Wastewater-based epidemiology predicts COVID-19 community prevalence

Blythe Layton
Clean Water Services
Devrim Kaya
Oregon State University https://orcid.org/0000-0002-7115-0845
Christine Kelly
Oregon State University
Kenneth Williamson
Clean Water Services
Silke Bachhuber
Oregon State University
Peter Banwarth
Benton County Health Department
Jeffrey Bethel
Oregon State University
Katherine Carter
Oregon State University
Benjamin Dalziel
Oregon State University https://orcid.org/0000-0002-9674-2783
Mark Dasenko
Oregon State University
Matthew Geniza
Oregon State University https://orcid.org/0000-0003-4828-7891
Dana Gibbon
Oregon State University
Anne-Marie Girard
Oregon State University
Roy Haggerty
Oregon State University
Kathryn Higley
Oregon State University
Denise Hynes
Oregon State University
Jane Lubchenco

Oregon State University https://orcid.org/0000-0003-3540-5879
Katherine McLaughlin
Oregon State University
F. Javier Nieto
Oregon State University
Aslan Noakes
Oregon State University
Matthew Peterson
Oregon State University
Adriana Piemonti
Clean Water Services
Justin Sanders
Oregon State University
Brett Tyler
Oregon State University https://orcid.org/0000-0003-1549-2987
Tyler Radniecki (tyler.radniecki@oregonstate.edu)
tyler.radniecki@oregonstate.edu https://orcid.org/0000-0002-5295-3562

Biological Sciences - Article

Keywords: COVID-19, community prevalence, wastewater

Posted Date: July 8th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-690031/v1

Wastewater-based epidemiology predicts COVID-19 community prevalence

Blythe A. Layton[1,2], Devrim Kaya[1], Christine Kelly[1], Kenneth J. Williamson[2], Silke M. Bachhuber[3], Peter G. Banwarth[4], Jeffrey W. Bethel[5], Katherine Carter[6], Benjamin D. Dalziel[3,7], Mark Dasenko[6], Matthew Geniza[6], Dana Gibbon[6], Anne-Marie Girard[6], Roy Haggerty[8], Kathryn A. Higley[9], Denise Hynes[3,10], Jane Lubchenco[3], Katherine R. McLaughlin[11], F. Javier Nieto[12], Aslan Noakes[8], Matthew Peterson[6], Adriana D. Piemonti[2], Justin L. Sanders[13], Brett M. Tyler[6,14], and Tyler S. Radniecki[1*]

[1]School of Chemical, Biological, and Environmental Engineering, Oregon State University, Corvallis, OR, USA

[2]Department of Research and Innovation, Clean Water Services, Hillsboro, OR, USA

[3]School of Biological and Population Health Sciences, Oregon State University, Corvallis, OR, USA

[3]Department of Integrative Biology, Oregon State University, Corvallis, OR, USA

[4]Benton County Health Department, Corvallis, OR, USA

[5]School of Biological and Population Health Sciences, Oregon State University, Corvallis, OR, USA

[6]Center for Genome Research and Biocomputing, Oregon State University, Corvallis, OR, USA

[7]Department of Mathematics, Oregon State University, Corvallis, OR, USA

[8]College of Science, Oregon State University, Corvallis, OR, USA

[9]School of Nuclear Science and Engineering, Oregon State University, Corvallis, OR, USA

[10]School of Social and Behavioral Health Sciences, Oregon State University, Corvallis, OR, USA

[11]Department of Statistics, Oregon State University, Corvallis, OR, USA

[12]College of Public Health and Human Sciences, Oregon State University, Corvallis, OR, USA
[13]Carlson College of Veterinary Medicine, Oregon State University, Corvallis, OR, USA
[14]Department of Botany and Plant Pathology, Oregon State University, Corvallis, OR, USA

*Correspondence to: tyler.radniecki@oregonstate.edu

Abstract

Wastewater-based epidemiology uses pooled wastewater samples to monitor community health and has been used extensively during the COVID-19 pandemic to track SARS-CoV-2 RNA shed by infected individuals into wastewater. Wastewater concentrations of SARS-CoV-2 RNA have been positively correlated with contemporaneous counts of COVID-19 cases, making it useful for following relative disease burden trends within a community. However, the statistical associations are too weak for wastewater-based epidemiology to reliably predict reported case counts, limiting its potential. Here we show that wastewater SARS-CoV-2 concentrations are highly correlated with the community prevalence estimated from 8 randomized household community surveys in 6 Oregon communities over a 10-month period. We found that wastewater-based epidemiology is a significantly better predictor of COVID-19 community prevalence than reported case counts, which suffer from systematic biases including variations in access to testing and underreporting of asymptomatic cases, even after accounting for uncertainty inherent in the wastewater and prevalence estimates by using Monte Carlo simulations. Additionally, our results show that wastewater-based epidemiology can identify the rise and fall of neighborhood-scale COVID-19 hot spots and provide rapid information about the presence of SARS-CoV-2 variants at the neighborhood- and city-scale through sequence analyses of the wastewater. These results validate the potential of wastewater-based epidemiology to be a

quantitative method to predict the prevalence of SARS-CoV-2 and identify the presence of variants of concern in a given community or neighborhood, independent of availability and access to individual-level testing. These advantages in combination with its scalability, relatively modest cost and low labor requirements, makes integrating permanent wastewater-based epidemiology infrastructure into public health systems a key component in creating pandemic-resilient cities in the future.

Main

Wastewater-based epidemiology (WBE) has emerged as an effective and sensitive approach for monitoring COVID-19 presence in a community through the detection of the novel coronavirus (SARS-CoV-2) shed by infected individuals into wastewater[1–5]. While methods for COVID-19 WBE are still being refined, particularly with respect to optimizing sampling and virus concentration methods[6–9], this approach has shown promise in monitoring COVID-19 infection trends and detecting community infections prior to reported cases[5,10]. WBE also has clear advantages in terms of cost (compared to traditional surveillance methods) and in areas where clinical testing is limited or residents are hesitant to participate[1–5].

What has remained elusive is the quantitative relationship between viral concentrations in wastewater and community infection rates, as well as the representativeness of community viral genotype profiles from wastewater sequencing. These limitations are due to biological variability of SARS-CoV-2 infections, physical uncertainties of wastewater sampling, and inherent variability in case reporting. Biological variability comes from uncertainty and encompasses latent variation in the magnitude and duration of viral shedding in both symptomatic and asymptomatic individuals[11–13]. Physical uncertainties include representativeness of the

wastewater samples, virus concentration and extraction methodologies, molecular detection methods and inhibition as well as genetic marker persistence and decay rates in sewage conveyance systems[6,8,14–16]. The inherent variability in case reporting results from underreporting of infections due to limited testing capacity, barriers in access to testing, testing avoidance, self-isolation of individuals with mild symptoms, and widespread asymptomatic transmission of the virus[17].

In this paper, we examine SARS-CoV-2 burdens on communities through four lines of evidence: WBE data, clinically reported COVID-19 cases, sequence data obtained through nasal swabs as well as wastewater samples, and prevalence data estimated via nasal swabs of residents in randomly selected census blocks (trace.oregonstate.edu/community). From this examination, **we demonstrate that WBE is a significantly better predictor of community prevalence than reported COVID-19 cases despite its inherent biological and physical variability.** As such, we show that WBE can predict COVID-19 community prevalence (including both symptomatic and asymptomatic individuals), identify COVID-19 hot spots at the neighborhood scale, and estimate the community SARS-CoV-2 variant profile.

Community Prevalence

Over the course of 10 months, from May 2020 through March 2021, 8 COVID-19 prevalence surveys were conducted in 6 communities in Oregon, USA by Oregon State University's TRACE (Team-based Rapid Assessment of Community-level coronavirus Epidemics) project[18] (**Extended Data Table 1**). The selected communities represent a diverse cross-section of Oregon, ranging from a small coastal commercial fishing community (Newport, OR), to mid-sized and large university communities in the temperate Willamette Valley

(Corvallis and Eugene, OR), mid-sized and large arid high desert communities (Redmond and Bend, OR), and a small agricultural community in eastern Oregon (Hermiston, OR).

An average of 517 individuals from 315 households (60% average household participation rate) from 30 randomly selected census blocks were surveyed for the presence of SARS-CoV-2 in nasal swabs for each prevalence study. Nasal swabs were analyzed at the Oregon Veterinary Diagnostic Laboratory using the TaqPath COVID-19 Combo Kit (Applied Biosystems, Foster City, CA). The prevalence of SARS-CoV-2 within each community was calculated using design-weighted estimates accounting for the multi-stage sampling design and imperfect test sensitivity and specificity when at least one case was detected. A Bayesian approach combining observed data with active case counts within the community was used when zero cases were detected. The estimated SARS-CoV-2 prevalence, including both symptomatic and asymptomatic infections, ranged from 8 per 10,000 to 1,687 per 10,000 (**Extended Data Table 2**).

During each prevalence study, composite samples (24-h time-weighted) were taken from the influent of each community's wastewater treatment plant. Samples were filtered through a mixed cellulose ester (HA-type) filter and stabilized with DNA/RNA Shield (Zymo Research, Irvine, CA). RNA was extracted from the filters and analyzed for SARS-CoV-2 RNA using U.S. CDC primers (N1 and N2) using reverse-transcriptase droplet digital PCR (RT-ddPCR). RT-ddPCR was chosen due to its superior sensitivity[19,20], robustness against inhibitors[21], and its wide application in WBE[15,22,23]. The SARS-CoV-2 concentrations in the influent wastewater corresponding to the door-to-door nasal swab samples ranged from 2.92 to 5.13 Log_{10} gene copies per liter (gc/L) (**Extended Data Table 2**).

Additional influent wastewater samples from the same cities were collected 1–3 times per week for 6 to 11 months following the same protocol as described above (**Extended Data Figure 1**). A comparison of the wastewater SARS-CoV-2 concentrations with the Log_{10} of reported cases per 10,000 people (to normalize for differences in population) showed a general moderate positive correlation (Pearson's $r = 0.71$). The accuracy of predicting reported COVID-19 cases using WBE was also moderate (Root Mean Square Logarithmic Error (RMSLE) = 0.14 and Mean Absolute Percentage Error (MAPE) = 0.29), with wastewater concentrations differing by up to 1.8 orders of magnitude representing the same number of reported COVID-19 cases per 10,000 people (**Figure 1A**). Similar correlation strengths and levels of accuracy were observed when each city was analyzed individually (**Extended Data Figure 2**).

Other studies have reported similar results, with WBE providing a relative sense of viral burden in a community and tracking its trend over time[2,24,25]. However, the moderate correlation between wastewater SARS-CoV-2 concentrations and reported COVID-19 cases limits the ability of WBE to predict community infections[26,27]. This moderate association has been attributed to a variety of factors related to the wastewater sample, including virus decay in sewage conveyance systems, variability in sampling and concentrating techniques, and variability in the magnitude and duration of shedding by infected individuals[6,11,14]. However, when the wastewater virus concentrations were compared with prevalence estimates that include both symptomatic and asymptomatic cases (**Figure 1B**), the positive correlation was much stronger (Pearson's $r = 0.96$) and the accuracy was much higher, as demonstrated by the lower RMSLE (0.03) and MAPE (0.17) values. This correlation also suggests that the detection limit of our WBE method is 3 infections per 10,000 people.

The strong correlation between estimated prevalence and wastewater concentrations suggests that the major source of variability in **Figure 1A** may not be due to the wastewater but rather the uncertainty associated with reported COVID-19 cases. To explore this possibility further, reported COVID-19 cases were correlated with estimated prevalence (**Figure 1C**). This correlation was moderately weaker (Pearson's $r = 0.85$) and the accuracy was lower (RMSLE = 0.05 and MAPE = 0.31) than the wastewater versus prevalence correlation. In addition, we found wastewater concentrations to be a significantly better predictor of COVID-19 community prevalence than reported COVID-19 case counts, even after accounting for uncertainty inherent in the wastewater and prevalence estimates by using Monte Carlo simulations (**Extended Data Figure 3**). This demonstrates that it is the uncertainty of reported COVID-19 cases, rather than the wastewater data, that is the most significant source of variability.

Thus, when compared to a metric with much smaller uncertainties, such as the prevalence estimates presented in this study, wastewater data is quantitative and can accurately predict COVID-19 prevalence (**Figure 1B**). This may also explain why wastewater has shown significant predictive capabilities when compared to COVID-19 hospitalizations and deaths[2,5,24]. These metrics also have lower uncertainties than reported COVID-19 cases, which is likely to be impacted by unequal distribution of testing availability, asymptomatic individuals and individuals who do not seek testing for various reasons.

