# An allele-specific primer extension assay to quantify the proportion of B.1.1.7-specific SARS-CoV-2 RNA in wastewater.

Tyson E. Graber[1], Élisabeth Mercier[2], Patrick M. D'Aoust[3], Huy-Dung Hoang[1,4], Xin Tian[3], Syeda Tasneem[3], Kamya Bhatnagar[3], Robert Delatolla[3*]

[1] Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada, K1H 8L1
[2] Department of Chemical Engineering, University of Ottawa, Canada, K1N 6N5
[3] Department of Civil Engineering, University of Ottawa, Ottawa, Canada, K1N 6N5
[4] Department of Biochemistry, Microbiology and Immunology, University of Ottawa, Ottawa, Canada, K1H 8M5

Corresponding author:

**Dr. Robert Delatolla**

Associate Professor

Work E-mail: robert.delatolla@uottawa.ca

## Abstract

The coronavirus disease 2019 (COVID-19) pandemic caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has claimed millions of lives globally to date. Rapid accumulation of co-occurring mutations has led to the emergence of viral variants which appear to be more transmissible, virulent, or both. Variants of concern (VOC) now include those belonging to the B.1.1.7, B.1.351, and P.1 lineages. Early detection of VOC and the ability to retrospectively follow their respective, longitudinal prevalence in communities is wanting. Wastewater-based epidemiology (WBE) allows tracking of disease prevalence in the general population using RT-qPCR to detect viral fragments, but ongoing longitudinal studies have yet to differentiate between these variants. Here, we describe and validate a primer extension strategy to amplify and distinguish B.1.1.7-specific, from non-B.1.1.7 alleles by combining new forward, and existing CDC 2019-nCoV_N1 qPCR probes and reverse primers. This assay can be quickly implemented within a current SARS-CoV-2 WBE framework with minimal cost with the goal of providing early detection of increasing B.1.1.7 transmission in a community prior to identification through clinical testing and confirmation via secondary screening strategies. As such, this assay can provide public health units with an additional and much needed metric to be able to rapidly triangulate B.1.1.7 prevalence.

**Keywords:** COVID-19; SARS-CoV-2; wastewater; variant; B.1.1.7; WBE; qPCR

## Introduction

Since the beginning of the COVID-19 pandemic, the possibility of viral mutations within the SARS-CoV-2 genome was outlined as a potential risk[1]. SARS-CoV-2, being a single-stranded RNA (ssRNA) virus, is susceptible to frequent mutations[2–5]. Generally, the majority of viral mutations have a low impact on disease in populations[6], but in late 2020, several concurrent mutations emerged which increase infectivity[7,8] and potentially decrease sensitivity to neutralizing antibodies[9,10]. As of February 2021, these variants of concern (VOC) include: 1) the B.1.1.7/501Y.V1 variant (also known as the UK variant); 2) the B.1.351/501Y.V2 variant (also known as the South African variant) and 3) the P.1/501Y.V3 variant (also known as the Brazilian variant). With communities around the world already reeling from the effects of the initial waves of the pandemic, the onset and widespread resurgence of newer variants, which may be more infectious and might exhibit a higher case fatality rate than ancestral variants, are exhausting public health resources and effecting the mental health of communities.

SARS-CoV-2 WBE is being adopted as a means of surveillance in many jurisdictions around the world (reviewed in [11]). The common method currently used for SARS-CoV-2 WBE is RT-qPCR based on probe and primer sets used for clinical testing[12]. These commonly used probe and primer sets to date do not distinguish VOCs, including B.1.1.7. The high specificity and sensitivity of the CDC "2019-nCoV N1" (N1) and "2019-nCoV N2" (N2) probe/primer sets seen in clinical samples[12], appears to be mirrored in wastewater matrices[11] although relative sensitivities haven't been formally investigated. N1 and N2 assays are thus used by many WBE research groups including those participating in the wastewater-based surveillance initiative in Ontario, Canada. Here, we describe and validate an RT-qPCR assay that targets a B.1.1.7-specific allele that lies adjacent to the N1 region.

