

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, Rob Knight^{1,3,4,5}

6

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

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

10 ³ Department of Bioengineering, University of California, San Diego, La Jolla, CA, USA

11 ⁴ Department of Computer Science & Engineering, University of California, San Diego, La
12 Jolla, CA, USA

13 ⁵ Center for Microbiome Innovation, University of California, San Diego, La Jolla, CA, USA

14

15 **Address correspondence to:**

16

17 Rob Knight

18 Department of Pediatrics

19 University of California San Diego

20 9500 Gilman Drive, MC 0763

21 La Jolla, CA 92093, USA

22 robknight@ucsd.edu

23 +1 858-246-1184

24

25 **Abstract**

26 Large-scale wastewater surveillance has the ability to greatly augment the tracking of
27 infection dynamics especially in communities where the prevalence rates far exceed the
28 testing capacity. However, current methods for viral detection in wastewater are severely
29 lacking in terms of scaling up for high throughput. In the present study, we employed an
30 automated magnetic-bead based concentration approach for viral detection in sewage that can
31 effectively be scaled up for processing 24 samples in a single 40-minute run. The method
32 compared favorably to conventionally used methods for viral wastewater concentrations with
33 a limit of detection of 8.809 viral gene copies/ml from input sample volumes as low as 10ml
34 and can enable the processing of over 100 wastewater samples in a day. Using the high
35 throughput pipeline, samples from the influent stream of the primary wastewater treatment
36 plant of San Diego county (serving 2.3 million residents) were processed for a period of 13
37 weeks. Wastewater estimates of SARS-CoV-2 viral genome copies in raw untreated
38 wastewater correlated strongly with clinically reported cases by the county, and enabled
39 prediction of cases up to 3 weeks in advance. Taken together, the results show that the high-
40 throughput surveillance could greatly ameliorate comprehensive community prevalence
41 assessments by providing robust, rapid estimates.

42

43 **Importance**

44 Wastewater monitoring has a lot of potential for revealing COVID-19 outbreaks before they
45 happen because the virus is found in the wastewater before people have clinical symptoms.
46 However, application of wastewater based surveillance has been limited by long processing
47 times specifically at the concentration step. Here we introduce a much faster method of
48 processing the samples, and show that its robustness by demonstrating direct comparisons
49 with existing methods and showing that we can predict cases in San Diego by a week with

50 excellent accuracy, and three weeks with fair accuracy, using city sewage. The automated
51 viral concentration method will greatly alleviate the major bottleneck in wastewater
52 processing by reducing the turnaround time during epidemics.

53

54 **Main**

55 Wastewater-based epidemiology (WBE) can facilitate detailed mapping of the extent
56 and spread of SARS-CoV-2 in a community and has seen a rapid rise in recent months owing
57 to its cost effectiveness as well as its ability to foreshadow trends ahead of diagnostic testing
58 [1-4] . With over 46 million cases reported globally, 9.3 million of which are from the
59 United States, there is an imminent need for rapid, community-level surveillance in order to
60 identify potential outbreak clusters ahead of diagnostic data. Previous studies have reported
61 high levels of correlation between viral concentration in sewage to clinically reported cases in
62 a community with trends appearing 2-8 days ahead in wastewater [2, 5]. A major bottleneck
63 in large scale wastewater surveillance is the lack of robust, high throughput viral
64 concentration methodology. Conventional techniques for viral concentration from wastewater
65 typically employ laborious or time-consuming processes, namely polyethylene glycol (PEG)
66 based precipitation, direct filtration or ultra-filtration methods that severely limit throughput
67 [6]. In the present study, we employed a magnetic bead-based viral concentration method
68 which was incorporated on the KingFisher Flex liquid-handling robot platform robots
69 (Thermo Fisher Scientific, USA), using a 24-plex head to process 24 samples at once in a 40
70 min run. RNA is then extracted on the same Kingfisher system for rapid sample processing.
71 Reverse transcription-quantitative polymerase chain reaction (RT-qPCR) targeting N1, N2,
72 E-gene targets was used for detection and quantification of SARS-CoV-2 RNA. Using the
73 above pipeline, 96 raw sewage samples were processed in a period of 5 hours (concentration

74 to RT-qPCR detection/quantification) effectively reducing the processing time by at least 20-
75 fold.

