

1 **High throughput wastewater SARS-CoV-2 detection enables forecasting of community**
2 **infection dynamics in San Diego county**

3

4 Smruthi Karthikeyan¹, Nancy Ronquillo², Pedro Belda-Ferre¹, Destiny Alvarado², Tara
5 Javidi², Christopher A. Longhurst^{1,3}, Rob Knight^{1,4,5,6}

6

7 ¹Department of Pediatrics, University of California, San Diego, La Jolla, California, USA

8 ²Department. of Electrical and Computer Engineering, University of California, San Diego,
9 La Jolla, CA 92093, USA

10 ³Department of Biomedical Informatics, University of California, San Diego, La Jolla, CA,
11 USA

12 ⁴Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA

13 ⁵Department of Computer Science & Engineering, University of California, San Diego, La
14 Jolla, CA, USA

15 ⁶Center for Microbiome Innovation, University of California, San Diego, La Jolla, CA, USA

16

17 **Address correspondence to:**

18

19 Rob Knight

20 Department of Pediatrics

21 University of California San Diego

22 9500 Gilman Drive, MC 0763

23 La Jolla, CA 92093, USA

24 robknight@ucsd.edu

25 +1 858-246-1184

26

27 **Abstract**

28

29 Large-scale wastewater surveillance has the ability to greatly augment the tracking of
30 infection dynamics especially in communities where the prevalence rates far exceed the
31 testing capacity. However, current methods for viral detection in wastewater are severely
32 lacking in terms of scaling up for high throughput. In the present study, we employed an
33 automated magnetic-bead based concentration approach for viral detection in sewage that can
34 effectively be scaled up for processing 24 samples in a single 40-minute run. The method
35 compared favorably to conventionally used methods for viral wastewater concentrations with
36 higher recovery efficiencies from input sample volumes as low as 10ml and can enable the
37 processing of over 100 wastewater samples in a day. The sensitivity of the high-throughput
38 protocol was shown to detect cases as low as 2 in a hospital building with a known COVID-
39 19 caseload. Using the high throughput pipeline, samples from the influent stream of the
40 primary wastewater treatment plant of San Diego county (serving 2.3 million residents) were
41 processed for a period of 13 weeks. Wastewater estimates of SARS-CoV-2 viral genome
42 copies in raw untreated wastewater correlated strongly with clinically reported cases by the
43 county, and when used alongside past reported case numbers and temporal information in an
44 autoregressive integrated moving average (ARIMA) model enabled prediction of new
45 reported cases up to 3 weeks in advance. Taken together, the results show that the high-
46 throughput surveillance could greatly ameliorate comprehensive community prevalence
47 assessments by providing robust, rapid estimates.

48

49 **Importance**

50 Wastewater monitoring has a lot of potential for revealing COVID-19 outbreaks before they
51 happen because the virus is found in the wastewater before people have clinical symptoms.
52 However, application of wastewater-based surveillance has been limited by long processing

53 times specifically at the concentration step. Here we introduce a much faster method of
54 processing the samples, and show that its robustness by demonstrating direct comparisons
55 with existing methods and showing that we can predict cases in San Diego by a week with
56 excellent accuracy, and three weeks with fair accuracy, using city sewage. The automated
57 viral concentration method will greatly alleviate the major bottleneck in wastewater
58 processing by reducing the turnaround time during epidemics.

59

60 **Main**

61 Wastewater-based epidemiology (WBE) can facilitate detailed mapping of the extent
62 and spread of SARS-CoV-2 in a community and has seen a rapid rise in recent months owing
63 to its cost effectiveness as well as its ability to foreshadow trends ahead of diagnostic testing
64 (1-4). With over 46 million cases reported globally, 9.3 million of which are from the United
65 States, there is an imminent need for rapid, community-level surveillance in order to identify
66 potential outbreak clusters ahead of diagnostic data. Previous studies have reported high
67 levels of correlation between viral concentration in sewage to clinically reported cases in a
68 community with trends appearing 2-8 days ahead in wastewater (2, 5). A major bottleneck in
69 large scale wastewater surveillance is the lack of robust, high throughput viral concentration
70 methodology. Conventional techniques for viral concentration from wastewater typically
71 employ laborious or time-consuming processes, namely polyethylene glycol (PEG) based
72 precipitation, direct filtration or ultra-filtration methods that severely limit throughput (6). In
73 the present study, we employed an affinity-capture magnetic hydrogel particle (Nanotrap) -
74 based viral concentration method which was incorporated on the KingFisher Flex liquid-
75 handling robot platform robots (Thermo Fisher Scientific, USA), using a 24-plex head to
76 process 24 samples at once in a 40 min run. RNA is then extracted on the same Kingfisher
77 system for rapid sample processing. Reverse transcription-quantitative polymerase chain