## Results and Discussion

### *Targeting the N gene D3L mutation defines a B.1.1.7-specific RT-qPCR assay in wastewater.*

Diagnosing a B.1.1.7 VOC case requires an assay with a very high positive predictive value. This is difficult to achieve in a singleplex qPCR assay, as targeting multiple, co-occurring mutations at different loci may be needed to unambiguously define a VOC. Thus, multiplexed assays or secondary screens incorporating Sanger sequencing and/or whole-genome sequencing are required to confirm the genetic lineage of a virus isolated from clinical specimens. Unlike individual testing, wastewater contains mixtures of genomic fragments sourced from multiple infections that can be probed across the course of disease in a given population of individuals. In essence, wastewater surveillance of SARS-CoV-2 allows unbiased sampling of population-level allele frequencies in a single test tube. Thus, additional granularity that is represented in wastewater samples includes genomes from multiple individuals collected over continuous periods of time. This property provides

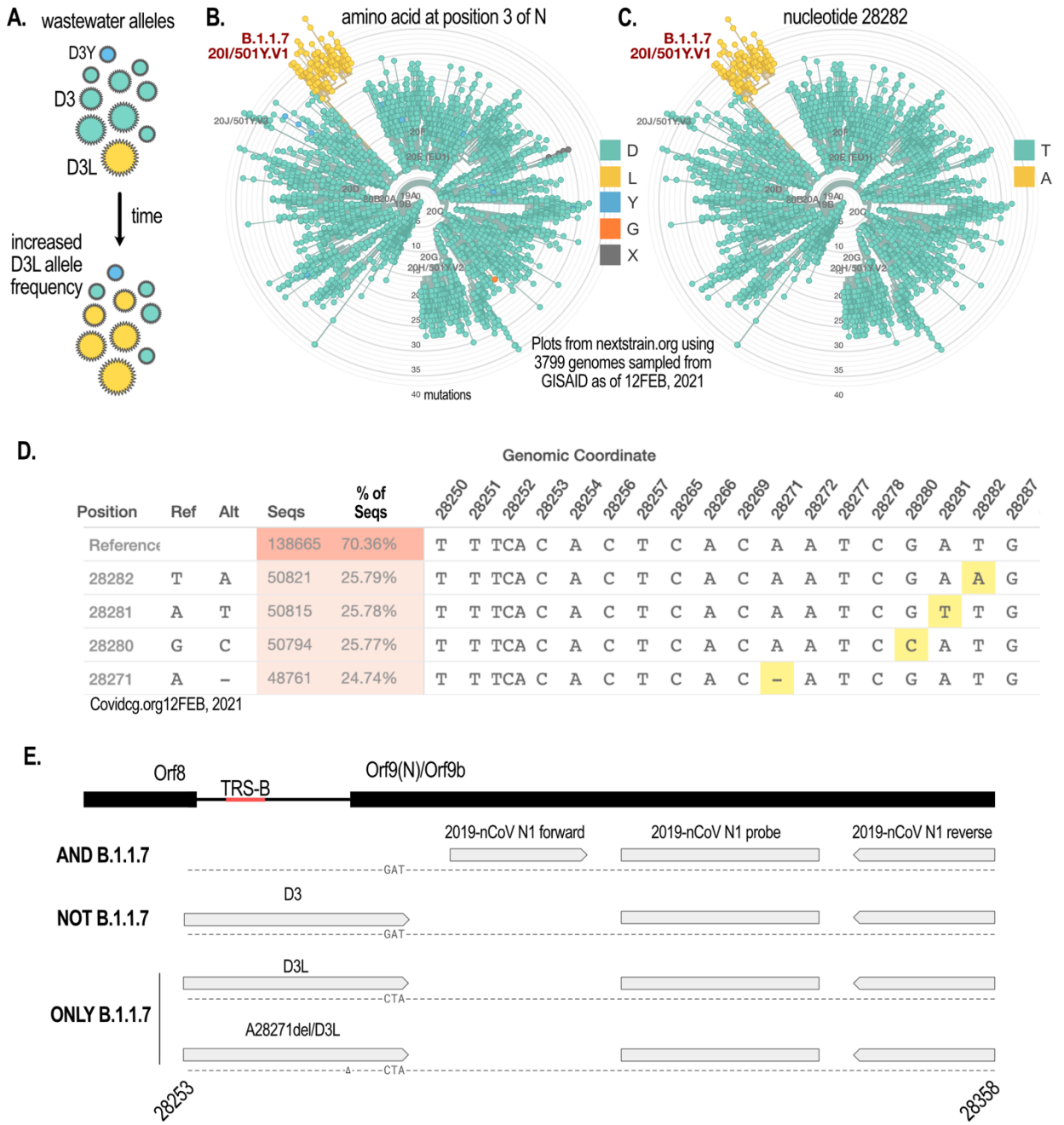


**Figure 1. The N gene D3L mutation serves as a signature target to detect VOC of the B.1.1.7 lineage.** (**A**) D3L allele frequency is hypothesized to rapidly increase in wastewater samples from communities in which B.1.1.7 has been introduced. Other mutations at the same genomic locus which are at low prevalence in a community would not be detected and/or would not change over time. (**B**) A radial plot of mutations amongst globally sampled GISAID SARS-CoV-2 sequences at the N:D3 locus. The D3L mutation is a signature of the B.1.1.7 lineage. (**C**) A radial plot as in **B** but at the nucleotide level showing that the T28282A SNV is unique to B.1.1.7. (**D**) Allele frequency table of all GISAID genomes sequenced from UK cases. Note the A28271del SNV which co-occurs with D3L in the vast majority of genomes classified as B.1.1.7. (**E**) Genomic neighbourhood and primer/probe locations for the 3 different assays discussed in the text: N1 which amplifies 99+% of viral variants (AND B.1.1.7), D3 which amplifies alleles other than the B.1.1.7-specific allele (NOT B.1.1.7), and D3L or A28271del/D3L which target B.1.1.7-specific alleles. TRS-B=Transcriptional regulatory site-body. Radial plots were generated on nextstrain.org[13] with sequencing data from GISAID[14]. Allele frequency table generated at covidcg.org using GISAID sequencing data.

substantial power to probe and follow the dynamics