Micro-sewershed Surveillance

In addition to predicting community-wide prevalence with high precision, we have also shown that WBE is a powerful method for detecting infection "hot spots" at the micro-sewershed (*i.e.,* neighborhood) scale. In Newport, OR, USA, a community that experienced a significant

COVID-19 outbreak during our study in mid-June 2020, wastewater SARS-CoV-2 concentrations in micro-sewersheds were significantly correlated with the positivity rate of random swab samples (Pearson's product-moment correlation = 0.64, 95% CI = 0.32 – 0.83, df = 22, $p < 0.001$) and reported cases (Pearson's product-moment correlation = 0.61, 95% CI = 0.39 – 0.76, df = 44, $p < 0.001$). Similarly, the presence of SARS-CoV-2 in wastewater was significantly correlated with reported cases within each micro-sewershed (Fisher's exact test, odds ratio (OR) = 24.3, 95% CI = 3.1 – 341, $p < 0.001$, $n = 33$). The presence/absence of SARS-CoV-2 was also significantly correlated between wastewater and random nasal swabs within each micro-sewershed (Fisher's exact test, OR could not be calculated, 95% CI = 2.16 – ∞, $p = 0.003$, $n = 22$).

The outbreak in Newport involved 185 reported cases linked to a local seafood processing plant in mid-June 2020[28]. During the peak of the outbreak, the TRACE team randomly selected and surveyed 569 individuals across the city using nasal swab assays and estimated a prevalence of 3.4% (**Extended Data Table 2**). Simultaneously, wastewater samples were collected from the wastewater treatment plant influent and 22 pump stations located throughout the community of 10,853 people (**Extended Data Figure 4**). Each pump station drained a micro-sewershed with estimated populations ranging from 15 to 9,426 people (**Extended Data Table 3**). The random nasal swab and wastewater sampling were repeated three weeks later in July 2020. The outcome of this survey indicated a decrease in the community prevalence to an estimated 0.6% (**Extended Data Table 2**).

The decrease in community prevalence corresponded with marked decreases in nasal swab positivity rates, reported COVID-19 cases, and wastewater SARS-CoV-2 concentrations in each micro-sewershed (**Figure 2**). It should be noted that due to the hierarchical flows between

pump stations, the Northside micro-sewershed is receiving flow from the Bayfront micro-sewershed, which contains the Samaritan Pacific Communities Hospital (**Extended Data Figure 4**). Thus, the Northside micro-sewershed wastewater results may be capturing individuals who would not be linked to the Northside micro-sewershed by case reporting or random household surveillance (*e.g.,* infected individuals at the Samaritan Pacific Communities Hospital).

Similar to our observations at the community scale, our WBE data correlated slightly more strongly with nasal swab positivity rates than with reported cases at the neighborhood scale. Thus, WBE may be generally more accurate at identifying hot spots of COVID-19 infections within a community than reported clinical cases. Other studies have used similar approaches to demonstrate the utility of wastewater monitoring at the building scale on college and university campuses[29,30].

Newport Community SARS-CoV-2 Genotype Profile

Using Multi-Locus Sequence Typing[31], two distinct viral variants, a B.1.399 lineage variant designated NA and a B.1 lineage variant designated NB, were detected both in the wastewater and among the individuals who tested positive during the June 2020 Newport survey (**Extended Data Table 4**). The B.1.399/NA consensus most closely matched sequences found in Europe and then in California in March and April, 2020, respectively. The B.1/NB consensus sequence most closely matched that from Yakima County, Washington, on April 29, 2020, and sequences found in Europe in March 2020.

During the June survey (from a total of 569 samples), B.1.399/NA was detected in 77% of the individuals tested positive (10/13) whereas B.1/NB was detected in 23% of the individuals

tested positive (3/13) (**Figure 3 and Extended Data Table 5**). In the July survey, two individuals tested positive (from 550 samples). Although both samples both yielded low coverage sequence data, they were identified as B.1.1.291 by the international database, GISAID (https://www.gisaid.org/) and thus, were different than the individuals from the June survey.

In the wastewater samples, the fraction of viral RNA accounted for by each variant was estimated from the total fraction of variant reads summed across all single nucleotide polymorphism (SNP) positions specific to that variant. During the June survey, the viral variant distribution in the wastewater treatment plant influent was dominated by B.1.399/NA, accounting for 70% of the reads, while B.1/NB accounted for a minority of the reads at 4% (**Figure 3**). This mirrors the observations made through the random nasal swab assays (77% B.1.399/NA and 23% B.1/NB), suggesting that wastewater sequences can both detect the type of viral variants present in a community and provide information on the relative distribution of the SARS-CoV-2 variants within that community.

B.1.399/NA was also dominant in 11 of the 12 positive micro-sewershed wastewater samples collected. During the June survey, B.1.399/NA accounted for 30-98% of the total viral sequence reads across the micro-sewersheds, while during the July survey it accounted for 48-92% of the viral sequence reads across the micro-sewersheds (**Extended Data Table 5**). Additionally, while the abundance of B.1/NB was always in the minority, ranging from 0-52% of total virus sequence reads during the June survey and 0-11% of total virus sequence reads during the July survey, it was always detected in the wastewater of micro-sewersheds where B.1/NB was detected among individuals via random nasal swabs (**Figure 3**).

The only exception to this relationship was the Bayfront micro-sewershed (**Extended Data Figure 4)**. In Bayfront, B.1/NB was not detected in the wastewater, while the single

positive individual discovered in Bayfront carried B.1/NB. However, the Bayfront micro-sewershed houses the Samaritan Pacific Communities Hospital, which likely was releasing wastewater with a high concentration of B.1.399/NA from patients.

Finally, from the weeks of June 8 to November 30, 2020, five SARS-CoV-2 variants were detected consistently in wastewater samples from the Newport wastewater treatment plant (**Figure 4A**), representing at least 5% of the sequence reads in samples from at least three weeks. Variant B.1.399/NA was the dominant variant during the weeks of June 8 to July 6, with lesser amounts of B.1/NB detected. During the weeks of July 20 to 27, a B.1.369 sub-variant, designated EE (**Extended Data Table 4B**), was most abundant, and was also detected during the weeks of June 29 and August 3. From August 3 onwards, a B.1.2 sub-variant, designated FF (**Extended Data Table 4B**), was the most abundant variant detected. The detection of these variants in Newport broadly mirrored trends across the state of Oregon observed in individuals whose positive samples were sequenced and deposited in GISAID (**Figure 4B**). Interestingly, spikes observed in the wastewater treatment plant influent SARS-CoV-2 concentrations often corresponded to the appearance of a new dominant variant in the wastewater sequences (**Figure 4A**). This suggests that rises in viral RNA concentrations in WBE samples may signal the appearance of a new dominant variant. To our knowledge this is the first study of variant abundance in wastewater at the neighborhood scale, though others have sequenced SARS-CoV-2 in wastewater at the community scale[32–34].

Conclusion

We have demonstrated that WBE is a robust method that provides accurate quantitative predictions of COVID-19 prevalence in communities that are diverse in terms of size, location,

access to testing, and pandemic stage. This finding greatly expands the potential of WBE beyond its current qualitative uses to track SARS-CoV-2 concentration trends in wastewater. We also have shown that WBE is a more reliable indicator of COVID-19 prevalence than reported cases due the inherent variability and biases that are associated with reported cases. Additionally, we demonstrated that WBE can detect neighborhood-level COVID-19 hot spots, characterize the viral genetic diversity, and assess the importation of new cases through viral variant identification over time. Thus, we have shown that WBE is a powerful approach to accurately estimate COVID-19 prevalence and survey its genetic diversity. As such, WBE is clearly an essential tool in the efforts to bring the current pandemic to an end, and the establishment of national wastewater surveillance networks worldwide offers hope for improved surveillance of future epidemics caused by other pathogens.

Main Text References

1. Bivins, A. *et al.* Wastewater-Based Epidemiology: Global collaborative to maximize contributions in the fight against COVID-19. *Environmental Science and Technology* vol. 54 7754–7757 (2020).
2. Medema, G., Heijnen, L., Elsinga, G., Italiaander, R. & Brouwer, A. Presence of SARS-Coronavirus-2 RNA in sewage and correlation with reported COVID-19 prevalence in the early stage of the epidemic in the Netherlands. *Environ. Sci. Technol. Lett.* **7**, 511–516 (2020).
3. Ahmed, W. *et al.* First confirmed detection of SARS-CoV-2 in untreated wastewater in Australia: A proof of concept for the wastewater surveillance of COVID-19 in the community. *Sci. Total Environ.* **728**, 138764 (2020).
4. La Rosa, G. *et al.* First detection of SARS-CoV-2 in untreated wastewaters in Italy. *Sci. Total Environ.* **736**, 139652 (2020).
5. Wurtzer, S. *et al.* Evaluation of lockdown effect on SARS-CoV-2 dynamics through viral genome quantification in waste water, Greater Paris, France, 5 March to 23 April 2020. *Eurosurveillance* **25**, (2020).
6. Curtis, K., Keeling, D., Yetka, K., Larson, A. & Gonzalez, R. Wastewater SARS-CoV-2 concentration and loading variability from grab and 24-1 hour composite samples.

medRxiv 2020.07.10.20150607 (2020) doi:10.1101/2020.07.10.20150607.

7. Ahmed, W. *et al.* Comparison of virus concentration methods for the RT-qPCR-based recovery of murine hepatitis virus, a surrogate for SARS-CoV-2 from untreated wastewater. *Sci. Total Environ.* 139960 (2020) doi:10.1016/j.scitotenv.2020.139960.

8. LaTurner, Z. W. *et al.* Evaluating recovery, cost, and throughput of different concentration methods for SARS-CoV-2 wastewater-based epidemiology. *Water Res.* **197**, 117043 (2021).

9. Whitney, O. N. *et al.* Sewage, Salt, Silica and SARS-CoV-2 (4S): An economical kit-free method for direct capture of SARS-CoV-2 RNA from wastewater. *medRxiv* 2020.12.01.20242131 (2020) doi:10.1101/2020.12.01.20242131.

10. Randazzo, W. *et al.* SARS-CoV-2 RNA in wastewater anticipated COVID-19 occurrence in a low prevalence area. *Water Res.* **181**, 115942 (2020).

11. Jones, D. L. *et al.* Shedding of SARS-CoV-2 in feces and urine and its potential role in person-to-person transmission and the environment-based spread of COVID-19. *Sci. Total Environ.* **749**, 141364 (2020).

12. Cheung, K. S. *et al.* Gastrointestinal manifestations of SARS-CoV-2 infection and virus load in fecal samples from a Hong Kong cohort: Systematic review and meta-analysis. *Gastroenterology* **159**, 81–95 (2020).

13. Chen, Y. *et al.* The presence of SARS-CoV-2 RNA in the feces of COVID-19 patients. *J. Med. Virol.* **92**, 833–840 (2020).

14. Bivins, A. *et al.* Persistence of SARS-CoV-2 in water and wastewater. *Environ. Sci. Technol. Lett.* **7**, 937–942 (2020).

15. Bivins, A. *et al.* Within-day variability of SARS-CoV-2 RNA in municipal wastewater influent during periods of varying COVID-19 prevalence and positivity. *medRxiv* 2021.03.16.21253652 (2021) doi:10.1101/2021.03.16.21253652.

16. Pecson, B. M. *et al.* Reproducibility and sensitivity of 36 methods to quantify the SARS-CoV-2 genetic signal in raw wastewater: Findings from an interlaboratory methods evaluation in the U.S. *Environ. Sci. Water Res. Technol.* **7**, 504–520 (2021).

17. Gandhi, M., Yokoe, D. S. & Havlir, D. V. Asymptomatic transmission, the Achilles' heel of current strategies to control Covid-19. *N. Engl. J. Med.* **382**, 2158–2160 (2020).

18. Oregon State University TRACE COVID-19. trace.oregonstate.edu (2021).

19. Suo, T. *et al.* ddPCR: a more accurate tool for SARS-CoV-2 detection in low viral load specimens. *Emerg. Microbes Infect.* **9**, 1259–1268 (2020).

20. Dong, L. *et al.* Highly accurate and sensitive diagnostic detection of SARS-CoV-2 by digital PCR. *Talanta* **224**, 121726 (2021).

21. Cao, Y., Raith, M. R. & Griffith, J. F. Droplet digital PCR for simultaneous quantification of general and human-associated fecal indicators for water quality assessment. *Water Res.* **70**, 337–349 (2015).

22. Gonzalez, R. *et al.* COVID-19 surveillance in Southeastern Virginia using wastewater-based epidemiology. *Water Res.* **186**, 116296 (2020).

23. D'Aoust, P. M. *et al.* Quantitative analysis of SARS-CoV-2 RNA from wastewater solids in communities with low COVID-19 incidence and prevalence. *Water Res.* **188**, (2021).

24. Peccia, J. *et al.* Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nat. Biotechnol.* **38**, 1164–1167 (2020).

25. Graham, K. E. *et al.* SARS-CoV-2 RNA in wastewater settled solids is associated with COVID-19 cases in a large urban sewershed. *Environ. Sci. Technol.* **55**, 488–498 (2021).