78 reaction (RT-qPCR) targeting N1, N2, E-gene targets was used for detection and
79 quantification of SARS-CoV-2 RNA. Detection of all 3 genes in a sample and its replicate (at
80 C_q values <39) was considered positive. All steps from concentration to RT-qPCR plating
81 were carried hands-free and carried out by liquid handling robots to minimize human error.
82 Using the above pipeline, 96 raw sewage samples were processed in a period of 4.5 hours
83 (concentration to RT-qPCR detection/quantification) effectively reducing the processing time
84 by at least 20-fold. The sensitivity of detection of SARS-CoV-2 viral RNA in wastewater by
85 the high-throughput pipeline (i.e. to determine if individual buildings yield sufficient
86 wastewater signal for high-resolution spatial studies) was established by routinely monitoring
87 the SARS-CoV-2 signatures in the wastewater of a San Diego hospital building housing
88 active COVID-19 patients (7). This site was used as a positive control to test correlations
89 with caseload on a daily basis. Sewage samples were collected daily for a period of 12 weeks
90 during which time the hospital's case load (specific to Covid-19 patients) varied between 2-
91 26. SARS-CoV-2 viral gene copies correlated with the daily hospital caseload ($r=0.75$, Suppl.
92 Fig. S1) suggesting that the wastewater data could at least be used to identify the peaks.
93 SARS-CoV-2 viral RNA was detected in wastewater on all days sampled. Furthermore, the
94 high-throughput protocol has been used as a part of an on-campus wastewater surveillance
95 for the last few months where data from 70 autosamplers covering individual buildings are
96 sampled and analyzed on a daily basis. The method enabled detection of cases as low as 1 in
97 large buildings (*manuscript in preparation, data available on request*).

98

99 The efficiency of high throughput concentration method was compared to the two
100 other commonly implemented concentration methods (electronegative membrane filtration
101 and Polyethylene glycol, PEG-based) using nine-fold serial dilutions of heat-inactivated
102 SARS-CoV-2 viral particles spiked into 10 mL volumes of raw sewage which was previously

103 verified to be from a location with no SARS-CoV-2 prevalence and verified by RT-qPCR
104 (Suppl. Methods). The high-throughput protocol compared favorably to the conventionally
105 used protocols demonstrating its potential implications for large scale sample processing
106 (>100 samples/day) (Suppl. Fig. S2.A, Suppl. Fig. S3, Table S1). The average viral recovery
107 efficiencies were 27% (SD 8%), 15% (SD 9%) and 13% (SD 8%) respectively for the high-
108 throughput, PEG and HA filtration protocol (Suppl. Methods S1). The improved recovery
109 could also be attributed in part due to the specificity of the magnetic beads in recovering viral
110 particles coupled with a magnetic bead-based nucleic acid extraction protocol as well. The
111 protocol was optimized on the KingFisher for optimal bead-binding through homogenization
112 ([dx.doi.org/10.17504/protocols.io.bptemnje](https://doi.org/10.17504/protocols.io.bptemnje)). Furthermore, the automated protocol was
113 hands-free and less prone to user discrepancies than the other 2 methods which were carried
114 out manually.