of allele frequencies in a population as infections wax and wane. Additionally, following a single, signature mutation in this context gives sufficient power to distinguish VOC lineages, as other variants that are less transmissible, and which have a very low prevalence in communities, might carry a mutation at the same locus but whose frequency and prevalence does not change substantially over time (**Figure 1A**).

We found that one such signature mutation on the N gene at the amino acid level (D3L) is present in the B.1.1.7 lineage (**Figure 1B**). Moreover, a single nucleotide variant (SNV) in the wobble position of N:D3, T28282A appears to be unique to the B.1.1.7 lineage (**Figure 1C**). Serendipitously, D3L lies only 5 bp upstream from the start of the CDC 2019-nCoV N1 amplicon region, which we currently target to detect SARS-CoV-2 in wastewater[15,16]. Taking advantage of this, we designed forward primers with mutations at the 3' end complementary to the B.1.1.7 signature allele at this locus. This allele-specific primer extension strategy works by virtue of the inability of DNA polymerase to extend mismatched nucleotides at the 3' end of the primer[17]. This method affords high allele specificity, while at the same time obviating the complexities associated with the development of competition probe assays. We then combined this forward primer with the existing N1 probe and reverse primer. We call these two new assays "D3" and "D3L" and it distinguishes non-B.1.1.7 from B.1.1.7 lineages when applied in the wastewater context. Note that this assay may not have a high positive predictive value in clinical samples for viruses of the B.1.1.7 lineage as a rare mutation may occur at this location in individual cases. Part way into designing and testing this new assay, we noted the emergence of a SNV in B.1.1.7 sequences deposited at GISAID that co-occurs with the D3L mutation and is located in the middle of the D3L forward primer. This deletion mutant (A28271del) now appears to be the dominant SNV in the B.1.1.7 lineage, co-occurring with D3L at similar frequencies in the United Kingdom (**Figure 1D**). Therefore, we designed an additional forward primer (A28271del/D3L) incorporating this SNV. The primer locations are shown in **Figure 1E**.

***The D3 and D3L assays have high specificity on IVT RNA templates.***

We in vitro transcribed (IVT) RNA template to determine the allelic specificity of the D3 and D3L assays. We synthesized the dsDNA with sequence complementary to either the D3 or D3L amplicons.

