Short Communication

# SARS-CoV-2 variant detection at a university dormitory using wastewater genomic tools

Van Vo [a,1], Richard L. Tillett [b,1], Ching-Lan Chang [a], Daniel Gerrity [c], Walter Q. Betancourt [d,*], Edwin C. Oh [a,b,e,**]

[a] Laboratory of Neurogenetics and Precision Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA
[b] Nevada Institute of Personalized Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA
[c] Southern Nevada Water Authority, P.O. Box 99954, Las Vegas, NV 89193, USA
[d] Water & Energy Sustainable Technology (WEST) Center, University of Arizona, 2959 W. Calle Agua Nueva, Tucson, AZ 85745, USA
[e] Department of Internal Medicine, UNLV School of Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

HIGHLIGHTS

- SARS-CoV-2 sequencing of dormitory wastewater grab samples identifies viral variants.
- 1.1×10^6 gene copies/L in wastewater sufficient for WGS
- Viral shedding from 3 infected students suggests circulation of 2 cluster strains.
- WGS paired with WBE can help address declining testing challenges.

GRAPHICAL ABSTRACT
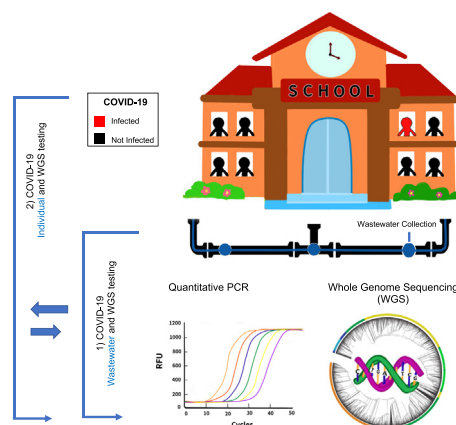
ARTICLE INFO

ABSTRACT

In the Fall of 2020, university campuses in the United States resumed on-campus instruction and implemented wastewater monitoring for the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). While quantitative polymerase chain reaction (qPCR) tests were deployed successfully to detect viral RNA in wastewater across campuses, the feasibility of detecting viral variants from a residential building like a dormitory was unclear. Here, we demonstrate that wastewater surveillance from a dormitory with at least three infected students could lead to the identification of viral genomes with more than 95% coverage. Our results indicate that viral variant detection from wastewater is achievable at a dormitory and that coronavirus disease 2019 (COVID-19) wastewater surveillance programs will benefit from the implementation of viral whole genome sequencing at universities.

© 2021 Elsevier B.V. All rights reserved.

* Corresponding author.
** Correspondence to: E.C. Oh, Laboratory of Neurogenetics and Precision Medicine, University of Nevada Las Vegas, Las Vegas, NV 89154, USA.
 E-mail addresses: wbetancourt@arizona.edu (W.Q. Betancourt), edwin.oh@unlv.edu (E.C. Oh).
 [1] These first authors contributed equally to this article.

1. Introduction

Whole genome sequencing (WGS) of SARS-CoV-2 from large wastewater treatment plants has proven to be an emerging technology for tracing viral evolution (Nemudryi et al., 2020; Crits-Christoph et al., 2021; Izquierdo-Lara et al., 2021); however, the utility of such tools at the level of a single building is only starting to emerge (Spurbeck et al., 2021). Several studies have suggested that successful genome sequencing requires at least 2.8×10^5 viral copies per liter of wastewater (Nemudryi et al., 2020; Crits-Christoph et al., 2021; Izquierdo-Lara et al., 2021), yet this calculation can be variable and dependent on sample preparation and concentration procedures.

To test whether WGS can be used to identify SARS-CoV-2 genomes and emerging variants from a dormitory, we sequenced and analyzed wastewater samples (Betancourt et al., 2021) during the reopening of the University of Arizona in the Fall semester of 2020. Using RNA samples that contained on average 1.1×10^6 SARS-CoV-2 copies/L, we identified viral signatures in the dormitory wastewater sample that included clade 20C and a 20A+20268G subclade. Our findings intimate that transmission of at least two cluster strains in college students was occurring from the onset of the potential outbreak. Obtaining wastewater viral genomes from a dormitory with at least three infected individuals out of 311 residents supports the hypothesis that wastewater WGS has the potential to complement other public health resources to monitor patterns of viral transmission at the level of a single building. In addition, the utility of wastewater WGS for variant surveillance may serve an even more important role in 2021/2022 due to 1) the decline of clinical testing, and 2) testing and vaccination hesitancy in communities.