115

116 24-hour flow weighted composites were collected each day for 3 months between
117 July 20th-October 21st, 2020 from the influent stream of Point Loma wastewater treatment
118 plant. The plant processes over 175 MGD (million gallons per day) of raw sewage and is the
119 primary treatment center for the greater San Diego area serving over 2.3 million residents.
120 SARS-CoV-2 viral RNA was detected in all of the samples processed at an average
121 concentration of 2,010,104 gene copies/L in the influent. Over the course of the study, the
122 clinically reported cases in the county increased by 29,375 (Fig. 1). Peaks in the wastewater
123 data were frequently followed by peaks in the clinically confirmed cases at a later date. This
124 suggests a correlation between wastewater and the number of new cases with the caveat of a
125 time delay, where the wastewater data predicts future trends in the new number of cases.
126 Although informative, this time-lagged correlation alone is not enough for robust predictions.
127 This served as the main motivation to build a predictive model for forecasting the number of

128 new cases per day in San Diego County. In order to model the number of new reported cases,
129 which correlates, not only to wastewater but other complex prevalence and spreading
130 dynamics, we embed useful predictors as separate time series to take advantage of any side
131 information that may improve our forecasting results. Here, the day of the week was tracked
132 for each data point in a third time series to capture any weekly trends for a total of three time
133 series with 88 data points each. We used a data-driven approach to train a prediction model
134 that utilizes prior reported cases, wastewater data and temporal correlations (embedded in the
135 day of the week) in order to forecast the number of new positive cases in San Diego County.
136 The multivariate Autoregressive Integrated Moving Average (ARIMA) model [7] was
137 applied to build a prediction model for the number of new positive cases. The (predicted)
138 number of new cases consist of lagged past values from all three series (number of new cases,
139 wastewater data, day of the week) and each term can be thought of as the influence of that
140 lagged time series on the number of new cases. 65 data points corresponding to data collected
141 from 07/07/2020 to 09/28/2020, were used for developing the predictive model and
142 forecasting estimates. The remaining 24 data points, corresponding to data collected from
143 09/29/2020 to 10/25/2020, were reserved for validating our forecast estimates. Fig. 2B shows
144 the results of the model (prediction and forecast) compared to the observed data. The Pearson
145 correlation coefficients and root mean squared error between the observed data and the
146 predicted model were $r = 0.84, 0.79, 0.69, 0.47$, $RMSE = 57, 50, 59, 70$ for the trained model,
147 1, 2 and 3 week advance forecast values respectively (Suppl. Table S2). Our data-driven
148 approach obtains forecasts successfully capturing general trends on the number of new cases
149 (shown here up to 3 weeks in advance), ultimately reinforcing that wastewater analysis can
150 be expository of previously undetected SARS-CoV-2 infections in the population which
151 could provide useful insight for county officials for the purpose of early public health
152 interventions. As with most environmental samples, there are biases and caveats associated

153 with interpreting data from wastewater quantitatively, there are trade-offs while attempting to
154 scale-up the process to enable the screening of hundreds of samples on a daily basis with a
155 quick turn-around time. However, our data shows that we can still capture low prevalence
156 cases enabling studies at high spatial resolution. Our study demonstrates that high-throughput
157 wastewater-based surveillance can be successfully leveraged to enable creation of a rapid,
158 large-scale early alert system for counties/districts and could be particularly useful in
159 community surveillance in the more vulnerable populations.

160

161 **ACKNOWLEDGEMENTS**

162

163 The authors would like to thank the plant supervisor and the lab team at the Point Loma
164 Wastewater Treatment plant in San Diego for providing us with samples and are grateful for
165 their support in this effort. We thank Dr. Jack Gilbert and the Microbiome Sample Processing
166 Core at UC San Diego for access to qPCR equipment. We also thank Jennifer Holland and
167 the UC San Diego Health analytics team for generating the operational data for analysis. The
168 authors would also like to thank Dr. Aaron Carlin from the Division of Infectious Diseases at
169 UCSD for providing us with heat-inactivated SARS-CoV-2 viral particles. This work was
170 supported by The University of California San Diego Return to Learn program (UCSD-RTL).