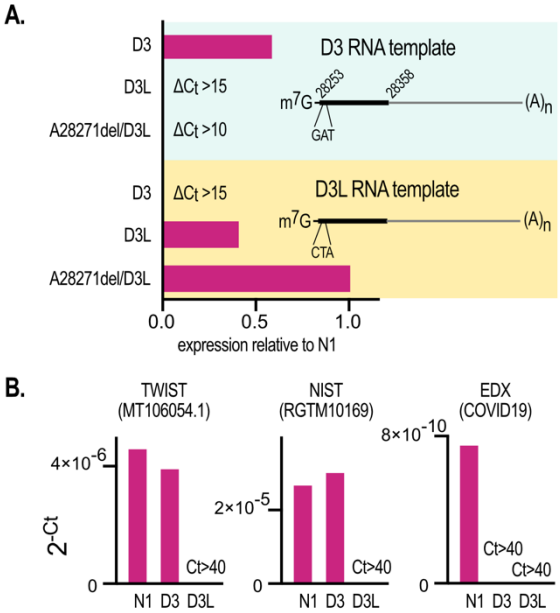


**Figure 2. The D3, D3L, and A28271del/D3L RT-qPCR assays sufficiently discriminate targeted alleles on IVT RNA templates.** (**A**) RT-qPCR assays targeting indicated alleles on 2 different IVT RNA templates. Expression is normalized to the "classical" CDC 2019-nCoV N1 assay performed in parallel. The N1 amplicon is nested within the D3, D3L, A28271del/D3L amplicons. SARS-CoV-2 genomic coordinates (Genbank NC_045512v2) of this region and the 3-nucleotide mutation at the 3rd codon of N which is specific to the B.1.1.7 lineage are shown. The RNAs have additional downstream sequence not relevant to this study and were capped and polyadenylated. (**B**) Allele-specific RT-qPCR showing expression on commercially available RNA templates (standards). Note that the EDX(COVID19) standard, unlike the other two shown, does not contain the intergenic sequence that is complementary to D3 forward primers and thus no amplification is observed.

Note that this sequence also contains the nested N1 amplicon. An additional ~250nt of downstream sequence was incorporated to include amplicons corresponding to other PCR standards used in the laboratory. A T7 promoter sequence was incorporated at the 5' end of the sequence allowing in vitro transcription of a ~350 bp RNA (**Figure 2A**). We performed D3, D3L, and A28271del/D3L assays on either the D3 or D3L RNA templates, and, in parallel performed the N1 assay whose expression allowed normalization of

copies of each RNA standard. Critically, the D3 assay showed positivity on the D3 RNA template, albeit at a lower sensitivity than N1; while both D3L and 28271del/D3L assays amplified at least 10 Ct's higher (**Figure 2A**). Similarly, both the D3L and A28271del/D3L assays showed positivity on the D3L template, while the D3 assay amplified more than 15 cycles later on the same template. This demonstrates high template specificity allowing quantification of the relative proportion of alleles in the sample with high confidence. Additionally, we found that the D3L assay failed to amplify on 3 commercially available synthetic RNA templates (**Figure 2B**), while the "wildtype" D3 allele was detected at similar levels as the N1 target. Notably, D3 failed to detect on the EDX standard, which lacks the *Orf8-Orf9* intergenic sequence that is complementary to the D3 forward primer. These data indicate high allele specificity for distinguishing D3 from D3L. We are working to also determine the specificity of D3L vs. A28271del/D3L assays, although we expect this will be less specific, with less Ct differential as the deletion occurs in the middle of the forward primer.

***The D3 and D3L assays have high specificity in patient stool samples and sensitivity decreases with annealing temperature.***

We wished to confirm the specificity and sensitivity of D3 and D3L assays using fecal RNA donated by a COVID-19 patient. The patient in question was admitted to the ICU in 2020, prior to the emergence of B.1.1.7 viruses in Canada. Confirming the specificity of the assays, we did not see amplification with the D3L assay in this sample, while it was seen with the D3 assay. We performed a comparison between N1 and D3 assays at different annealing temperatures to optimize sensitivity (**Figure 3**). At 55ºC, the D3 assay appeared to be 2-3X less sensitive than the N1 assay. Surprisingly, increasing annealing temperature to 60ºC enhanced sensitivity of the D3 assay, bringing it closer to that of N1.