2. Methods

2.1. SARS-CoV-2 wastewater sample collection and preparation

The University of Nevada, Las Vegas Institutional Review Board (IRB) reviewed this project and determined this study to be not human subject research according to federal regulations and University policy. The Institutional Biosafety Committee (IBC) of the University of Nevada, Las Vegas approved methods and techniques used in this study. Two 24-hour composite wastewater influent samples (post-screening but prior to primary sedimentation) were collected from a wastewater treatment plant (WWTP) in Tucson, Arizona serving a population of ~500,000. Five sewage grab samples were collected on the morning of August 26, 2020 between 8:30 am and 8:50 am from a manhole servicing a dormitory at the University of Arizona. The dormitory housed approximately 300 students and staff at the time of sampling. Aliquots of wastewater (70 ml) were concentrated using stepwise vacuum filtration through membrane filters of 0.8, 0.65, 0.45 and 0.22 μm pore sizes (EMD Millipore, Billerica, MA) followed by centrifugal ultrafiltration of the filtrate using the CentriconPlus-70 filter, 100 kDa cutoff (EMD Millipore, Billerica, MA). The final concentrate was used for RNA extraction, as described previously (Betancourt et al., 2021).

Viral RNA was extracted from wastewater using the QIAGEN QIAmp Viral Mini Kit AllPrep kit (Qiagen Cat# 52906 for dormitory samples), and the PowerViral DNA/RNA Kit (Qiagen Cat# 28000-50 for WWTP samples), according to the manufacturer's instructions and eluted in RNAse free dH₂O. DNA was digested using DNase I and RNA was purified using E.Z.N.A. MicroElute RNA Clean-up Kit (Omega Bio-Tek Cat# R6247-01) according to the manufacturer's instructions and eluted in RNAse free dH₂O.

2.2. Targeted whole genome sequencing and variant analysis

Whole genome sequencing libraries were constructed using the CleanPlex SARS-CoV-2 FLEX Panel from Paragon Genomics by following the manufacturer's instructions. Total RNA (10 ng) was processed for first-strand cDNA synthesis. Libraries were sequenced using an Illumina NextSeq 500 platform and a mid-output v2.5 (300 cycles) flow cell. Illumina adapter sequences were trimmed from reads using cutadapt v3.2 (Martin, 2011). All sequencing reads were mapped to SARS-CoV-2 genome (NC_045512.2) using bwa mem v0.7.17-r1188 (Li, 2013). Amplicon primers were trimmed from aligned reads using fgbio TrimPrimers v1.3.0 and variants were called with iVar variants v1.3 (Grubaugh et al., 2019). Genome coverages were calculated by samtools coverage v1.10 (Li et al., 2009). Coverage of the SARS-CoV-2 genomes was above 95% with a median depth above 50×. Phylogenetic analysis was conducted on March 1, 2021, by using Nextstrain and open-source tools for visualizing viral genomes (Hadfield et al., 2018). Genome sequences of 3988 isolates were obtained from GISAID and aligned to produce a phylogenetic tree using the NGphylogeny.fr FastME workflow (Lemoine et al., 2019) (Supplementary Table 1). Nextstrain clade memberships of genomes were inferred using nextstrain-cli v1.16.5 and applied to the FastME tree.