26. Cao, Y. & Francis, R. On forecasting the community-level COVID-19 cases from the concentration of SARS-CoV-2 in wastewater. *Sci. Total Environ.* **786**, (2021).

27. Greenwald, H. D. *et al.* Interpretation of temporal and spatial trends of SARS-CoV-2 RNA in San Francisco Bay Area wastewater. *medRxiv* 2021.05.04.21256418 (2021) doi:10.1101/2021.05.04.21256418.

28. Oregon Health Authority COVID-19 weekly Report: Oregon's Weekly Surveillance Summary of the Nonvel Coronavirus (COVID-19). *COVID-19 weekly Report: Oregon's Weekly Surveillance Summary of the Nonvel Coronavirus (COVID-19)* https://www.oregon.gov/oha/PH/DISEASESCONDITIONS/DISEASESAZ/Emerging Respitory Infections/COVID-19-Weekly-Report-2020-08-05-FINAL.pdf (2020).

29. Spurbeck, R. R., Minard-Smith, A. & Catlin, L. Feasibility of neighborhood and building scale wastewater-based genomic epidemiology for pathogen surveillance. *Sci. Total Environ.* **789**, 147829 (2021).

30. Harris-Lovett, S. *et al.* Wastewater surveillance for SARS-CoV-2 on college campuses: Initial efforts, lessons learned and research needs. *Int. J. Environ. Res. Public Health* **18**, 4455 (2021).

31. Ibarz Pavón, A. B. & Maiden, M. C. J. Multilocus sequence typing. *Methods Mol. Biol.* **551**, 129–140 (2009).

32. Crits-Christoph, A. *et al.* Genome sequencing of sewage detects regionally prevalent SARS-CoV-2 variants. *MBio* **12**, 1–9 (2021).

33. Izquierdo-Lara, R. *et al.* Monitoring SARS-CoV-2 circulation and diversity through community wastewater sequencing, the Netherlands and Belgium. *Emerg. Infect. Dis.* **27**, 1405–1415 (2021).

34. Agrawal, S., Orschler, L. & Lackner, S. Metatranscriptomic analysis reveals SARS-CoV-2 mutations in wastewater of the Frankfurt Metropolitan Area in Southern Germany. *Microbiol. Resour. Announc.* **10**, (2021).

Tables and Figures

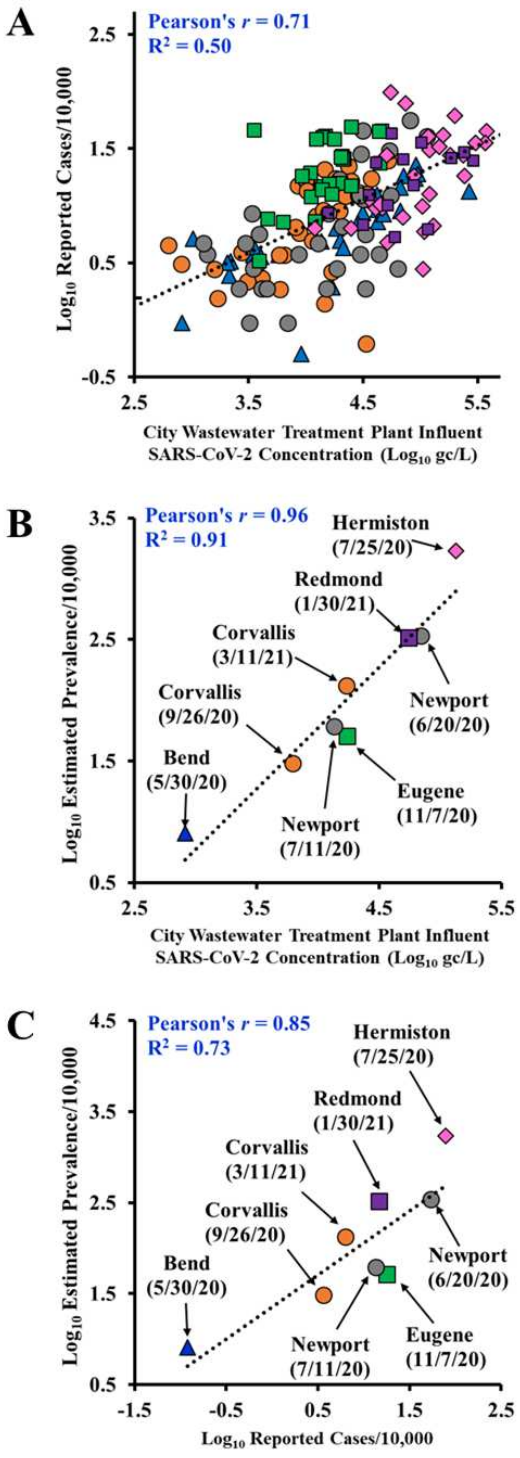


Figure 1. Wastewater Concentrations vs Reported COVID 19 Cases or Estimated Prevalence. (A) Log_{10} of wastewater SARS-CoV-2 concentrations versus the Log_{10} of weekly reported COVID-19 cases (reported by ZIP code) **(B)** Log_{10} of wastewater SARS-CoV-2 concentrations versus the Log_{10} of estimated prevalence and **(C)** Log_{10} of weekly reported COVID-19 cases (reported by ZIP code) versus the Log_{10} of estimated prevalence for Bend (▲), Corvallis (●), Eugene (■), Hermiston (♦), Newport (●) and Redmond (■), OR.

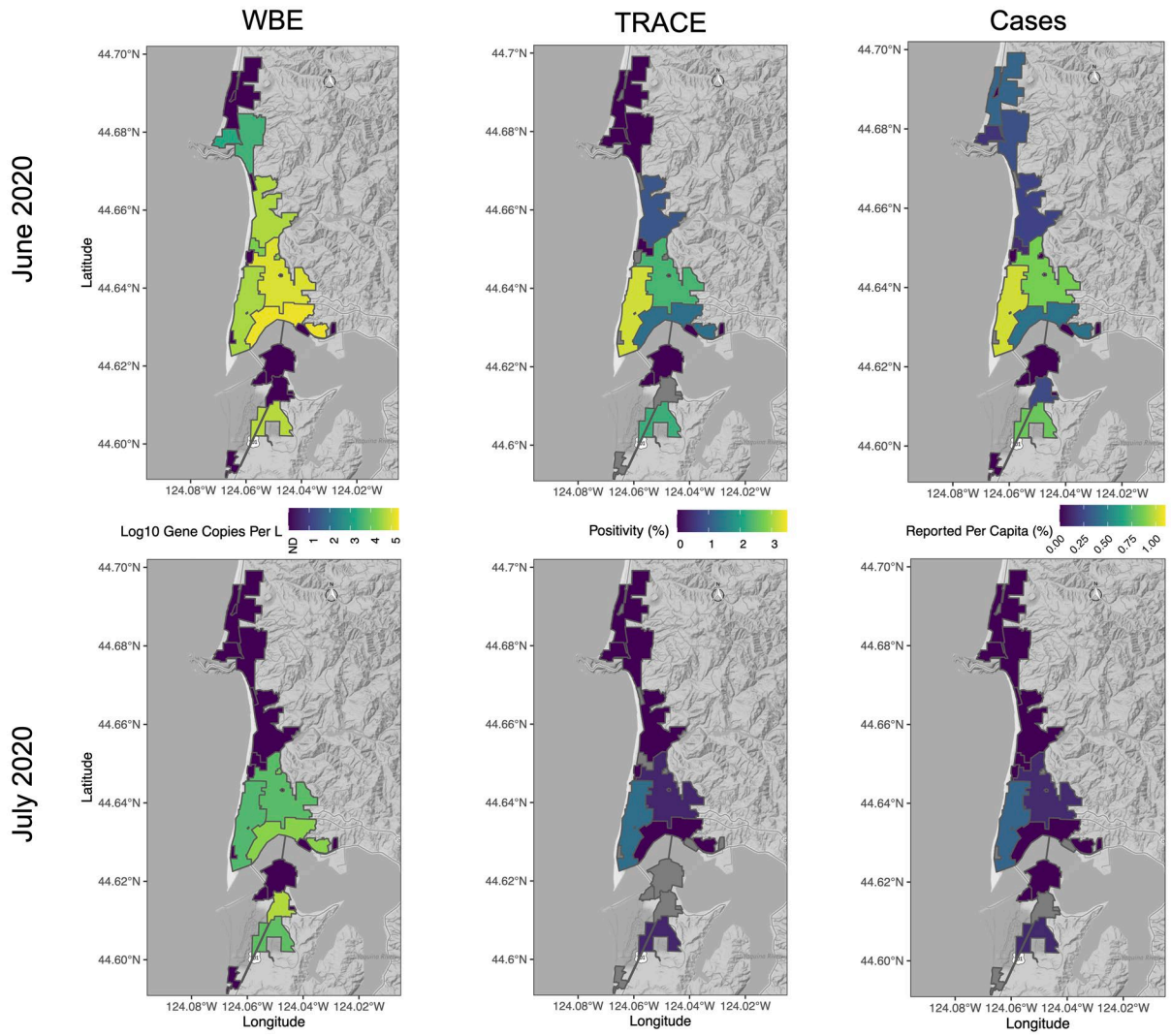


Figure 2. COVID-19 Burden Heat Maps. Wastewater SARS-CoV-2 concentrations (left column), % positivity from community surveys (middle column) and % reported cases per capita (right column) of the 22 micro-sewersheds sampled in Newport, OR. Wastewater and community surveys were conducted during June 18-19, 2020 (top row) and July 8-9, 2020 (bottom row); reported case windows were the 10 days prior to and including the sampling periods.

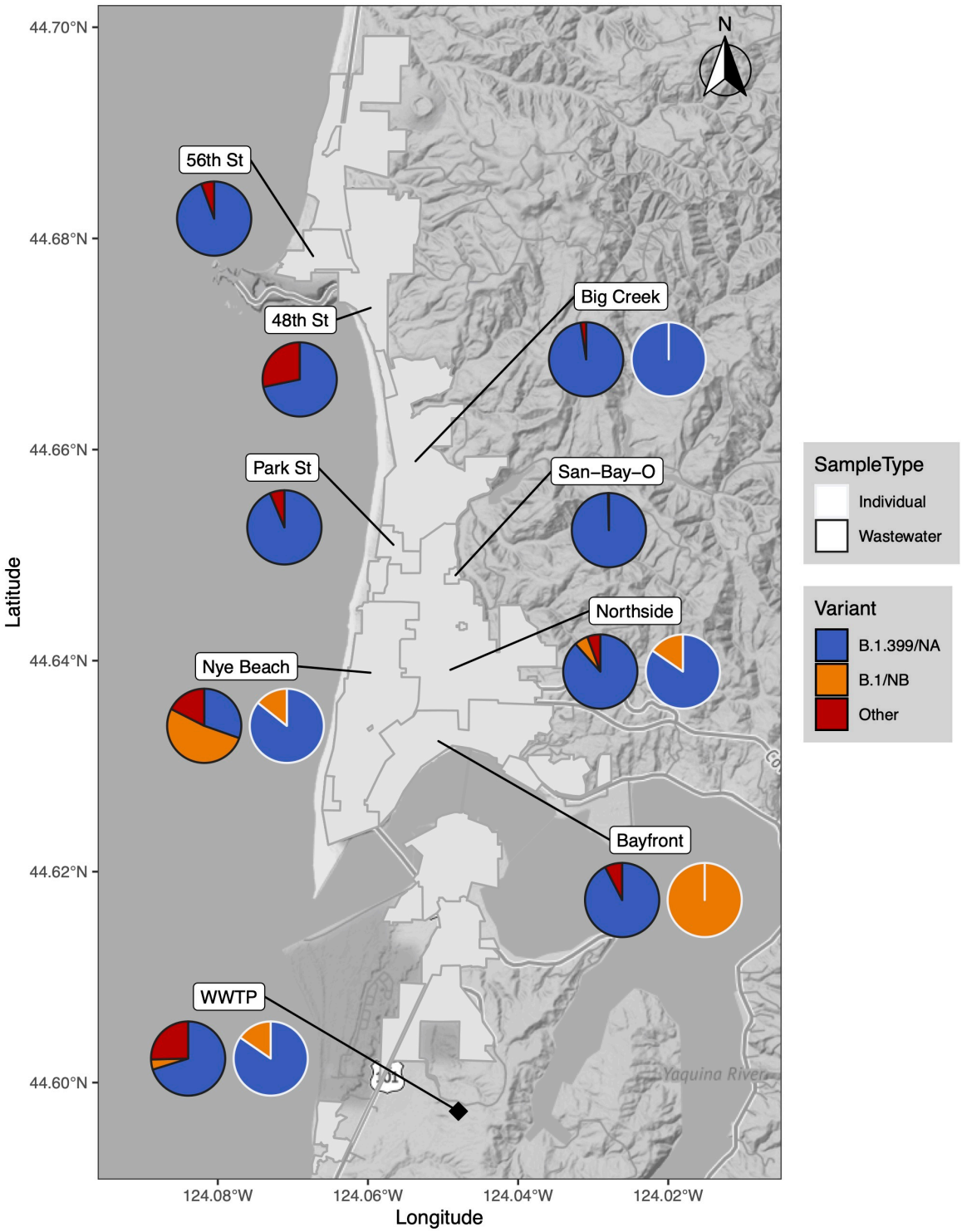


Figure 3. Spatial Distribution of SARS-CoV-2 Variants. The percent sequence reads of B.1.399/NA (A) and B.1/NB (B) located within the various Newport, OR micro-sewershed boundaries during the June 18-19, 2020 surveillance study. Sequences were obtained from both micro-sewershed wastewater as well as random household community surveys.