171

172 **FIGURE LEGENDS**

173

174 **Figure 1: Tracking infection dynamics in SD county. A.** Map showing the San Diego
175 sewer mains (depicted in purple) that feed into the influent stream at the primary WWTP at
176 Point Loma. Overlaid are the cumulative cases recorded from the different zip codes in the
177 county during the course of the study. The caseload was counted by cases per zip code from

178 areas draining into the WWTP. The circles are proportional to the diagnostic cases reported
179 from each zone and the color gradient shows the cases per 100,000 residents. **B.** Daily new
180 cases reported by the county of San Diego. **C.** SARS-CoV-2 viral gene copies detected per L
181 of raw sewage determined from N1 Cq values corrected for PMMoV concentration (Pepper
182 mild mottle virus). All viral concentration estimates were derived from the processing of 2
183 sample replicates and 2 PCR replicates for each sample.

184

185 **Figure 2: A.** Daily caseload and wastewater viral concentration data shown for a period of 13
186 weeks, where a spline smoothing is applied to each time series to demonstrate general trends.
187 **B.** Predictive model showing the predicted data (yellow) compared to the observed caseload
188 (blue) and the 4-week forecast (red). Data collected from 07/07/2020 to 09/28/2020 were
189 used as the training dataset to predict the caseload for the following weeks (up to
190 10/25/2020). Data (wastewater + county testing data) gathered from 09/29/2020-10/21/2020
191 were used for model validation. MATLAB Systems Identification toolbox was used to
192 estimate the model order and parameters and calculate the forecasted values.

193

194 **Supplemental Data**

195 **Fig. S1:** Plot showing the calculated viral gene copies/L of wastewater compared to the daily
196 active COVID-19 caseload.

197

198 **Fig. S2:** A. Viral concentration method comparison. RT-qPCR N1 Cq values for nine-fold
199 serial dilutions of heat-inactivated SARS-CoV-2 viral particles seeded into 10 mL volumes of
200 raw sewage processed using the PEG concentration protocol (blue), Magnetic-bead based,
201 Nanotrap protocol (purple) and electronegative membrane filtration method (gray). B,C,D.
202 Standard curves for N1,N2 and E gene respectively for 6 fold-serial dilution of heat-

203 inactivated SARS-CoV-2 viral particles spiked into 10ml of raw sewage and concentrated
204 using the high-throughput pipeline. Results for 3 replicates shown.

205

206 **Fig. S3:** Average N1 Cq values for 24 samples run with all 3 concentration methods.

207

208 **Fig. S4:** 1D Amplitude pictures of 2019nCOV ORF1a_ FAM

209

210 **Table S1:** N1 Cq values for the 24 samples concentrated with the 3 methods. SD (%) shows
211 the standard deviation of the 2 replicates per sample run.

212

213 **Table S2:** Observed and predicted response values for the daily number of cases in San
214 Diego county.

215

216 **Table S3:** A. Cycling conditions for RT-qPCR used for SARS-CoV-2 viral RNA detection
217 using the multiplex Promega protocol. B. Cycling conditions for one-step ddPCR for viral
218 quantification of the ORF1ab gene in the heat-inactivated SARS-CoV-2 viral particles used
219 in recovery experiments. C. Dilution series of the heat-inactivated SARS-CoV-2 viral
220 particles used in recovery experiments.

