2 from faeces to
406 wastewater treatment: What do we know? A review. *Sci Total Environ* 2020; 743:
407 140444.

408 Galloway SE, Paul P, MacCannell D, et al. Emergence of SARS-CoV-2 B.1.1.7 Lineage —
409 United States, December 29, 2020–January 12, 2021. *MMWR Morb Mortal Wkly Rep.* .
410 CDC, 2021.

411 Gonzalez R, Curtis K, Bivins A, Bibby K, Weir MH, Yetka K, et al. COVID-19 surveillance in
412 Southeastern Virginia using wastewater-based epidemiology. *Water Research* 2020; 186:
413 116296.

414 Gormley M, Aspray TJ, Kelly DA. COVID-19: mitigating transmission via wastewater plumbing
415 systems. *Lancet Glob Health* 2020; 8: e643.

416 Hamouda M, Mustafa F, Maraqa M, Rizvi T, Aly Hassan A. Wastewater surveillance for SARS-
417 CoV-2: Lessons learnt from recent studies to define future applications. *Sci Total*
418 *Environ* 2020: 143493.

419 Haramoto E, Malla B, Thakali O, Kitajima M. First environmental surveillance for the presence
420 of SARS-CoV-2 RNA in wastewater and river water in Japan. *Sci Total Environ* 2020;
421 737: 140405.

- 422 Hart OE, Halden RU. Computational analysis of SARS-CoV-2/COVID-19 surveillance by
423 wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities
424 and challenges. *Sci Total Environ* 2020; 730: 138875.
- 425 Kemp SA, Collier DA, Datir R, Ferreira I, Gayed S, Jahun A, et al. Neutralising antibodies in
426 Spike mediated SARS-CoV-2 adaptation. *medRxiv* 2020.
- 427 Kitajima M, Ahmed W, Bibby K, Carducci A, Gerba CP, Hamilton KA, et al. SARS-CoV-2 in
428 wastewater: State of the knowledge and research needs. *Sci Total Environ* 2020; 739:
429 139076.
- 430 Medema G, Heijnen L, Elsinga G, Italiaander R, Brouwer A. Presence of SARS-Coronavirus-2
431 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in the Early Stage
432 of the Epidemic in The Netherlands. *Environmental Science & Technology Letters* 2020;
433 7: 511-516.
- 434 Murakami M, Hata A, Honda R, Watanabe T. Letter to the Editor: Wastewater-Based
435 Epidemiology Can Overcome Representativeness and Stigma Issues Related to COVID-
436 19. *Environ Sci Technol* 2020; 54: 5311.
- 437 Sengupta A, Hassan SS, Choudhury PP. Clade GR and Clade GH Isolates in Asia Show Highest
438 Amount of SNPs. *bioRxiv* 2020: 2020.11.30.402487.
- 439 Tu H, Avenarius MR, Kubatko L, Hunt M, Pan X, Ru P, et al. Distinct Patterns of Emergence of
440 SARS-CoV-2 Spike Variants including N501Y in Clinical Samples in Columbus Ohio.
441 *bioRxiv* 2021: 2021.01.12.426407.
- 442 Vlok M, Lang AS, Suttle CA. Marine RNA Virus Quasispecies Are Distributed throughout the
443 Oceans. *mSphere* 2019; 4: e00157-19.

- 444 Wang R, Hozumi Y, Yin C, Wei GW. Mutations on COVID-19 diagnostic targets. *Genomics*
445 2020; 112: 5204-5213.
- 446 Weidhaas J, Aanderud Z, Roper D, VanDerslice J, Gaddis E, Ostermiller J, et al. Correlation of
447 SARS-CoV-2 RNA in wastewater with COVID-19 disease burden in sewersheds.
448 Research Square, 2020.
- 449 Westhaus S, Weber FA, Schiwy S, Linnemann V, Brinkmann M, Widera M, et al. Detection of
450 SARS-CoV-2 in raw and treated wastewater in Germany - Suitability for COVID-19
451 surveillance and potential transmission risks. *Sci Total Environ* 2021; 751: 141750.
- 452 Wu F, Zhang J, Xiao A, Gu X, Lee WL, Armas F, et al. SARS-CoV-2 Titers in Wastewater Are
453 Higher than Expected from Clinically Confirmed Cases. *mSystems* 2020a; 5: e00614-20.
- 454 Wu Y, Guo C, Tang L, Hong Z, Zhou J, Dong X, et al. Prolonged presence of SARS-CoV-2
455 viral RNA in faecal samples. *Lancet Gastroenterol Hepatol* 2020b; 5: 434-435.
- 456