3. Results and discussion

As part of a larger public health surveillance effort, morning wastewater samples were obtained daily from dormitory manholes at the University of Arizona. On the morning of August 26, 2020, five grab samples collected from a single manhole were determined to be positive for SARS-CoV-2 at an average concentration of 1.1×10^6 gene copies/L (Betancourt et al., 2021). In response to the wastewater result, clinical tests for COVID-19 diagnosis were conducted for all dormitory residents on the same day, leading to the identification and isolation of three infected students (two asymptomatic and one symptomatic) (Betancourt et al., 2021). Total RNA from wastewater was used to generate viral libraries and sequenced on an Illumina NextSeq 500 platform. Analysis of the sequencing results identified a mixture of two viral signatures that included clade 20C, with signature 20C mutations C1059T and G25563T, and another emerging viral strain corresponding to a 20A+20268G subclade (Fig. 1). The nucleotide change in the 20A+20268G strain is synonymous at the amino acid level and is not predicted to influence biological function of SARS-CoV-2.

We analyzed wastewater samples obtained from wastewater treatment plants (WWTP) in Arizona during June 2020 to determine if the 20A+20268G subclade was present in the Tucson metropolitan area prior to the reopening of the dormitory. Analysis of the WWTP viral genomes showed no evidence for the 20A+20268G subclade (Supplementary Fig. 1; Appendix), suggesting that the 20A+20268G subclade was not present in Tucson, Arizona prior to Fall of 2020, or that the limit of detection of our sequencing techniques was insufficient to detect this lineage in the treatment plant samples. The A20268G mutation discovered in our dormitory sample has been characterized previously in Europe (Lobiuc et al., 2021). In fact, predictions through the Pangolin webserver demonstrate that SARS-CoV-2 genomes with this mutation belong to the B.1.5 sub-lineage. Interestingly, this mutation has been characterized as outbreak-defining signature of SARS-CoV-2 strains in Suceava County, Romania (Lobiuc et al., 2021).

WGS of SARS-CoV-2 genomes from municipal sewage treatment plants servicing large residential and commercial area has been achieved previously (Nemudryi et al., 2020; Crits-Christoph et al., 2021; Izquierdo-Lara et al., 2021). In these studies, hybrid capture (Illumina respiratory virus panel) (Crits-Christoph et al., 2021) and amplicon-based (Oxford Nanopore) (Nemudryi et al., 2020; Izquierdo-Lara et al., 2021) library preparation kits were used to amplify the target genomes before sequencing on an Illumina instrument. Here, we demonstrate that our processing (Betancourt et al., 2021; Gerrity et al., 2021) and genomic tools, using another amplicon-based approach (Paragon Genomics), have the specificity and sensitivity to sequence single viral genomes from an individual housing unit like a university dormitory. Interestingly, the observation of two viral signatures in the wastewater sample (Fig. 1) suggests that transmission of at least two cluster strains in college students was occurring from the onset of the potential
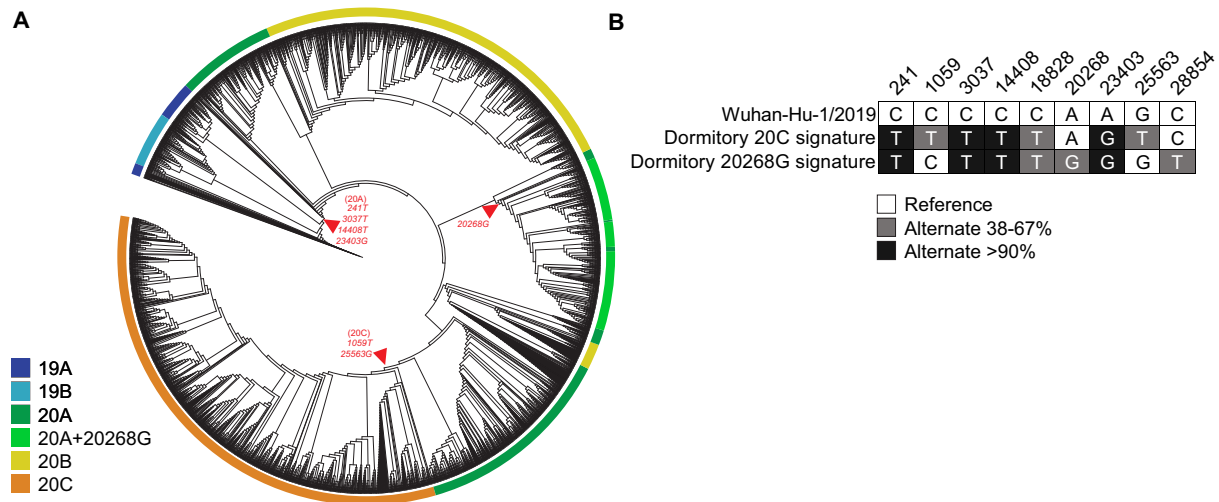
Fig. 1. Phylogenetic analysis of SARS-CoV-2 sequence isolated from a dormitory building. (A) Maximum-likelihood phylogeny of the SARS-CoV-2-related lineages. The outer ring is colored according to assigned Nextstrain clade. The tree is rooted relative to the Wuhan-Hu-1/2019 genome. Characteristic mutations of clades of interest are indicated at arrows. (B) Sequences isolated from the Arizona dormitory wastewater clade and the sequences of the reference Wuhan-Hu-1/2019 reference strain. A comparison of mutations at nine genome positions is shown in the inset. The reference sequence is shown across the top. Observed allele frequencies in wastewater sample are indicated by color in either grey (38–67% alternate frequency) or black (>90% alternate frequency). Signature mutations of clades 20C and 20A+20268G observed in dormitory sample are shown on separate rows.