221

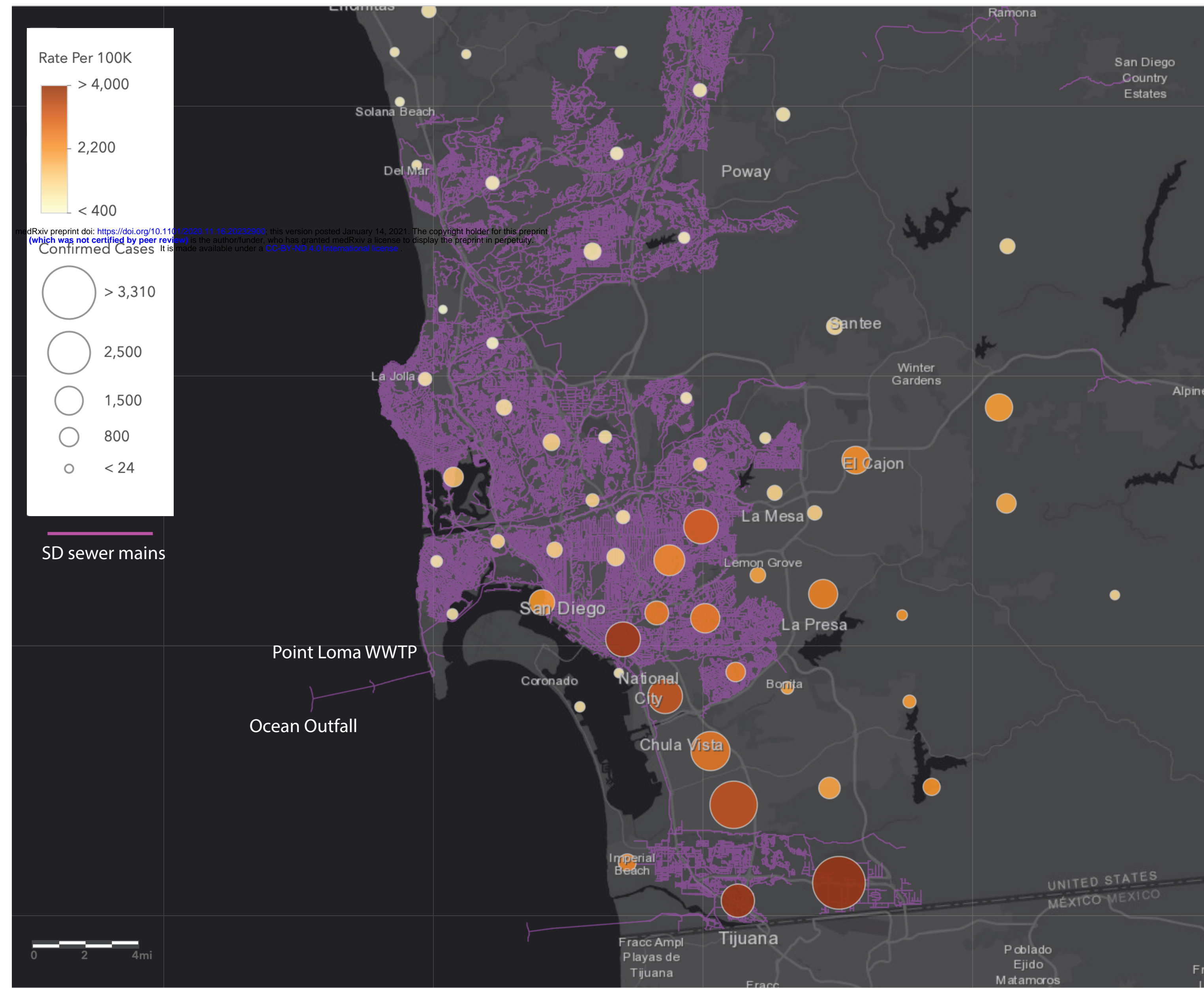
222

223 REFERENCES

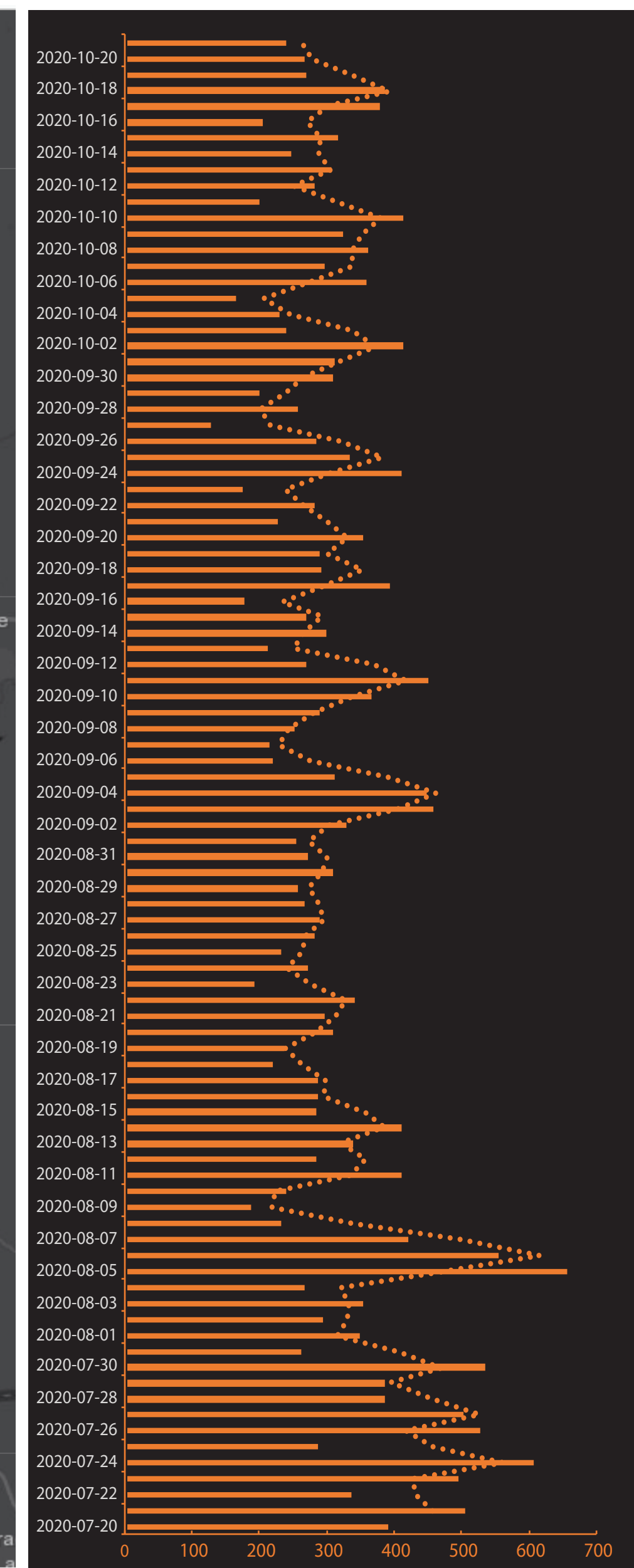
- 224 1. Larsen DA, Wigginton KR. 2020. Tracking COVID-19 with wastewater. *Nature*
225 *Biotechnology* 38:1151-1153.
- 226 2. Peccia J, Zulli A, Brackney DE, Grubaugh ND, Kaplan EH, Casanovas-Massana A,
227 Ko AI, Malik AA, Wang D, Wang M, Warren JL, Weinberger DM, Arnold W, Omer
228 SB. 2020. Measurement of SARS-CoV-2 RNA in wastewater tracks community
229 infection dynamics. *Nature Biotechnology* 38:1164-1167.

- 230 3. Wu F, Xiao A, Zhang J, Moniz K, Endo N, Armas F, Bonneau R, Brown MA,
231 Bushman M, Chai PR, Duvallet C, Erickson TB, Foppe K, Ghaeli N, Gu X, Hanage
232 WP, Huang KH, Lee WL, Matus M, McElroy KA, Nagler J, Rhode SF, Santillana M,
233 Tucker JA, Wuertz S, Zhao S, Thompson J, Alm EJ. 2020. SARS-CoV-2 titers in
234 wastewater foreshadow dynamics and clinical presentation of new COVID-19 cases.
235 medRxiv : the preprint server for health sciences:2020.06.15.20117747.
- 236 4. Ahmed W, Angel N, Edson J, Bibby K, Bivins A, O'Brien JW, Choi PM, Kitajima M,
237 Simpson SL, Li J, Tschärke B, Verhagen R, Smith WJM, Zaugg J, Dierens L,
238 Hugenholtz P, Thomas KV, Mueller JF. 2020. First confirmed detection of SARS-