76

77 The efficiency of high throughput concentration method was compared to the two
78 other commonly implemented concentration methods (electronegative membrane filtration
79 and Polyethylene glycol, PEG-based) using nine-fold serial dilutions of heat-inactivated
80 SARS-CoV-2 viral particles spiked into 10 mL volumes of raw sewage which was previously
81 verified to be from a location with no SARS-CoV-2 prevalence (Suppl. Methods). The high-
82 throughput protocol compared favorably to the conventionally used protocols demonstrating
83 its potential implications for large scale sample processing (>100 samples/day) (Suppl. Fig.
84 S1.A).

85

86 24-hour flow weighted composites were collected each day for 3 months between
87 July 20th-October 21st, 2020 from the influent stream of Point Loma wastewater treatment
88 plant. The plant processes over 175 MGD (million gallons per day) of raw sewage and is the
89 primary treatment center for the greater San Diego area serving over 2.3 million residents.
90 SARS-CoV-2 viral RNA was detected in all of the samples processed at an average
91 concentration of 2,010,104.45 gene copies/L in the influent and 420,000–580,000 gene
92 copies/L at ocean outfall (effluent). Over the course of the study, the clinically reported cases
93 in the county increased by 29,375 (Fig. 1). Peaks in the wastewater data were frequently
94 followed by peaks in the clinically confirmed cases at a later date. This suggests a correlation
95 between wastewater and the number of new cases with the caveat of a time delay, where the
96 wastewater data predicts future trends in the new number of cases. Although informative, this
97 time-lagged correlation alone is not enough for robust predictions. This served as the main
98 motivation to build a predictive model for forecasting the number of new cases per day in

99 San Diego County. Additionally, the day of the week was tracked for each data point in a
100 third time series to capture any weekly trends for a total of three time series with 88 data
101 points each. We used a data-driven approach to train a prediction model that utilizes
102 wastewater data and temporal correlations (embedded in the day of the week) in order to
103 forecast the number of new positive cases in San Diego County. The multivariate
104 Autoregressive Moving Average (ARMA) model [7] was applied to build a prediction model
105 for the number of new positive cases. The (predicted) number of new cases consist of lagged
106 past values from all three series (number of new cases, wastewater data, day of the week) and
107 each term can be thought of as the influence of that lagged time series on the number of new
108 cases. 65 data points were used for developing the predictive model and the remaining 24
109 were reserved for validating our forecast estimates. Fig. 2B shows the results of the model
110 (prediction and forecast) compared to the observed data. The Pearson correlation coefficients
111 between the observed data and the predicted model were $r = 0.84, 0.79, 0.69, 0.47$ for current,
112 1, 2 and 3 week advance forecast values respectively (Suppl. Table S1). Our data-driven
113 approach obtains forecasts successfully capturing general trends on the number of new cases
114 (shown here up to 3 weeks in advance), ultimately reinforcing that wastewater analysis can
115 be expository of previously undetected SARS-CoV-2 infections in the population which
116 could provide useful insight for county officials for the purpose of early public health
117 interventions. Our study demonstrates that high-throughput wastewater-based surveillance
118 can be successfully leveraged to enable creation of a rapid, large-scale early alert system for
119 counties/districts and could be particularly useful in community surveillance in the more
120 vulnerable populations.

121

122

123

124 **ACKNOWLEDGEMENTS**

125

126 The authors would like to thank the plant supervisor and the lab team at the Point Loma
127 Wastewater Treatment plant in San Diego for providing us with samples and are grateful for
128 their support in this effort. We thank Dr. Jack Gilbert and the Microbiome Sample Processing
129 Core at UC San Diego for access to qPCR equipment. This work was supported by The
130 University of California San Diego Return to Learn program (UCSD-RTL).

131

132 **FIGURE LEGENDS**

133

134 **Figure 1: Tracking infection dynamics in SD county. A.** Map showing the San Diego
135 sewer mains (depicted in purple) that feed into the influent stream at the primary WWTP at
136 Point Loma. Overlaid are the cumulative cases recorded from the different zip codes in the
137 county during the course of the study. The circles are proportional to the diagnostic cases
138 reported from each zone and the color gradient shows the cases per 100,000 residents. **B.**
139 Daily new cases reported by the county of San Diego. **C.** SARS-CoV-2 viral gene copies
140 detected per L of raw sewage. All viral concentration estimates were derived from the
141 processing of 2 biological replicates.