outbreak (Betancourt et al., 2021). However, our study did not compare clinical genomes from infected students and our conclusions about potential transmission remain limited.

The findings in this study demonstrate that WGS of wastewater samples can be performed in a laboratory setting and does not require extensive resources that are typically found in large universities or biotechnology companies with comprehensive core services. In addition, due to the relatively small size of the SARS-CoV-2 genome, smaller desktop Illumina instruments can be utilized for sequencing hundreds of viral genomes using a single flow cell. Given the availability of such resources, we propose the inclusion of whole genome sequencing in wastewater surveillance programs with other public health measures to identify viral variants that may be present in a university dormitory. Currently, both qPCR and digital PCR (dPCR) strategies are utilized to determine whether viral load can be detected and whether potential variants of concern may be circulating in a community; while costs can be somewhat limited by using PCR techniques to identify signature mutations in viral strains, the implementation of WGS enables public health authorities to definitively identify viral strains and manage local outbreaks. Such sequencing techniques are especially important given the emergence and evolution of new lineages and sub-lineages of variants over the last 12 months (Chen et al., 2021; Bugembe et al., 2021). Moving forward, we can envision a situation when qPCR/dPCR strategies will be used to triage samples for more detailed analyses like WGS. In addition, the combination of such genomic resources, especially when diagnostic testing may be limited, will have the potential to link the earliest encounter of a viral variant to the index case(s), thereby limiting the spread of an infectious disease.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.scitotenv.2021.149930.

CRediT authorship contribution statement

Van Vo: Methodology, Supervision, Formal analysis, Writing – original draft. **Richard** L. **Tillett:** Methodology, Supervision, Formal analysis, Writing – original draft. **Ching-Lan Chang:** Resource, Writing – reviewing & editing. **Daniel Gerrity:** Resource, Writing – reviewing & editing. **Walter Q. Betancourt:** Formal analysis, Visualization, Writing – original draft. **Edwin C. Oh:** Formal analysis, Funding acquisition, Project administration, Visualization, Supervision, Investigation, Writing – original draft. **Van Vo** and **Richard** L. **Tillett** are co-first authors and contributed equally. **Walter Q. Betancourt** and **Edwin Oh** are co-corresponding authors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Betancourt, W.Q., Schmitz, B.W., Innes, G.K., Brown, K.M., Prasek, S.M., Brown, K.M., Stark, E.R., Foster, A.R., Sprissler, R.S., Harris, D.T., Sherchan, S.P., Gerba, C.P., Pepper, I.L., 2021. COVID-19 containment on a college campus via wastewater-based epidemiology, targeted clinical testing and an intervention. Published online 2021 Mar 13Sci Total Environ. 779, 146408. https://doi.org/10.1016/j.scitotenv.2021.146408 (Jul 20; PMCID: PMC7954642).