239 CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater
240 surveillance of COVID-19 in the community. *Science of The Total Environment*
241 728:138764.
- 242 5. Medema G, Heijnen L, Elsinga G, Italiaander R, Brouwer A. 2020. Presence of
243 SARS-Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19
244 Prevalence in the Early Stage of the Epidemic in The Netherlands. *Environmental*
245 *Science & Technology Letters* 7:511-516.
- 246 6. Ahmed W, Bertsch PM, Bivins A, Bibby K, Farkas K, Gathercole A, Haramoto E,
247 Gyawali P, Korajkic A, McMinn BR, Mueller JF, Simpson SL, Smith WJM,
248 Symonds EM, Thomas KV, Verhagen R, Kitajima M. 2020. Comparison of virus
249 concentration methods for the RT-qPCR-based recovery of murine hepatitis virus, a
250 surrogate for SARS-CoV-2 from untreated wastewater. *Science of The Total*
251 *Environment* 739:139960.
- 252 7. Reeves JJ, Hollandsworth HM, Torriani FJ, Taplitz R, Abeles S, Tai-Seale M, Millen
253 M, Clay BJ, Longhurst CA. 2020. Rapid response to COVID-19: health informatics
254 support for outbreak management in an academic health system. *Journal of the*
255 *American Medical Informatics Association* 27:853-859.
256