142

143 **Figure 2: A.** Daily caseload and wastewater viral concentration data shown for a period of 13
144 weeks, where a spline smoothing is applied to each time series to demonstrate general trends.
145 **B.** Predictive model showing the predicted data (yellow) compared to the observed caseload
146 (blue) and the 4-week forecast (red). Data collected from 07/07/2020 to 09/28/2020 were
147 used as the training dataset to predict the caseload for the following weeks (up to
148 10/25/2020). Data (wastewater + county testing data) gathered from 09/29/2020-10/21/2020

149 were used for model validation. MATLAB Systems Identification toolbox was used to
150 estimate the model order and parameters and calculate the forecasted values.

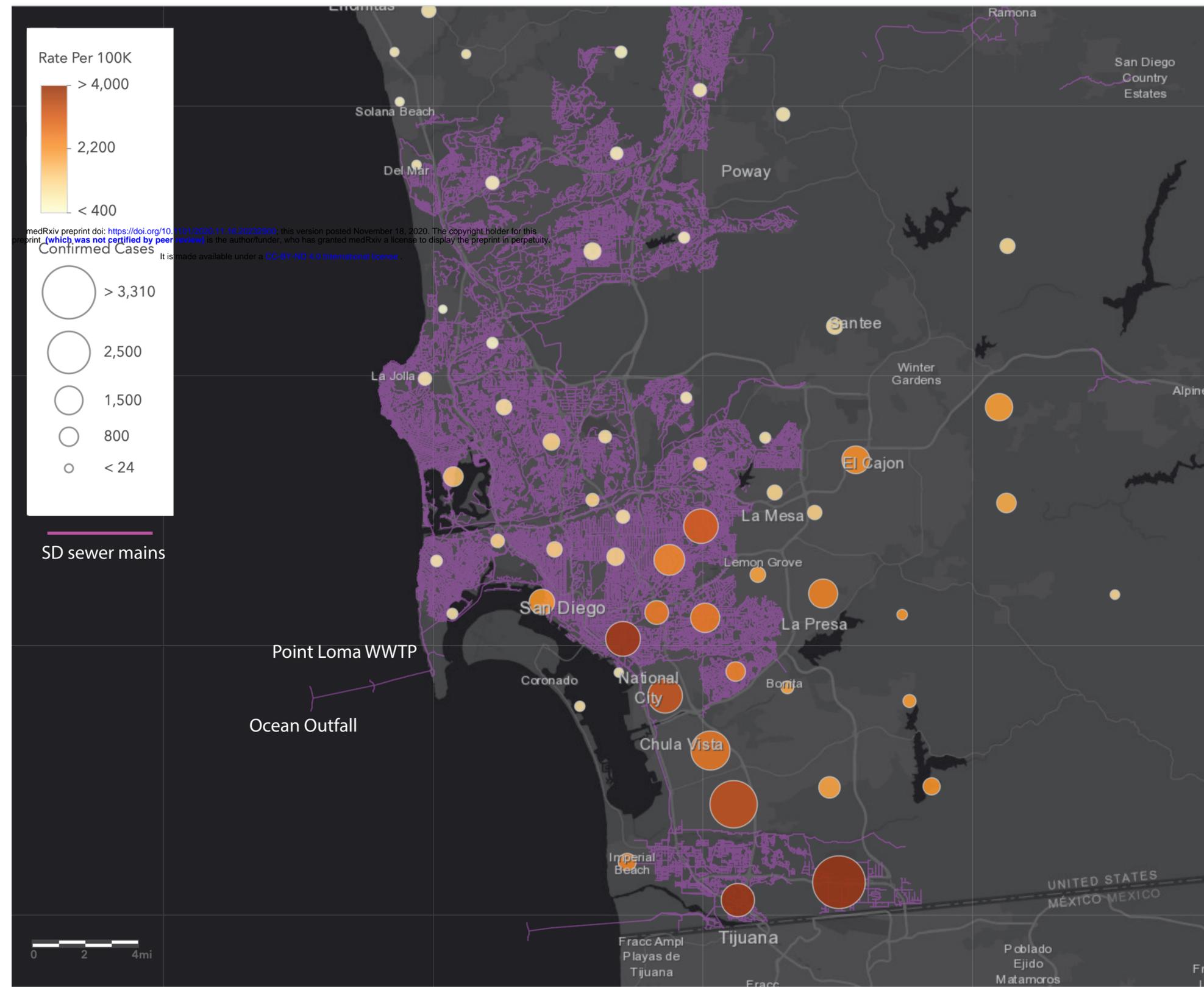
151

152

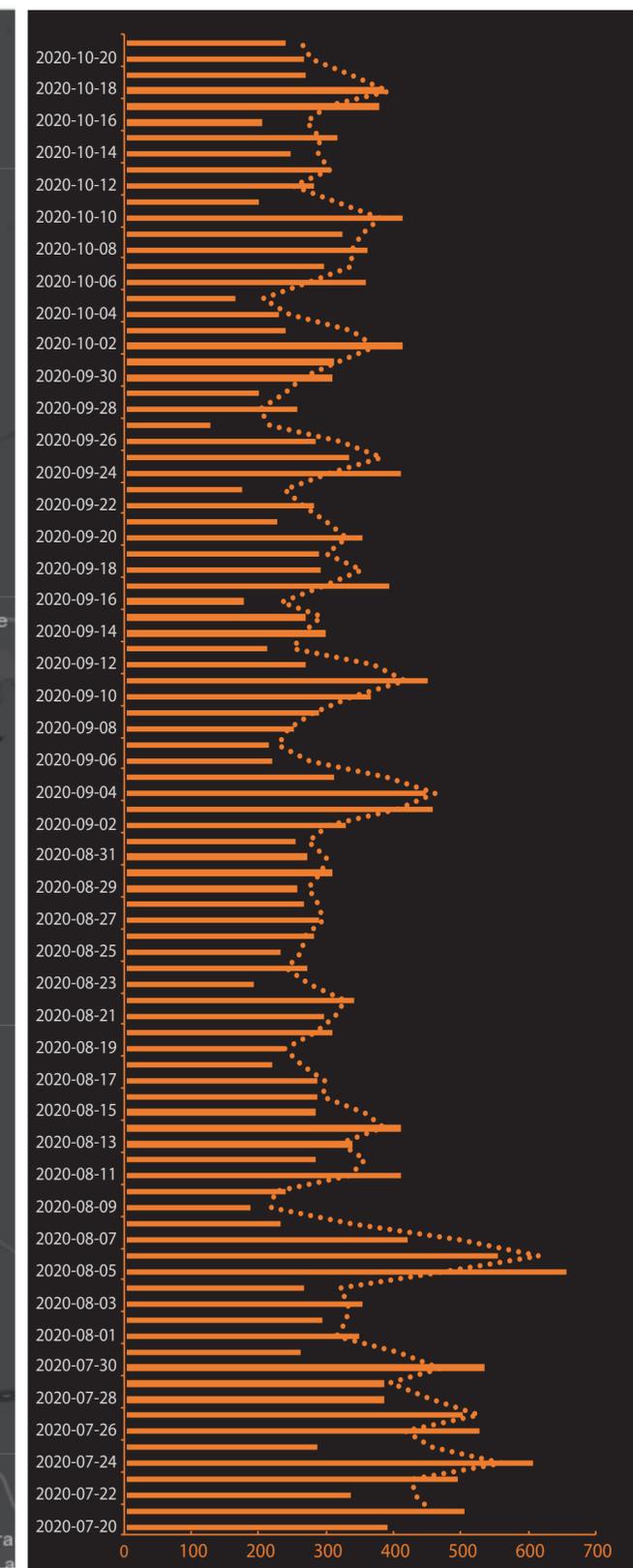
153 REFERENCES

- 154 1. Larsen, D.A. and K.R. Wigginton, *Tracking COVID-19 with wastewater*. Nature
155 Biotechnology, 2020. **38**(10): p. 1151-1153.
- 156 2. Peccia, J., et al., *Measurement of SARS-CoV-2 RNA in wastewater tracks community*
157 *infection dynamics*. Nature Biotechnology, 2020. **38**(10): p. 1164-1167.
- 158 3. Wu, F., et al., *SARS-CoV-2 titers in wastewater foreshadow dynamics and clinical*
159 *presentation of new COVID-19 cases*. medRxiv : the preprint server for health