Bugembe, D.L., Phan, M.V.T., Ssewanyana, I., Semanda, P., Nansumba, H., Dhaala, B., Nabadda, S., O'Toole, Á.N., Rambaut, A., Kaleebu, P., Cotten, M., 2021. Emergence and spread of a SARS-CoV-2 lineage a variant (A.23.1) with altered spike protein in Uganda. Nat. Microbiol. 6 (8), 1094–1101. https://doi.org/10.1038/s41564-021-00933-9 (Aug).

Chen, A.T., Altshuler, K., Zhan, S.H., Chan, Y.A., Deverman, B.E., 2021. COVID-19 CG enables SARS-CoV-2 mutation and lineage tracking by locations and dates of interest. elife 23 (10), e63409. https://doi.org/10.7554/eLife.63409 (Feb).

Crits-Christoph, A., Kantor, R.S., Olm, M.R., Whitney, O.N., Al-Shayeb, B., Lou, Y.C., Flamholz, A., Kennedy, L.C., Greenwald, H., Hinkle, A., Hetzel, J., Spitzer, S., Koble, J., Tan, A., Hyde, F., Schroth, G., Kuersten, S., Banfield, J.F., Nelson, K.L., 2021. Genome sequencing of sewage detects regionally prevalent SARS-CoV-2 variants. mBio 12 (1), e02703–e02720. https://doi.org/10.1128/mBio.02703-20 (Jan 19).

Gerrity, D., Papp, K., Stoker, M., Sims, A., Frehner, W., 2021. Early-pandemic wastewater surveillance of SARS-CoV-2 in Southern Nevada: methodology, occurrence, and incidence/prevalence considerations. Water Res. X 10, 100086. https://doi.org/10.1016/j.wroa.2020.100086 (Jan 1; Epub 2020 Dec 31. PMID: 33398255).

Grubaugh, N.D., Gangavarapu, K., Quick, J., Matteson, N.L., De Jesus, J.G., Main, B.J., Tan, A.L., Paul, L.M., Brackney, D.E., Grewal, S., Gurfield, N., 2019. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol. 20 (1), 1–9 (Dec).

Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., Neher, R.A., 2018. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 34 (23), 4121–4123 (Dec 1).

Izquierdo-Lara, R., Elsinga, G., Heijnen, L., Munnink, B.B.O., Schapendonk, C.M.E., Nieuwenhuijse, D., Kon, M., Lu, L., Aarestrup, F.M., Lycett, S., Medema, G., Koopmans, M.P.G., de Graaf, M., 2021. Monitoring SARS-CoV-2 circulation and diversity through community wastewater sequencing, the Netherlands and Belgium.

Emerg. Infect. Dis. 27 (5), 1405–1415. https://doi.org/10.3201/eid2705.204410 (May).

Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S., Gascuel, O., 2019. NGPhylogeny. fr: new generation phylogenetic services for non-specialists. Nucleic Acids Res. 47 (W1), W260-5 (Jul 2).

Li, H., 2013. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM (Mar 16; arXiv preprint arXiv:1303.3997).

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25 (16), 2078–2079 (Aug 15).

Lobiuc, A., Dimian, M., Gheorghita, R., OAC, Sturdza, Covasa, M., 2021. Introduction and characteristics of SARS-CoV-2 in North-East of Romania during the first COVID-19

outbreak. 12, 654417. https://doi.org/10.3389/fmicb.2021.654417 (Jul 7; eCollection 2021. PMID: 34305826).

Martin, M., 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. J. 17 (1), 10–12 (May 2).

Nemudryi, A., Nemudraia, A., Wiegand, T., Surya, K., Buyukyoruk, M., Cicha, C., Vanderwood, K.K., Wilkinson, R., Wiedenheft, B., 2020. Temporal detection and phylogenetic assessment of SARS-CoV-2 in municipal wastewater. Cell Rep. Med. 1 (6), 100098. https://doi.org/10.1016/j.xcrm.2020.100098 (Sep 22; Epub 2020 Aug 31. PMID: 32904687).

Spurbeck, R.R., Minard-Smith, A., Catlin, L., 2021. Feasibility of neighborhood and building scale wastewater-based genomic epidemiology for pathogen surveillance. Sci. Total Environ. 1 (789), 147829. https://doi.org/10.1016/j.scitotenv.2021.147829 (Oct).