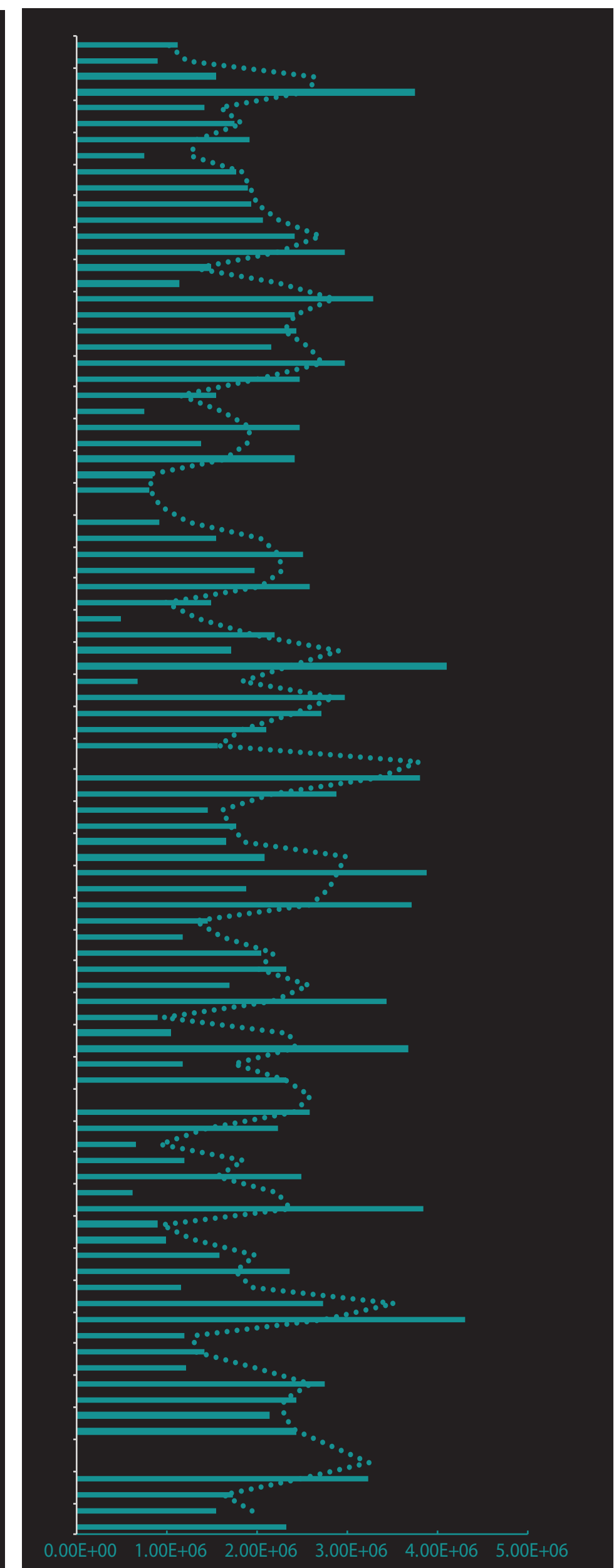
A