160 sciences, 2020: p. 2020.06.15.20117747.
- 161 4. Ahmed, W., et al., *First confirmed detection of SARS-CoV-2 in untreated wastewater*
162 *in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the*
163 *community*. Science of The Total Environment, 2020. **728**: p. 138764.
- 164 5. Medema, G., et al., *Presence of SARS-Coronavirus-2 RNA in Sewage and Correlation*
165 *with Reported COVID-19 Prevalence in the Early Stage of the Epidemic in The*
166 *Netherlands*. Environmental Science & Technology Letters, 2020. **7**(7): p. 511-516.
- 167 6. Ahmed, W., et al., *Comparison of virus concentration methods for the RT-qPCR-*
168 *based recovery of murine hepatitis virus, a surrogate for SARS-CoV-2 from untreated*
169 *wastewater*. Science of The Total Environment, 2020. **739**: p. 139960.
- 170 7. Brockwell, P.J., R.A. Davis, and S.E. Fienberg, *Time series: theory and methods:*
171 *theory and methods*. 1991: Springer Science & Business Media.
- 172 8. Barclay, R.A., et al., *Nanotrap particles improve detection of SARS-CoV-2 for pooled*
173 *sample methods, extraction-free saliva methods, and extraction-free transport*
174 *medium methods*. bioRxiv, 2020: p. 2020.06.25.172510.
- 175 9. Corman, V.M., et al., *Detection of 2019 novel coronavirus (2019-nCoV) by real-time*
176 *RT-PCR*. Eurosurveillance, 2020. **25**(3): p. 2000045.
- 177