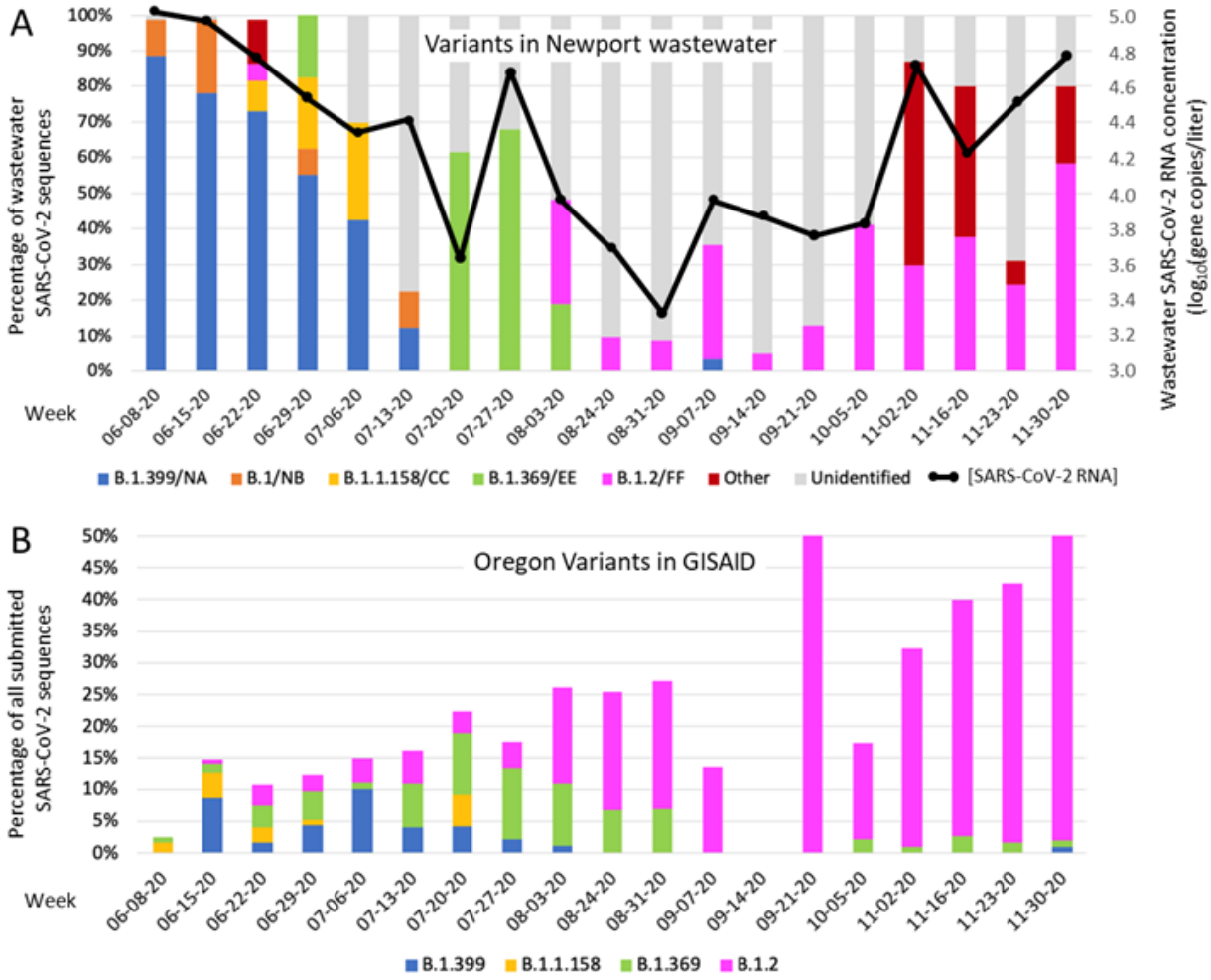


Figure 4. SARS-CoV-2 Variant Temporal Distribution. (**A**) The average estimated percent viral sequence reads of the indicated SARS-CoV-2 variant RNAs and the Log_{10} SARS-CoV-2 concentrations quantified in the Vance Avery Wastewater Treatment Plant influent (Newport, OR) from June 10 to December 2, 2020. Each data point represents the mean of measurements collected during the week beginning on the date shown. Sequence data from some dates derived from a single measurement, and errors are expected to be comparable to those shown in Extended Data Table 5. (**B**) Percentage of all variants of the indicated lineages among SARS-CoV-2 sequences from samples collected in Oregon during the week beginning on the indicated date and deposited in GISAID.

Figure Legends

Figure 1. Wastewater Concentrations vs Reported COVID 19 Cases or Estimated Prevalence. (A) Log_{10} of wastewater SARS-CoV-2 concentrations versus the Log_{10} of weekly reported COVID-19 cases (reported by ZIP code) **(B)** Log_{10} of wastewater SARS-CoV-2 concentrations versus the Log_{10} of estimated prevalence and **(C)** Log_{10} of weekly reported COVID-19 cases (reported by ZIP code) versus the Log_{10} of estimated prevalence for Bend (▲), Corvallis (●), Eugene (■), Hermiston (♦), Newport (●) and Redmond (■), OR.

Figure 2. COVID-19 Burden Heat Maps. Wastewater SARS-CoV-2 concentration (left column), % positivity from community surveys (middle column) and % reported cases per capita (right column) of the 22 micro-sewersheds sampled in Newport, OR. Wastewater and community surveys were conducted during June 17-20, 2020 (top row) and July 8-9, 2020 (bottom row); reported case windows were the 10 days prior to and including the sampling periods.

Figure 3. Spatial Distribution of SARS-CoV-2 Variants. The percent sequence reads of B.1.399/NA and B.1/NB located within the various Newport, OR micro-sewershed boundaries during the June 17-20, 2020 surveillance study. Sequences were obtained from both micro-sewershed wastewater and random household community surveys.

Figure 4. SARS-CoV-2 Variant Temporal Distribution. (**A**) The average estimated percent viral sequence reads of the indicated SARS-CoV-2 variants and the Log_{10} SARS-CoV-2 concentrations quantified in the Vance Avery Wastewater Treatment Plant influent (Newport, OR) from June 10 to December 2, 2020. Each data point represents the mean of measurements collected during the week beginning on the date shown. Error bars on the RNA concentrations represent the standard error of quadruplicate RT-ddPCR reactions from duplicate samples. Sequence data from some dates derived from a single measurement, and errors are expected to be comparable to those shown in Extended Data Table 5. Other indicates % reads corresponding to known variants other than those shown. Unidentified indicates % reads not attributable to any known variants (**B**) Percentage of all variants of the indicated lineages among SARS-CoV-2 sequences from samples collected in Oregon during the week beginning on the indicated date and deposited in GISAID.

Methods

Community swab sampling

In total, TRACE collected 4,136 nasal swab samples from residents in 2,521 randomly selected households in Bend, Corvallis, Eugene, Hermiston, Newport and Redmond, Oregon from May 30, 2020 to March 14, 2021. The response rates ranged from 38-71%, with an average response rate of 60% (**Extended Data Table 2**). The methods used for the random household sampling and prevalence estimate models are presented at the TRACE team website (trace.oregonstate.edu)[18].

Wastewater treatment plant sampling

All wastewater treatment plant (WWTP) influent samples were comprised of 24-h time-weighted composites taken prior to primary treatment. The characteristics of each wastewater treatment plant is given in **Extended Data Table 1**. Twenty-two pump stations serving Newport, OR's Vance Avery WWTP sewershed were sampled hourly for 24 h from June 20-21 and July 11-12, 2020. Some pump stations were sampled twice within a single weekend, for a total of 52 pump station samples. The characteristics of each micro-sewershed (*i.e.*, the area served by each pump station) and the distribution of samples are given in **Extended Data Table 3**.

For both WWTPs and pump stations, the 24-h composites consisted of hourly samples and were kept on ice during sampling. Samples collected on or before July 31, 2020 were frozen in 200 mL aliquots and stored for up to 33 days prior to concentration (median 6 days). After thawing in cold water baths, samples were concentrated using electronegative filtration, as previously described[35]. Briefly, all samples collected on or before July 10, 2020 were acidified to

461 a final pH of 3.5 and magnesium chloride was added to a final concentration of 25 mM. Samples
462 (30-40 mL) were vacuum filtered through a 0.45 μm pore size, 47 mm diameter mixed cellulose
463 ester electronegative filter (HAWP, Millipore, Bedford, MA, USA).

464 Influent samples collected after July 10, 2020 were filtered with no amendments as
465 preliminary data showed no statistical difference in viral recovery with unamended wastewater
466 (data not shown). Additionally, influent samples collected after July 31, 2020 were neither frozen
467 nor amended prior to filtration. In these samples, filtration occurred within 8 h of sample
468 collection.

469 Once filtration was complete, the electronegative membranes were placed into 2 mL
470 tubes containing a mix of 0.7 mm garnet and 0.5 mm glass beads, stabilized in 1 mL of
471 DNA/RNA Shield (Zymo Research, Irvine, CA, USA), and frozen until analysis. Field blanks of
472 deionized water were processed with every batch of samples.
473
474 ***Molecular analysis: Nasal swabs***
475 Participant nasal swab samples were analyzed using the TaqPath COVID-19 Combo Kit
476 (ThermoFisher), in accordance with the Instructions for Use as required by the Emergency Use
477 Authorization under strict biosafety level 2 (BSL2) conditions. Nucleic acid isolation was
478 performed using the MagMax Viral/Pathogen II Nucleic Acid Isolation Kit (Applied
479 Biosystems/ThermoFisher, USA). Briefly, 200 μL of transport medium from the swab sample
480 was added to a single well of a KingFisher Deepwell 96-well plate containing 5 μL Proteinase K.
481 Each 96-well plate held 94 participant samples and one negative control well containing water.
482 The last well was left empty to allow for the positive control to be added during the RT-PCR
483 detection step. After sample addition, nucleic acid magnetic beads were resuspended, 10 μL was

added to 265 µL binding solution and then added to the wells. MS2 phage control (5 µL) was added to all wells as an extraction control and processed on a KingFisher Flex magnetic particle processor. Purified nucleic acids were eluted in 50 µL MagMax elution solution. Eluted nucleic acid was stored at -80 ºC unless RT-PCR was run within 2 h of extraction.

Detection of SARS-CoV-2 viral RNA was performed using the TaqPath RT-PCR COVID-19 Kit on a 7500 Fast Real-Time PCR Instrument (Applied Biosystems). Reactions were run in multiplex with primers and probes specific for three gene sequences specific to SARS-CoV-2: ORF1ab, N Protein, and S Protein. A primer and probe set was included to detect the MS2 phage added during the initial sample processing as an internal control to verify RNA extraction. The thermal protocol for the one-step real-time PCR was as follows: 2 min at 25 ºC UNG incubation, 10 min at 53 ºC reverse transcription, 2 min at 95 ºC activation, followed by 40 cycles of 3 sec at 95 ºC denaturation and 30 sec at 60 ºC anneal/extension/detection. Results were analyzed using the SDS Software v1.5.1 and interpreted using the COVID-19 Interpretive Software (Applied Biosystems, v1.2).

Molecular analysis: Wastewater

Wastewater samples were homogenized with either 0.5 mm glass or 0.7 mm garnet beads in DNA/RNA Shield using either a Qiagen TissueLyser (Qiagen Inc, Germantown, MD) or BioSpec Mini-Beadbeater 16 (BioSpec Products, Inc, Bartlesville, OK) for 2 min. (Preliminary data indicated no difference in recovery between these homogenization methods; data not shown.) Beads and debris were pelleted by centrifugation at 12,000 rcf for 1 min. The lysate was transferred from each tube to a 96-well plate, and 200-400 µL was extracted using the MagMAX Viral/Pathogen kit on a KingFisher automated instrument (ThermoFisher Scientific, Waltham,

MA) as described above. Purified RNA was eluted in 50 µL MagMAX Elution Solution. Extraction recovery from RNA extraction step was quantified with a commercial standard (Exact Diagnostics, Fort Worth, TX), and extraction blanks were included with every run. RNA was stored at -80 ºC until analysis.