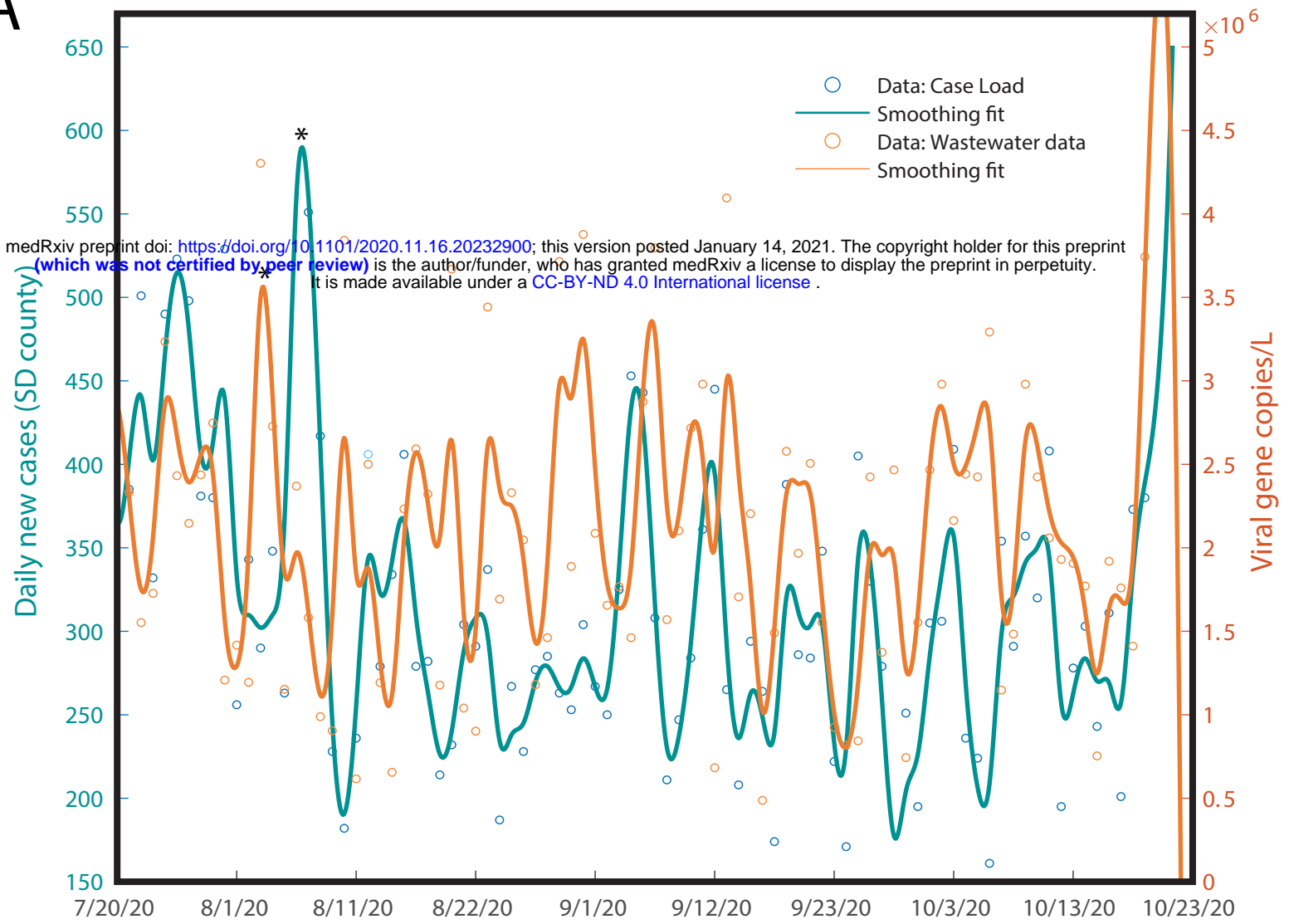


B



C



A**B**