A

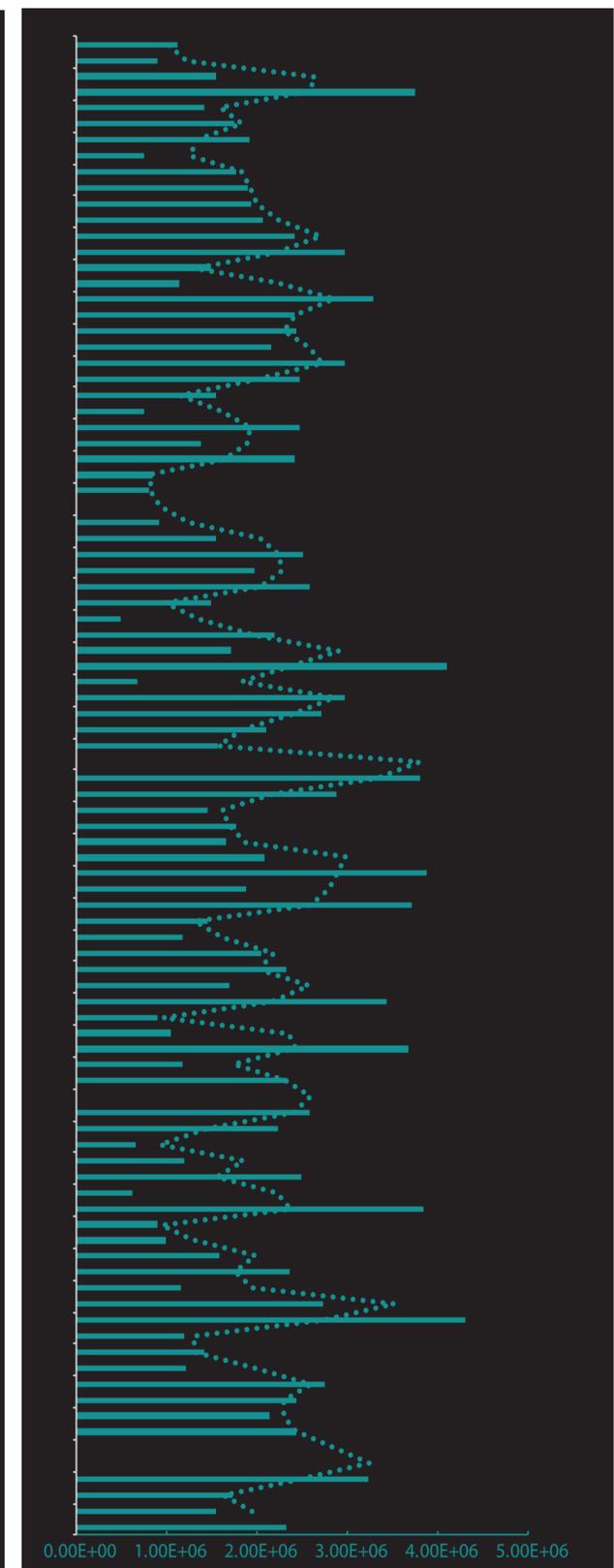


B

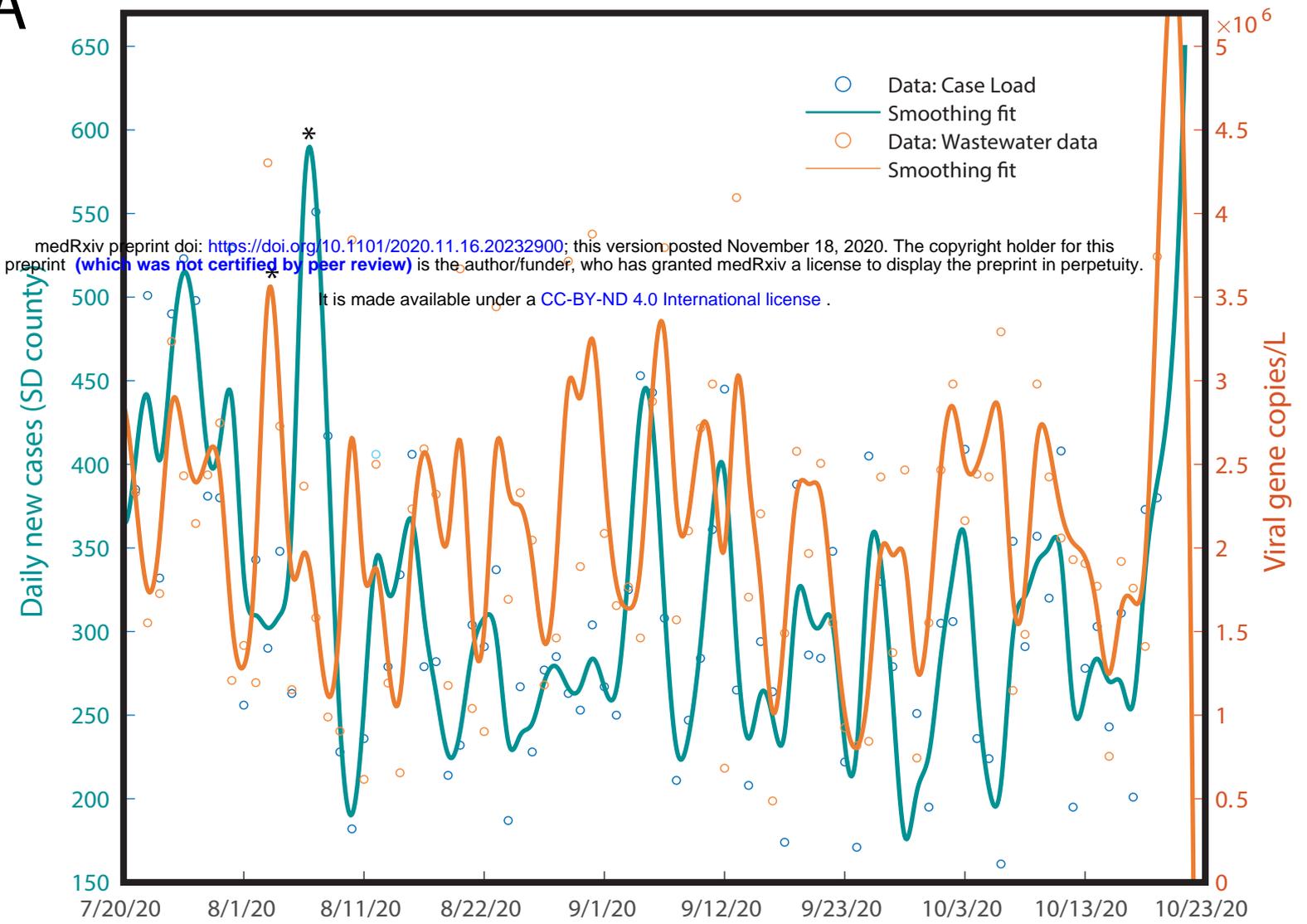


Reported cases

C



Viral gene copies/L

A**B**