Two SARS-CoV-2 RNA targets (N1/N2) and an internal control (Human RNase P) were measured via RT-ddPCR using a commercial triplex assay (2019-nCoV CDC ddPCR Triplex Probe Assay, Bio-Rad catalog no. 12008202) using the One-Step RT-ddPCR Advanced Kit for Probes and run on a QX-200 ddPCR system (Bio-Rad, Hercules, CA). The primer and probe sequences were published previously[36]. Reactions were partitioned into droplets using an automated droplet generator (ADG). Twenty-two microliter reactions were prepared with 5.5 µL template RNA, while the ADG only partitioned 20 µL, yielding an effective template volume of 5 µL. Each reaction had an average of 12,657 droplets (Std Dev = 1,783). Commercially prepared RNA standards and negative controls were included on each extraction plate and ddPCR plate (cat. no. COV019 and COV000, Exact Diagnostics, Fort Worth, TX). All samples and controls were analyzed in duplicate.

The one-step thermal cycling conditions were as follows: reverse transcription at 50 ºC for 60 min; enzyme activation at 95 ºC for 10 min; 40 cycles of denaturation at 94 ºC for 30 s followed by annealing/extension at 55 ºC for 60 s; enzyme inactivation at 98 ºC for 10 min; and lastly a 4 ºC hold for droplet stabilization, for a minimum of 30 min to a maximum of overnight. Finally, the amplification in the droplets was determined using the Bio-Rad droplet reader. All assay conditions were performed as specified in the Bio-Rad assay protocol[37].

RT-ddPCR data analysis

Wells with less than 6000 droplets were omitted. Sample data with positive reactions were accepted only if the corresponding extraction blank and field blanks as well as the PCR negative and no-template controls were all negative for the N1/N2 targets. When averaging sample data across replicates, a value of ½ the sample-specific limit of detection was substituted for non-detects. Reactions were regarded as positive if three or more droplets per well amplified in either target. Droplet clusters were manually called for each target using the QuantaSoft Analysis Pro software (Bio-Rad, Hercules, CA). All other analyses were conducted in R, version 4.1.0[38] with Rstudio Desktop, version 1.4.1717[39]. Spatial graphics were created using the sf[40], ggplot2[41], and ggmap[42] packages.

The N1 and N2 markers exhibited excellent agreement in the wastewater samples. The markers were concordant in 91.4% of reactions: 4.1% of reactions (n = 954) had a positive detection in N1 only (using a threshold of three positive droplets per reaction), and 4.5% of reactions were positive in N2 only. Quantitatively, the markers were also well aligned according to a simple linear model (slope = 0.94 with N2 as the "response" variable, adjusted $r^2 = 0.97$, RMSE = 0.58 copies per reaction). Accordingly, the N1 and N2 data were averaged together for all reported concentration values.

For limit of blank (LOB) determination, 104 reactions were run with the triplex assay across 3 plates using the Exact Diagnostics Negative Control as a template. Due to the non-normal distribution of the results, the LOB was determined using a non-parametric (rank order) method with a false positive probability (α) of 0.05. The LOB for N1 was 2.0 copies per reaction, and the LOB for N2 was 4.2 copies per reaction. Further details regarding the LOB results can be found in Extended Data and Supplementary Materials.

The limit of detection (LOD) was predicted for each target (N1 and N2) using Equation 1.

Equation 1: $LOD_{predicted} = LOB + 2 * stDev(CopiesPerReaction_{LOB})$

Where stDev($CopiesPerReaction_{LOB}$) is the standard deviation of the copies per reaction from the LOB assay. The predicted LOD was subsequently tested by running 60 test reactions at concentrations of 4 and 12 copies per reaction of target using the Exact Diagnostics Standard for SARS-CoV-2. The LOD of N1 and N2 was estimated to be 8 and 12 copies per reaction, respectively. Further details regarding the LOD results can be found in Extended Data and Supplementary Materials.

cDNA library preparation and sequencing

For cDNA synthesis, 11 µL of lysate from positive wastewater or participant samples, were used for single-strand cDNA synthesis using the Thermo Superscript IV kit with the following modifications: no host gDNA/RNA removal steps are performed, no RNase H step is performed, and the reverse transcriptase incubation step at 50 ºC is increased from 10 min to 30 min.
The cDNA was used for amplification and sequencing using the Swift Amplicon® SARS-CoV-2 Panel (AL-COV48) together with Swift Amplicon® Combinatorial Dual Indexed Adapters (AL-S1A96, AL-S1BA96, AL-S2A96, and AL-S2B96). The Swift Biosciences protocol was followed, except that, after optimization experiments, the volume of the G1 reagent was reduced to 25% of recommended level, resulting in an increase in read coverage for wastewater samples of 2- to 8-fold for the experiments described in this paper (with the release of version 2 of the Swift primer set, we no longer reduce this reagent). Individual libraries (30-96 samples) were quantified by fluorescence, normalized, and pooled for 2 x 150 bp sequencing on a lane of an

Illumina HiSeq3000 sequencer. Except for some initial sequencing experiments, libraries from individuals were prepared on different days and sequenced in different lanes than samples from wastewater, to reduce the possibility of contamination of sequences from low titer wastewater samples with sequences from high titer individual samples. Pooled libraries from wastewater samples were often run on 2 to 3 lanes of the Hi-Seq 3000 to increase read depth. Wastewater sequencing on the HiSeq3000 produced a median percentage of reads mapped of 0.86% compared to a median of 21.7% for nasal swab samples. In general, wastewater samples with an RNA concentration of $Log_{10} > 4.0$ gc/L reliably produced usable amounts of sequence. Samples with lower RNA concentrations produced variable amounts of sequence, and those below Log_{10} 3.0 gc/L were routinely unsuccessful.

Multi-Locus Sequence Typing

After demultiplexing, Illumina primer sequences were trimmed using BBDuk, sequences were aligned to the reference sequence (Wuhan-Hu-1; NC_045512.2) using BWA mem, then SARS-Cov-2 primers were removed using Swift Biosciences *Primerclip*™ package. GATK then was used identify variants compared to the reference sequence and count the numbers of reference and variants reads at each SNP site; ploidy was set to 4 and down-sampling was not employed. When required, Integrated Genome Viewer (IGV) was used to manually inspect sequence alignments and variant calls.

In order to identify viral genotypes represented in wastewater RNA by multi-locus sequencing typing, sequences from individual samples from Oregon were used to define multi-locus genotypes as a set of polymorphic sites that were unique to each variant and were not shared with any other variant known at the time (**Extended Data Table 4**).

To estimate the fraction of each variant present in the wastewater RNA sequences, the number of reads supporting each variant were summed across all variant-specific SNP sites, then divided by the total number of reads spanning those SNP sites. At SNP sites where the total number of reads was greater than 100, both the variant read number and total read number were scaled down to a total read number of 100 prior to the estimation calculation. At SNP sites with less than 100 total reads, the actual number of reads was used, thus decreasing the weight of those counts proportionally to the read coverage. After this estimation was conducted for each variant, the fractions were summed for all identified variants. If the sum was greater than 0.7, it was assumed that the identified variants comprised all of the RNA molecules present and each fraction was divided by the total to normalize the total to 1.0. If the initial sum was less than 0.7, it was assumed that the difference was comprised of RNA molecules from unidentified variants, and the fraction of RNA attributable to the unidentified variants was set to the difference between the total and 1.0. If the sum was greater than 1.5, the SNP data were manually reviewed to identify and remove artifacts such as SNPs miscalled by GATK (commonly 1 bp indels) or SNPs that were unexpectedly shared between the variants present and hence double-counted. The custom software to conduct these calculations was implemented in *R* and packaged into a *SnakeMake* pipeline[43].

All individual sequences were deposited in GISAID (see **Extended Data Table 6** for accession numbers) and all wastewater sequences were deposited in NCBI's short read archive, under BioProject PRJNA719837 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA719837)

Reported Cases

620 Weekly COVID-19 case data for the entire community was obtained for zip codes 97365 and
621 97366 from the Oregon Health Authority website[28]. For weeks when < 10 cases occurred in a
622 given zip code, they were reported as 1-9 cases, and a value of 5 cases was substituted in the
623 calculations. Reported test results with location data were obtained with permission from the
624 Lincoln County Health and Human Services Department, and public health officials anonymized
625 the data by aggregating according to micro-sewershed prior to analysis.

626

627 ***Monte Carlo Simulations***

628 To compare WBE and reported COVID-19 cases as predictors of community prevalence, Monte
629 Carlo simulations were performed to account for the uncertainty in the point estimates for each
630 sampling event. For each simulation, a new wastewater concentration (Log_{10} gc/L) was re-drawn
631 for each community sample from a Gaussian distribution with mean equal to the point estimate
632 and standard deviation equal to the standard error of the point estimate. Similarly, a new
633 prevalence was re-drawn for each community from a method-specific distribution: a truncated
634 Gaussian distribution with mean equal to the point estimate, standard deviation equal to the
635 standard error of the point estimate, and lower truncation bound equal to zero when the design-
636 based estimator was used; or a Beta distribution with shape and scale parameters estimated using
637 optimization to fit the posterior mean and 95% credible interval when the Bayesian estimator
638 was used[44]. The reported COVID-19 case numbers are not re-drawn. For each simulation, a
639 simple linear regression model is fit using the new wastewater concentration draw to predict the
640 Log_{10} of the new community prevalence draw, and a separate model is fit using the Log_{10} of the
641 observed COVID-19 case count to predict the Log_{10} of the new community prevalence draw. For
642 each model, the slope, intercept, and R^2 value are recorded.

10,000 simulations were performed, and the results are shown in **Extended Data Figure 3**. The horizontal lines about the observed points indicate ±1 SE. The vertical lines indicate ±1 SE for the design-based estimates and the 68% credible interval for the Bayesian estimates. Each of the middle 95% (based on slope) of the 10,000 different regression lines are shown in pale gray with high transparency, so the darker band indicates more lines. This gives a sense of the uncertainty of the relationships between the different variables.

The median R^2 for the WBE model is 0.82 compared with 0.71 for the reported cases model. A Wilcoxon rank-sum test for whether the two distributions for R^2 have a location shift of 0 versus the alternative that the WBE R^2 distribution has a positive shift gives a *p*-value of <0.0001. Thus, even after accounting for the uncertainty inherent in the WBE and community prevalence estimates, the difference between the WBE and reported cases median R^2 values is significant, and wastewater concentration has a larger median R^2. For each simulation, the two non-nested regression models can also be directly compared using Vuong's test, which is a likelihood-ratio-based test for model selection using the Kullback–Leibler information criterion[45,46]. Vuong's test finds evidence that the WBE model fits significantly better than the reported cases model ($p < 0.05$) 53% of the time. The reported cases model only fits significantly better 2% of the time; the models cannot be distinguished on the remaining occasions. Although Vuong's test cannot determine if the preferred model is the true model, the high R^2 value (0.91 using the observed data) and low RMSLE (0.03) and MAPE (0.17) are indicative of a good fit. Taken together, these two tests suggest that WBE is a significantly better predictor of community prevalence than reported COVID-19 case counts.

Methods References

35. Steele, J. A., Blackwood, A. D., Griffith, J. F., Noble, R. T. & Schiff, K. C. Quantification of pathogens and markers of fecal contamination during storm events along popular surfing beaches in San Diego, California. *Water Res.* **136**, 137–149 (2018).

36. Lu, X. *et al.* US CDC real-time reverse transcription PCR panel for detection of severe acute respiratory syndrome Coronavirus 2. *Emerg. Infect. Dis.* **26**, 1654–1665 (2020).

37. *Bio-Rad SARS-CoV-2 ddPCR Kit: Instructions for Use*. (2020).

38. R Code Team, R: A Language and Environment for Statistical Computing (R Foundation for Statitistical Computing). (2020).

39. RStudio Team, RStudio: Integrated Development Environment for R (RStudio, PBC). (2020).

40. Pebesma, E. Simple features for R: Standardized support for spatial vector data. *The R Journal* **10**, 439–446 (2018).

41. Wickham, H. *ggplot2: Elegant graphics for data analysis*. (Springer-Verlag, 2016).

42. Kahle, D. & Wickham, H. ggmap: Spatial visualization with ggplot2. *The R Jounral* **5**, 144–161 (2013).

43. Köster, J. *et al.* Sustainable data analysis with Snakemake. *F1000Research* **10**, 33 (2021).

44. Devleesschauwer, B. *et al.* prevalence: Tools for prevalence assessment studies. R package version 0.4.0. http://cran.r-project.org/package=prevalence.

45. Vuong, Q. H. Likelihood Ratio Tests for model selection and non-nested hypotheses. *Econometrica* **57**, 307 (1989).