***The D3L assay detects B.1.1.7 viral fragments in influent and primary sludge from a treatment plant serving a community with an isolated VOC outbreak.***

We performed N1 as well as the allele-specific assays on RNA extracted from 24h composites of primary sludge obtained from a wastewater resource recovery facility (WRRF) serving a large

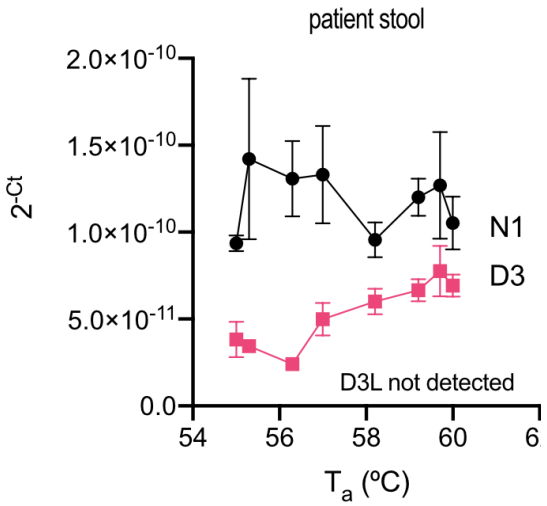


**Figure 3. The D3 assay has lower sensitivity than N1 at lower annealing temperature but high allelic specificity in RNA extracted from COVID-19 patient stool.** Stool from a single patient was collected prior to the emergence of B.1.1.7 in Canada. D3L, the B.1.1.7-specific allele was not detected. Lower annealing temperatures reduced sensitivity of the D3 assay relative to the N1 assay. Mean of RT-PCR replicates (n=3) presented with error bars representing s.d.

Canadian urban centre (Ottawa, Ontario; pop. ~1M) between Fall 2020 and January 16, 2021 (n=4). We performed that assay at 2 distinct annealing temperatures, however both D3L and the A28271del/D3L assays failed to detect B.1.1.7-specific RNA fragments (**Figure 4A**). Fewer than 5 clinically verified and travel-related B.1.1.7 cases were active in this community at the time of the last sampling of January 16, 2021. More surprisingly, we found a consistent decrease in sensitivity with increasing annealing temperature – the opposite of what was observed when RNA isolated from the faeces of one patient (**Figure 3**). Although the explanation is unclear, a possible reason may be reduced template specificity in a wastewater matrix at lower annealing temperatures.

We next obtained 24h composite influent and grab primary sludge samples from a WRRF serving the mid-size (pop. ~150K) community of Barrie, Ontario. We readily detected N1 amplicons in both samples, indicating that SARS-CoV-2 was prevalent in the community. Two weeks prior to sampling, an outbreak of B.1.1.7 infection was confirmed at a long-term care facility (LTCF) within the WRRF catchment. More than 100 cases were later confirmed by whole

genome sequencing of clinical samples at the provincial laboratories. Importantly, while the D3L and A28271del/D3L assays yielded non-detects in samples from Ottawa, with either no or very few known cases of B.1.1.7 (**Figure 4A, B**); Barrie, with a known active B.1.1.7 outbreak, yielded a readily detectable signal with all assays and in both sample types (**Figure 4B**). Surprisingly, D3, targeting non-B.1.1.7 was at very low levels compared to both D3L and A28271D3L alleles, suggesting that B.1.1.7 was the dominant viral lineage in the community at the time of sampling. Furthermore, we saw >2-fold greater signal with A28271D3L than D3L suggesting that this is the major allele in the faecally-shed viral RNA population. It's possible that unequal primer efficiencies may lead to erroneous conclusions regarding the relative proportion of each allele. We are thus currently defining the primer efficiencies for these assays.

Together, these data support the use of allele-specific qPCR strategies to follow the emergence and proportional changes of VOCs in populations served by WRRFs. Allele-specific RT-qPCR assays provide sufficiently high sensitivity and specificity in the wastewater context and can be rapidly designed and implemented within existing WBE surveillance architectures.

## Materials and methods

### *Water resource recovery facility description and sample collection*

24-hour composite samples of primary clarified sludge (PCS) were collected from the Ottawa, Ontario WRRF with no known community transmission of B.1.1.7. 24-hour composite samples of post-grit influent and grab samples of PCS were collected from the Barrie, Ontario WRRF during an isolated B.1.1.7 outbreak. Both sampled WRRFs are designed and operated as conventional activated sludge treatment systems.