46. Merkle, E. & You, D. nonnest2: Tests of non-nested models. R package version 0.5-5. (2020).

47. Decaro, N. *et al.* Detection of bovine coronavirus using a TaqMan-based real-time RT-PCR assay. *J. Virol. Methods* **151**, 167–171 (2008).

Acknowledgements

The authors are grateful to the Public Works Departments of the cities of Bend, Corvallis, Eugene, Hermiston, Newport and Redmond, the Benton County Health Department, the Deschutes County Health Services, the Lane County Public Health Department, the Lincoln

698 County Health and Human Services Department, the Umatilla County Public Health Department
699 and the Oregon Health Authority for their support in sample collection and public health data
700 acquisition. We also thank the staff of Clean Water Services, especially Leila Barker and interns
701 Andrea George and Kestrel Bailey, the Oregon Veterinary Diagnostic Laboratory (Andree
702 Hunkapiller, Noah Lawler, Donna Mulrooney, Brandy Nagamine, and Ali Al-Fotis) and Michael
703 Harry from OSU's School of Chemical, Biological and Environmental Engineering. Funding for
704 this work was provided by an NSF RAPID grant (#2027679), the David and Lucille Packard
705 Foundation, Pacific Source Health Plans, and the Oregon State University Foundation.

706

707 **Author Contributions**

708 B.A.L. generated and analyzed wastewater ddPCR data, wrote R scripts for automated ddPCR
709 data cleaning and processing, drafted the original version of the manuscript, and created the
710 spatial graphics.

711 D.K. generated wastewater ddPCR data, extracted COVID-19 cases from reports, developed
712 surrogate (BCoV) quantification assay and analyzed wastewater samples for recovery,
713 contributed significantly to the writing and editing of the manuscript.

714 C.K. is the co-PI of the wastewater portion of the project. She helped create the experimental
715 design for the sampling and testing of wastewater, including method development, and actively
716 participated in the collection of wastewater samples. She contributed significantly to data
717 interpretation, writing and editing of the manuscript.

K.J.W. helped create the experimental design for the sampling and testing of wastewater, including method development, and actively participated in the collection of wastewater samples. He contributed to writing and editing of the manuscript.

S.M.B contributed to the implementation of the work and edited the manuscript.

P.G.B. extracted and analyzed the case report data, and contributed to the analysis and interpretation of the data and to writing and editing the manuscript.

J.W.B. is the lead PI of the community survey data collection portion of the project. He co-developed the data collection protocol for the community survey and contributed to editing the manuscript.

K.C. developed and optimized the protocol used to extract RNA from wastewater, contributed to producing the sequence data shown in the paper and editing the manuscript.

B.D.D. contributed to the design and implementation of the work, to the analysis and interpretation of data and to writing and editing of the manuscript.

M.D. designed the sequence strategy used in the project and contributed to producing the sequence data shown in the paper and editing the manuscript.

M.G. co-developed the wastewater sequence analysis pipeline, jointly analyzed the sequence data shown in the paper, and contributed to editing the manuscript.

D.G. co-developed the wastewater sequence analysis pipeline, jointly analyzed the sequence data shown in the paper, and contributed to editing the manuscript.

737 A-M.G. developed and optimized the ddPCR and sequencing protocols used in the project, and
738 contributed to producing the sequence data shown in the paper and editing the manuscript.

739 R.H. contributed to the design and implementation of the work, to the analysis and interpretation
740 of data and to writing and editing the manuscript.

741 K.A.H. contributed to the design and implementation of the work, to the analysis and
742 interpretation of data, and to the editing of the manuscript.

743 D.H. contributed to the design and implementation of the data acquisition protocols for the
744 community surveys and contributed to editing the manuscript.

745 J.L. contributed to the design and implementation of the work, to the analysis and interpretation
746 of data and to writing and editing the manuscript.

747 K.R.M. co-developed the sampling design for the community prevalence surveys, produced
748 prevalence estimates, contributed to other statistical analyses, and contributed to editing the
749 manuscript.

750 F.J.N. contributed to the design and implementation of the work, to the analysis and
751 interpretation of data and to writing and editing the manuscript.

752 A.N. contributed to the design and implementation of the work, to the analysis and interpretation
753 of data and to writing and editing the manuscript.

754 M.P. designed and implemented the data acquisition and management pipelines for the
755 community surveys and for the Illumina sequencing and contributed to editing the manuscript.

A.D.P. created the Newport micro-sewershed shapefiles and generated the demographic data that corresponded with each micro-sewershed and contributed to the writing and editing of the manuscript.

J.L.S is the lead PI of the diagnostic laboratory component of the project. He contributed to the design and implementation of the work, to the analysis and interpretation of data and writing and editing of the manuscript.

B.M.T. designed the integrated sequencing and analysis of wastewater and individual samples, and contributed to the data analysis, writing and editing of the manuscript.

T.S.R. is the lead PI of the wastewater portion of the project. He helped create the experimental design for the sampling and testing of wastewater, including method development, and actively participated in the collection of wastewater samples. As corresponding author, he contributed significantly to data interpretation, writing and editing of the manuscript.

Competing Interest Declaration

The authors declare no competing interests.

Additional Information

Supplementary Information is available for this paper.

Correspondence and requests for materials should be addressed to Tyler S. Radniecki at tyler.radniecki@oregonstate.edu.

Reprints and permissions information is available at www.nature.com/reprints.

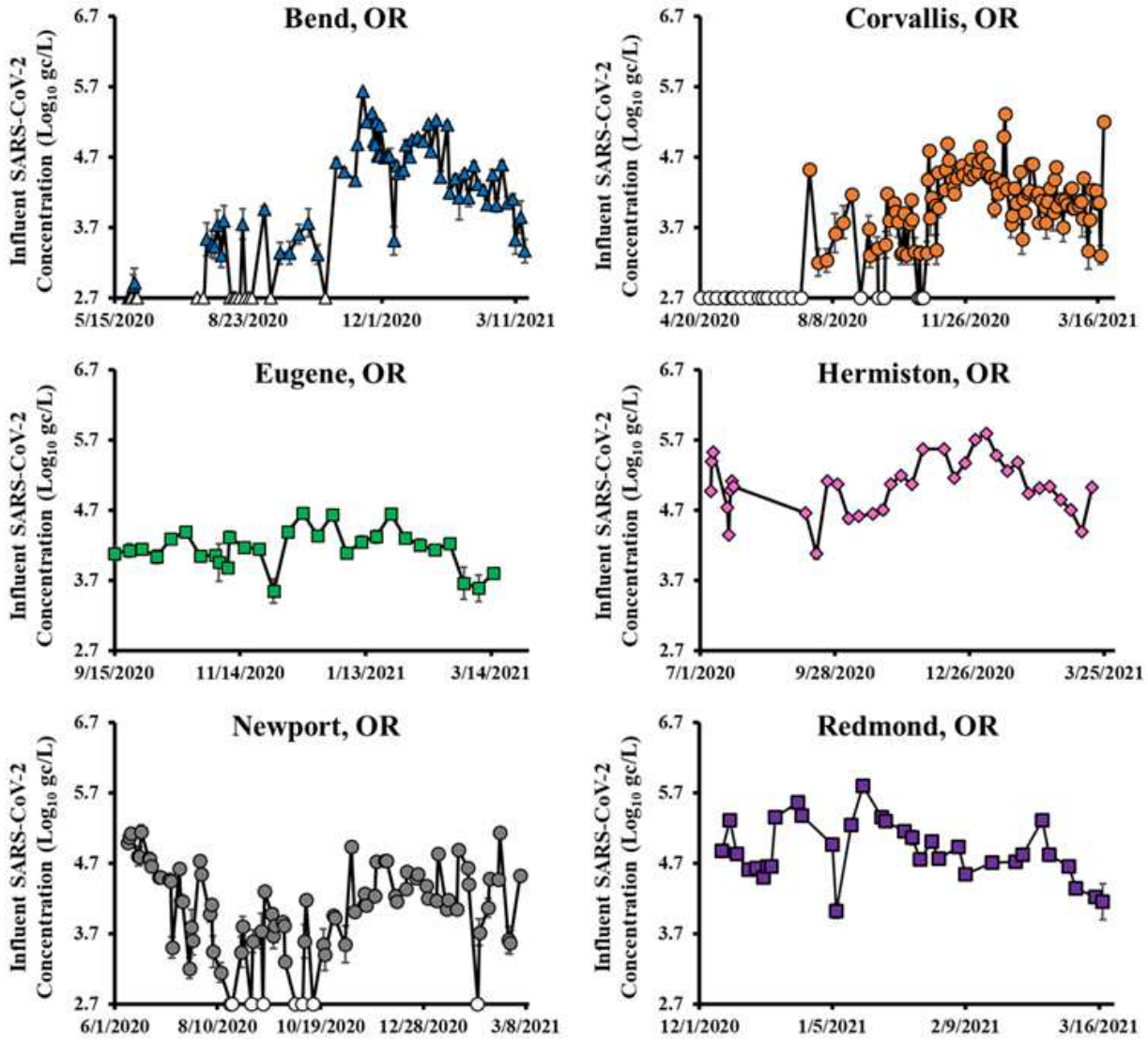
Data Availability

All individual sequences were deposited in GISAID (see **Extended Data Table 6** for accession numbers) and all wastewater sequences were deposited in NCBI's short read archive, under BioProject PRJNA719837 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA719837). Data from community prevalence surveys is in Extended Data Table 2. Wastewater ddPCR data supporting the findings of this study are located within the paper and its supplementary information files.

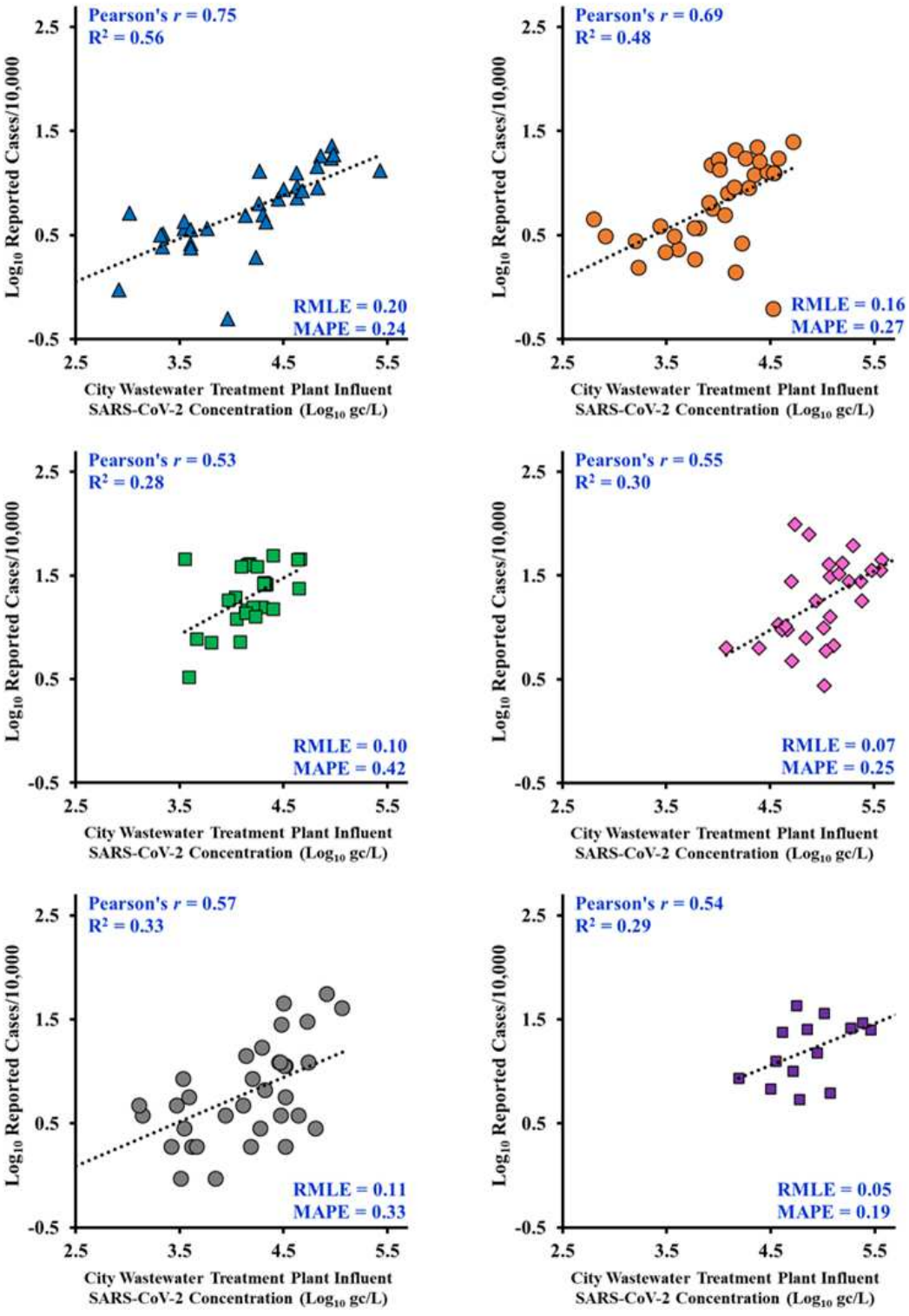
Code Availability

All R scripts used in this analysis will be made publicly available on GitHub at the time of publication. Scripts will be made available to editors and reviewers upon request prior to publication.