### *Sample concentration and extraction*

Influent samples were first decanted at 4°C for one hour and for each sample, 40 mL of the settled solids were pipetted to a 50 mL centrifuge tube. Similarly, for the PCS samples, 40 mL of well-mixed PCS was also pipetted to a 50 mL centrifuge tube. Samples were then centrifuged at 10,000 x g for 45 minutes at 4°C. The supernatant was decanted, taking care not to disturb the resulting pellet. Samples were then centrifuged a second time at 10,000 x g for another 5 minutes and the remaining supernatant decanted. The resulting influent solids and PGS pellets were transferred to a new, pre-weighed

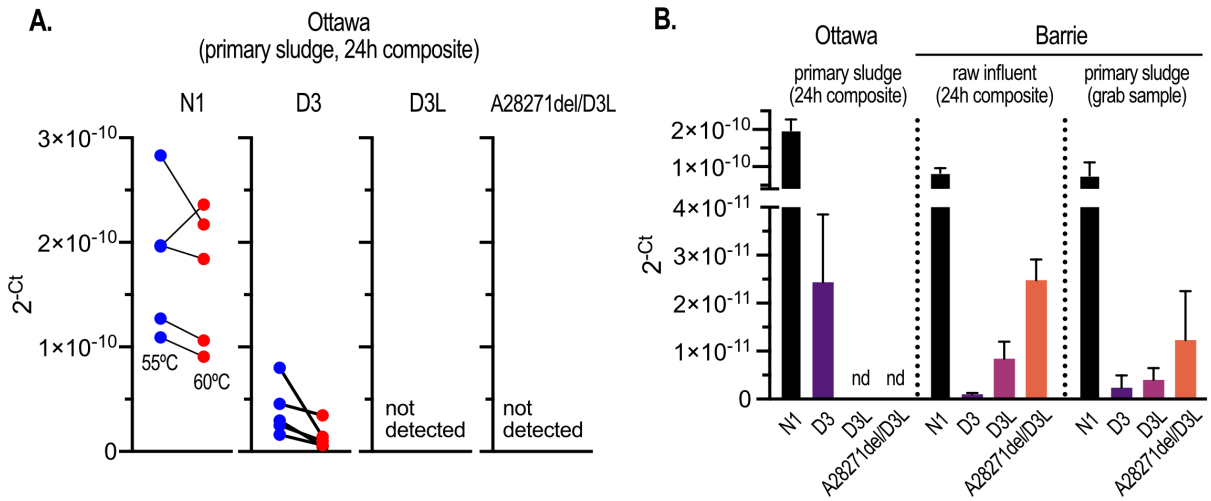


**Figure 4. Detection of B.1.1.7 alleles in wastewater samples following a confirmed community outbreak.** (**A**) Expression of alleles determined by RT-qPCR. Indicated assays were performed on sludge samples from a WRRF in a large community (Ottawa; pop. 1M) with no known outbreak of B.1.1.7-related COVID-19. Four different sampling dates are shown performed at 2 different annealing temperatures. Note that B.1.1.7-specific alleles were below the limit of detection (Ct>40). (**B**) Indicated assays were performed on raw influent solids or primary sludge samples collected from WRRFs in i) Ottawa on January 16, 2021 and with <5 known and active B.1.1.7 cases, and ii) Barrie, Ontario on January 26, 2021 and experiencing an outbreak of B.1.1.7 COVID-19 (approx. 100-200 cases). Mean of RT-PCR replicates (n=3) presented with error bars representing s.d.

RNase-free centrifuge tube before being weighed and immediately extracted.

Viral RNA was then extracted from the resulting pellets of influent and PCS solids using the RNeasy PowerMicrobiome Kit (Qiagen, Germantown, MD). The following changes to the manufacturer's protocol were performed: 1) 250 ± 15 mg of sample pellet was added to the initial extraction step in place of 200 µL of liquid sample, and 2) the optional phenol-chloroform step was instead performed with Trizol LS reagent (ThermoFisher, Ottawa, Canada) to maximize RNA protection and lysis prior to vortexing and centrifugation. The resulting aqueous phase was retained and processed as per the original manufacturer's protocol, including the DNase treatment step. RNA was eluted in 100 µl of RNase-free water.

***RT-qPCR conditions***

One-step RT-qPCR was performed with TaqMan® Fast Virus 1-Step Master Mix (ThermoFisher, Ottawa, Canada). N1 assay measurements, detecting SARS-CoV-2 RNA including B.1.1.7 lineages, were performed using the premixed 2019-nCoV N1 Assay-RUO probe/primers set (500/125nM, respectively) (IDT, Kanata, Canada). Allele-specific RT-qPCR was performed in parallel using either D3 (detecting non-B.1.1.7 alleles), D3L, or A28272del/D3L (both detecting B.1.1.7 alleles only) combined with the N1 probe and reverse primers at 500, 125 and 500nM, respectively. Sequences of the probes/primers are list in **Table 1**.