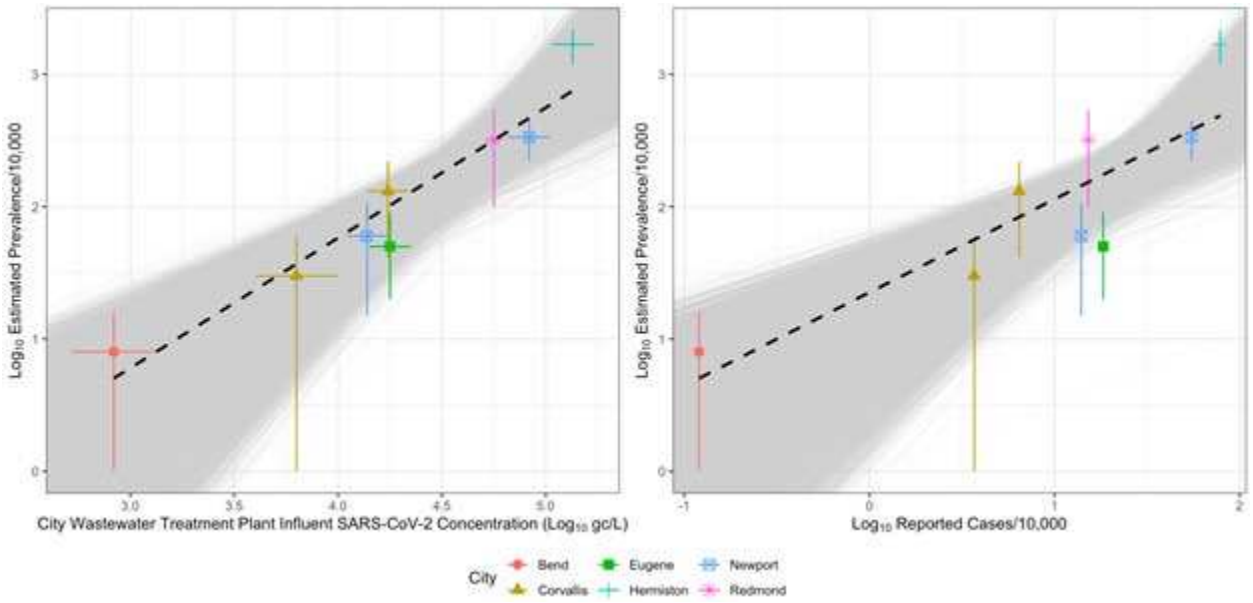
Extended Data Figure and Table Legends



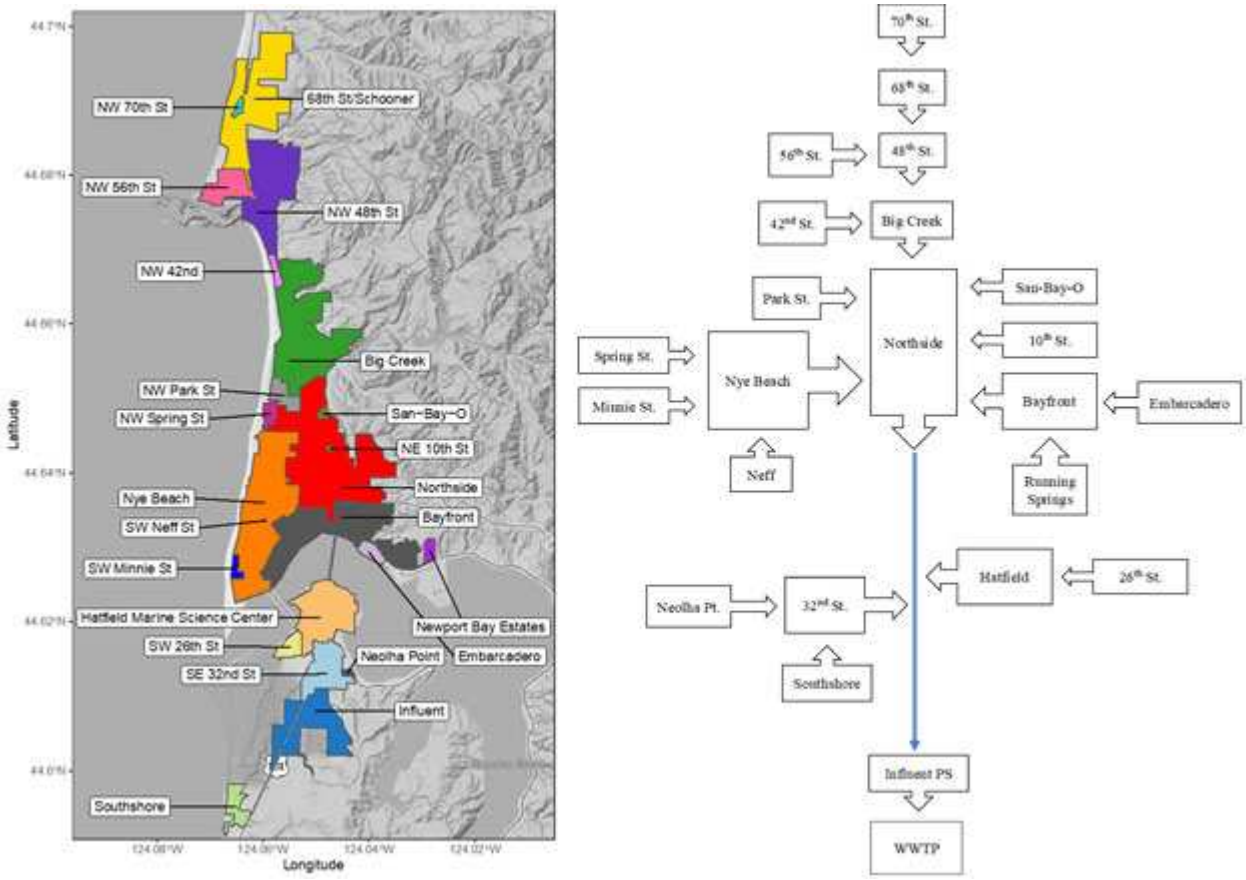
Extended Data Figure 1. Wastewater Time Series Data. The Log_{10} SARS-CoV-2 concentration (Log_{10} gc/L) over time for each participating city. Open symbols represent values below detection limits. Error bars represent the standard error of quadruplicate RT-ddPCR measurements from duplicate samples.



Extended Data Figure 2. Wastewater Concentrations vs Reported COVID 19 Cases for Individual Cities. Log_{10} of wastewater SARS-CoV-2 concentrations versus the Log_{10} of weekly reported COVID-19 cases normalized to each city's population and reported per 10,000 population (reported by ZIP code) for Bend (▲), Corvallis (●), Eugene (■), Hermiston (♦), Newport (●) and Redmond (■), OR.



Extended Data Figure 3. Estimated Prevalence vs. Wastewater Concentration and Reported Cases with Uncertainty from Monte Carlo Simulations. **(Left)** Log_{10} of wastewater SARS-CoV-2 concentrations versus the Log_{10} of estimated prevalence, where the regression line from observed data is shown in a dashed black line. Horizontal and vertical segments indicate one standard error or a 68% credible interval. The gray band is made up of individual regression lines from Monte Carlo simulations. **(Right)** Log_{10} of wastewater SARS-CoV-2 concentrations versus the Log_{10} of reported cases, where the regression line from observed data is shown in a dashed black line. Vertical segments indicate one standard error or a 68% credible interval. The gray band is made up of individual regression lines from Monte Carlo simulations. of individual regression lines from Monte Carlo simulations.



Extended Data Figure 4. Newport, OR Micro-sewersheds. (Left) Location and name of the 22 pump stations and their associated micro-sewersheds sampled in Newport, OR. **(Right)** Flow chart depicting the hierarchical relationships between pump stations. The blue arrow represents a forced main running under the Yaquina Bay and southward toward the wastewater treatment plant.

Extended Data Table 1. City and Wastewater Treatment Plant Statistics

City	County	Population Served	Average Daily Wastewater Flow	Sewer System Type	Collection Type	Sample Matrix
Bend, OR	Deschutes	100,421	6.1 MGD	Separated	24 h Composites	Wastewater Influent
Corvallis, OR	Benton	58,856	18.1 MGD	Combined*	24 h Composites	Wastewater Influent
Eugene, OR	Lane	235,852	52.3 MGD	Separated	24 h Composites	Wastewater Influent
Hermiston, OR	Umatilla	17,782	1.2 MGD	Separated	24 h Composites	Wastewater Influent
Newport, OR	Lincoln	10,853	1.4 MGD	Separated	24 h Composites	Wastewater Influent
Redmond, OR	Deschutes	32,421	2.3 MGD	Separated	24 h Composites	Wastewater Influent

*3.5 mi^2 are combined wastewater and stormwater out of a 8.6 mi^2 sewershed.

Extended Data Table 2. SARS-CoV-2 Prevalence Study Statistics. Samples with an (*) indicate zero positive cases were observed and so the prevalence estimate utilizes additional data on active infections in the community. s.e.: standard error

City	Date	Number of Participating Households	Household Participation Rate	Samples	Estimated Prevalence per 10,000	Lower Bounds per 10,000	Upper Bounds per 10,000	Wastewater SARS-CoV-2 Concentration ($Log_{10} \pm$ s.e.)
Bend, OR	May 30-31, 2020	342	68%	615	8*	1	38	2.9 ± 0.2
Newport, OR	June 20-21, 2020	336	71%	569	335	109	562	4.9 ± 0.1
Newport, OR	July 11-12, 2020	338	70%	550	60	0	150	4.1 ± 0.1
Hermiston, OR	July 25-26, 2020	249	44%	469	1687	714	2660	5.1 ± 0.1
Corvallis, OR	Sept. 26-27, 2020	354	71%	580	30	0	89	3.8 ± 0.2
Eugene, OR	Nov. 7-8, 2020	304	49%	463	50*	10	140	4.3 ± 0.1
Redmond, OR	Jan 29-31, 2021	251	38%	376	320	0	760	4.8 ± 0.0
Corvallis, OR	Mar 12-14, 2021	347	67%	514	131	0	310	4.2 ± 0.1

Extended Data Table 3. Newport Micro-sewershed Characteristics

Pump Station	Drainage Area (acres)	Percent Residential	Estimated Population	Collection Type	Sample Matrix	Sample Dates
10th St.	1.45	100	15	24 h Composites	Wastewater Conveyance	6/18/2020; 7/8/2020
26th St.	27.02	88	131	24 h Composites	Wastewater Conveyance	6/19/2020; 7/9/2020
32nd St.	157.54	51	391	24 h Composites	Wastewater Conveyance	6/19/2020; 7/9/2020
42nd St.	13.04	100	119	24 h Composites	Wastewater Conveyance	6/19/2020; 7/8/2020
48th St.	593.47	75	1771	24 h Composites	Wastewater Conveyance	6/19/2020; 7/8/2020
56th St.	68.49	93	611	24 h Composites	Wastewater Conveyance	6/19/2020; 7/8/2020
68th St./Schooner	297.56	74	1091	24 h Composites	Wastewater Conveyance	6/19/2020; 7/8/2020
70th St.	6.97	100	24	24 h Composites	Wastewater Conveyance	6/19/2020; 7/8/2020
Bayfront	330.41	67	1456	24 h Composites	Wastewater Conveyance	6/19/2020; 7/8/2020; 7/9/2020
Big Creek	952.72	66	3569	24 h Composites	Wastewater Conveyance	6/19/2020; 7/8/2020; 7/9/2020
Embarcadero	34.28	33	470	24 h Composites	Wastewater Conveyance	6/18/2020; 7/8/2020
Hatfield	196.93	12	222	24 h Composites	Wastewater Conveyance	6/18/2020; 7/9/2020
Minnie St.	10.68	100	53	24 h Composites	Wastewater Conveyance	6/18/2020; 7/8/2020
Neff	0.4	100	12	24 h Composites	Wastewater Conveyance	6/19/2020; 7/9/2020
Neolha Point	4.39	100	128	24 h Composites	Wastewater Conveyance	6/18/2020; 7/9/2020
Northside	2163.57	65	9426	24 h Composites	Wastewater Conveyance	6/18/2020; 7/8/2020; 7/9/2020
Nye Beach	381.78	52	3023	24 h Composites	Wastewater Conveyance	6/18/2020; 6/19/2020; 7/8/2020; 7/9/2020
Park St.	33.22	86	742	24 h Composites	Wastewater Conveyance	6/19/2020; 7/8/2020; 7/9/2020
Running Springs/Newport Bay Estates	11.46	100	30	24 h Composites	Wastewater Conveyance	6/18/2020; 7/9/2020
San-Bay-O	5.66	28	18	24 h Composites	Wastewater Conveyance	6/18/2020; 7/8/2020; 7/9/2020
Southshore	40.88	100	603	24 h Composites	Wastewater Conveyance	6/18/2020; 7/9/2020
NW Spring St.	18.47	100	402	24 h Composites	Wastewater Conveyance	6/19/2020; 7/8/2020

Extended Data Table 4. Newport SARS-CoV-2 Variants.

(A) List of the SARS-CoV-2 variants detected in samples from individuals and wastewater obtained during the prevalence studies carried out in Newport, OR on June 20-21 and July 11-12, 2020.

Variant (clade)	B.1.399/NA (G)	B.1/NB (GR)
Common SNPs	C241T, C3037T, C14408T, A23403G	C241T, C3037T, C14408T, A23403G
Distinctive SNPs	C2306T*, C11824T*, A18259G*, G18397A*, A20268G, C28854T, C28887T	G3995T*, C4809T*, A27373T*, T28318C*, G28881A, G28882A, G28883C
Amino acid substitutions	ORF1a:L681F, ORF1b:P314L, ORF1b:I1598V, ORF1b:V1644I, S:D614G, N:S194L, N:T205I	ORF1a:E1244Z, ORF1a:S1515F, ORF1b:P314L, S:D614G, ORF6:M58L, ORF9b:L12S, N:G204R
Closest matching previous sequences in GISAID	EPI_ISL_1231275 (Finland 2020-03-20) EPI_ISL_534859 (England 2020-03-27) EPI_ISL_534827 (England 2020-03-20) EPI_ISL_494444 (California 2020-04-16)	EPI_ISL_463510 (Washington 2020-04-29) EPI_ISL_1015032 (France 2020-03-16) EPI_ISL_895835 (Czech Repub. 2020-03-16) EPI_ISL_895820 (Czech Repub. 2020-04-01)

** unique SNPs used for quantifying variant RNAs*

(B) Additional SARS-CoV-2 variants commonly detected in Newport wastewater samples

Variant (clade)	B.1.1.158/CC (GR)	B.1.369/EE (GH)	B.1.2/FF (GH)
Common SNPs	C241Δ, C14408T, A23403G	C241T, C1059T, C3037T, C14408T, A23403G	C241T, C1059T, C3037T, C14408T, A23403G
Distinctive SNPs	C3042T, A10089G*, G11083T, C16393T*, C18086T, C19152T*, T19839C, C25207T, G28881A, C28807T*, G28882A, G28883C	C16260T*, C18695T*, C19928T*, C20946T*, G25563T, C28821T*	T1927T, C10319T, A18424G*, C21304T*, A22255T, G25563T, G25907T*, C26974T, C28472T*, C28869T
Amino acid substitutions	ORF1a:K3275R, ORF1a:L3606F, ORF1b:P314L, ORF1b:P976S, ORF1b:T1540I, S:D614G, N:G204R	ORF1a:T265I, ORF1b:P314L, ORF1b:T1743I, ORF1b:S2145F, S:D614G, ORF3a:Q57H, N:S183Y	ORF1a:T265I, ORF1a:L3352F, ORF1b:P314L, ORF1b:N1653D, ORF1b:R2613C, ORF3a:Q57H, ORF3a:G172V, ORF8:S24L, N:P67S, N:P199L
Representative sequences in GISAID	EPI_ISL_1004180 EPI_ISL_1004181 EPI_ISL_1004182 EPI_ISL_1004183	EPI_ISL_1004190 EPI_ISL_1364474	EPI_ISL_1017328 EPI_ISL_1017293

** unique SNPs used for quantifying variant RNAs*

Extended Data Table 5. SARS-CoV-2 Variant Relative Abundance. Relative abundances of variants detected in samples from individuals and wastewater across micro-sewersheds during the June 20-21, 2020 prevalence survey.

	Individuals (June 20-21, 2020)				Wastewater (June 18 or 19, 2020)				
Location	Numbers*		Fraction		Fraction	s.e.‡	Fraction	s.e.‡	Reps§
	NA†	NB†	NA†	NB†	NA†		NB†		
48th St	0	0	0	0	0.72	-	0	-	1
56th St	0	0	0	0	0.94	-	0	-	1
Bayfront	0	1	0	1	0.92	0.01	0	0	2
Big Creek	1	0	1	0	0.98	0.01	0	0	2
Northside	11	2	0.85	0.15	0.88	0.04	0.06	0.06	3
Nye Beach	6	1	0.86	0.14	0.3	0.16	0.52	0.08	3
Park St	0	0	0	0	0.94	-	0	-	1
San-Bay-O	0	0	0	0	1	-	0	-	1
Influent	11	2	0.85	0.15	0.7	0	0.04	0.04	3

** Numbers of individuals counted for each sewershed include not only those residing within the sewershed itself, but also those residing within upstream sewersheds flowing into the sewershed*

† NA = B.1.399/NA; NB = B.1/NB.

‡ Standard error calculated on untransformed fractions.

§ Measurements were from a single assay (1), from assays of two independent water samples (2) or from two assays of RNA from one sample and one assay of a second independent sample (3).

Extended Data Table 6. Accession Numbers of Individual SARS-CoV-2 RNA Sequences

B.1.399/NA	EPI_ISL_1004213
B.1.399/NA	EPI_ISL_1004214
B.1.399/NA	EPI_ISL_1004215
B.1.399/NA	EPI_ISL_1004216
B.1.399/NA	EPI_ISL_1004217
B.1.399/NA	EPI_ISL_1254704*
B.1.399/NA	EPI_ISL_1254705
B.1.399/NA	EPI_ISL_1254706
B.1.399/NA	EPI_ISL_1254709
B.1.399/NA	EPI_ISL_1254710
B.1/NB	EPI_ISL_1004178
B.1/NB	EPI_ISL_1254711
B.1/NB	EPI_ISL_1254712
B.1.1.289	EPI_ISL_1254713*
B.1.1.289	EPI_ISL_1254714*

**Low coverage sequence*

Extended Data Table 7. Summary of all RT-ddPCR QC Reactions. Improved plate handling procedures took effect starting with plate 8 (see Supplementary Information). Only non-detect sample data from plates < 8 were used. NTC = No-Template Control.

		Copies Per Reaction (all plates)					
Control Type	**Target**	**Min**	**Median**	**Max**	**% Fail (*n*, all plates)**	**% Fail (*n*, plates < 8)**	**% Fail (*n*, plates ≥8)**
Extraction Blank	N1	0.0	0.0	2.4	0.0 (158)	0.0 (6)	0.0 (152)
Extraction Blank	N2	0.0	0.0	7.9	1.3 (158)	16.7 (6)	0.7 (152)
Extraction Blank	RP	0.0	0.0	52.0	2.5 (158)	33.3 (6)	1.3 (152)
Field Blank	N1	0.0	0.0	1.9	0.0 (49)	0.0 (16)	0.0 (33)
Field Blank	N2	0.0	0.0	6.1	2.0 (49)	0.0 (16)	3.0 (33)
Field Blank	RP	0.0	0.0	27.0	4.1 (49)	12.5 (16)	0.0 (33)
Negative Control	N1	0.0	0.0	11.9	1.2 (165)	22.2 (9)	0.0 (156)
Negative Control	N2	0.0	0.0	12.6	3.0 (165)	55.6 (9)	0.0 (156)
Negative Control	RP	57.2	392.1	1142.9	0.0 (165)	0.0 (9)	0.0 (156)
NTC	N1	0.0	0.0	17.0	3.6 (169)	33.3 (18)	0.0 (151)
NTC	N2	0.0	0.0	18.5	8.9 (169)	77.8 (18)	0.7 (151)
NTC	RP	0.0	0.0	91.9	8.3 (169)	77.8 (18)	0.0 (151)
Positive Control	N1	43.2	407.8	2973.2	0.0 (164)	0.0 (9)	0.0 (155)
Positive Control	N2	42.1	482.4	2818.8	0.0 (164)	0.0 (9)	0.0 (155)
Positive Control	RP	86.8	399.8	1875.6	0.0 (164)	0.0 (9)	0.0 (155)

Supplementary Information

Limit of Blank and Limit of Detection

The limit of blank (LOB) for N1 and N2 was 2.0 and 4.2 copies per reaction, respectively. Note that both of these values are below the three-droplet threshold; only reactions with 1 or 2 droplets yielded copy numbers at that level, so we are confident in our choice of threshold for calling positive reactions. All LOB reactions were below the positive threshold for N1, and only 4 out of 104 non-target reactions had 3 or more droplets in N2, which is a false positive rate of 4%.

The predicted limits of detection (LOD) based on the LOB results were 4 and 12 copies per reaction for N1 and N2, respectively. In order for a LOD estimate to be valid, greater than 95% of test reactions at the predicted LOD value need to amplify above the LOB. The LOD of N2 was confirmed to be 12 copies per reaction, as 58/60 (97%) test reactions at that concentration had copy numbers above the N2 LOB. The N1 LOD apparently lies somewhere between 4 and 12 copies per reaction: at 12 copies per reaction, all 60 reactions amplified above the N1 LOB, but at 4 copies per reaction, 13/60 (22%) reactions amplified below the N1 LOB. Using a parametric method (which is an imperfect estimate in this case because the test reaction data are not normally distributed), the N1 LOD was estimated to be 8 copies per reaction.

RT-ddPCR Quality Control

The quality controls for the wastewater RT-ddPCR method included field blanks, extraction blanks, negative control reactions (containing only human gDNA), positive control reactions (containing synthetic RNA of SARS-CoV-2 assay targets and human gDNA), and no-template controls (NTCs). The results of all QC reactions are summarized in **Extended Data Table 7**.

Early on in our study we had a cross-contamination issue with the PCR sample plates resulting from pandemic distancing requirements to prepare the PCR plate in a separate laboratory from the ddPCR system. Initially, we were sealing the prepared PCR plates with an adhesive optical film for transport to the ddPCR laboratory. This seal was then removed upon arrival, and the plate was re-sealed with the Bio-Rad heat sealing foil. We hypothesize that the removal of this adhesive seal caused the cross-contamination of our plates due to aerosolized sample droplets.

Once we stopped the practice of using adhesive seals to transport the plates and instead exclusively used heat sealing foil, the contamination issue was resolved (**Extended Data Table 7**). We chose to use non-detect sample data from the contaminated plates, while samples with amplification were rerun using archived sample material stored in DNA/RNA Shield (Zymo Research Inc., Irvine, CA) at 80 ºC. Only positive detections which passed QA/QC were included in this study.

Bovine coronavirus (BCoV) process recovery control

BCoV solution was prepared from freeze-dried Calf Guard cattle vaccine (Zoetis, NJ, USA). After rehydrating in 3 mL of sterile diluent provided by the manufacturer, the BCOV solution was divided into 100 µL aliquots and stored at –20 ºC. To use, the BCoV stock solution aliquot was thawed on ice and vortexed thoroughly; each aliquot was used for a maximum of two freeze-thaw cycles. BCoV stock solution was spiked into wastewater samples at a ratio of 1/1000 (v/v) before the concentration step. To determine the concentration of BCoV stock solution, 10 µL of the stock was spiked in to 390 µL PBS and 200 µL of that mixture was extracted following the same protocols used for wastewater sample extractions as described in the main text. One

extraction blank, prepared with PBS, was included on each plate as an RNA extraction contamination control. The extracted RNA was then serially diluted (1:10) in nuclease-free water for six dilutions and ran in duplicate using a previously established BCoV assay by following the one-step RT-ddPCR procedure[47]. Stock concentration of BCoV was around 230,000 gc/µL.

Process efficiency (*i.e.,* viral recovery) was calculated by dividing the final quantity of BCoV measured in wastewater samples to the quantity of BCoV spiked to each wastewater sample before concentration. The BCoV recovery was found as 57 (± 4) % and used to assess SARS-CoV-2 RNA loss during sample processing. Non-spiked wastewater samples were also quantified for BCoV to assess background concentration and BCoV was not detected in non-spiked wastewater samples, indicating BCoV is an appropriate viral surrogate.

Viral Concentration Calculation

$$\frac{copies}{rxn} * \frac{1\,rxn}{Template\,Volume\,(\mu L)} * Elution\,Volume\,(\mu L) * \frac{1}{Lysate\,Volume\,(\mu L)} * Shield\,Volume\,(\mu L) * \frac{1}{Volume\,Filtered\,(mL)} * 1000\frac{mL}{L} = \frac{copies}{L}$$

Supplementary Equation 1. Conversion of Copies per Reaction to Gene Copies per Liter of Wastewater