**Table 1: List of PCR primer and probe sets.**

| Primer/probe & supplier | Sequence | Reference |
|---|---|---|
| 2019-nCoV_N1 forward primer (IDT) | GAC CCC AAA ATC AGC GAA AT | (CDC, 2020) |
| 2019-nCoV_N1 reverse primer (IDT) | TCT GGT TAC TGC CAG TTG AAT CTG | (CDC, 2020) |
| 2019-nCoV_N1 probe (IDT) | 6-FAM-ACC CCG CAT/ZEN/ TAC GTT TGG TGG ACC-IOWA BLACK FQ | (CDC, 2020) |
| D3 (forward primer) | CAT CTA AAC GAA CAA ACT AAA ATG TCT GAT | *This study* |
| D3L (forward primer) | CAT CTA AAC GAA CAA ACT AAA ATG TCT CTA | *This study* |
| A28271del/D3 (forward primer) | CAT CTA AAC GAA CAA ACT AAA TGT CTC TAA | *This study* |

RT-qPCR reactions were run in triplicate using 1.5 µl of RNA input in a final reaction volume of 10 µl. Cycling was performed on a CFX Connect qPCR thermocycler (Bio-Rad, Hercules, CA) as follows: RT at 50ºC, 5 minutes, followed by RT inactivation, polymerase activation and template denaturation at 95ºC for 20 seconds, and 45 cycles of denaturation (95ºC/3s), then annealing/extension (55ºC/45s). No template controls showed amplification. Reactions were considered positive when Ct<40.

**Declaration of competing interests**

The authors declare that no known competing financial interests or personal relationships could appear to influence the work reported in this manuscript.

**Acknowledgements**

The authors wish to acknowledge the leadership of Dr. Alex MacKenzie (CHEO-RI) in this continuing endeavour as well as the participation of the City of Ottawa, Ottawa Public Health, the City of Barrie, the Simcoe Muskoka District Public Health Unit, Public Health Ontario and the Ministry of the Environment, Conservation and Parks, and their respective employees involved in the project during this study. Their time, facilities, resources, and assistance provided throughout the study greatly contributed to this work. We gratefully acknowledge the originating and submitting laboratories responsible for sharing genetic sequence data via the GISAID Initiative, which was used in this project.

**References**

1. Pachetti, M. *et al.* Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Translational Medicine* **18**, (2020).

2. Duffy, S. Why are RNA virus mutation rates so damn high? *PLoS Biology* **16**, (2018).

3. Holmes, E. G. & Rambaut, A. Viral evolution and the emergence of SARS coronavirus. in *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 359 1059–1065 (Royal Society, 2004).

4. Kustin, T. & Stern, A. Biased Mutation and Selection in RNA Viruses. *Molecular Biology and Evolution* **38**, 575–588 (2021).

5. Sanjuán, R., Nebot, M. R., Chirico, N., Mansky, L. M. & Belshaw, R. Viral Mutation Rates. *Journal of Virology* **84**, 9733–9748 (2010).

6. Grubaugh, N. D., Petrone, M. E. & Holmes, E. C. We shouldn't worry when a virus mutates during disease outbreaks. *Nature Microbiology* vol. 5 529–530 (2020).

7. Chen, J., Wang, R., Wang, M. & Wei, G. W. Mutations Strengthened SARS-CoV-2 Infectivity. *Journal of Molecular Biology* **432**, 5212–5226 (2020).

8. Korber, B. *et al.* Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812-827.e19 (2020).

9. Gaebler, C. *et al.* Evolution of antibody immunity to SARS-CoV-2. *Nature* (2021) doi:10.1038/s41586-021-03207-w.

10. Wang, Z. *et al.* mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* (2021) doi:10.1038/s41586-021-03324-6.

11. Hamouda, M., Mustafa, F., Maraqa, M., Rizvi, T. & Aly Hassan, A. Wastewater surveillance for SARS-CoV-2: Lessons learnt from recent studies to define future applications. *Science of the Total Environment* vol. 759 (2021).

12. Vogels, C. B. F. *et al.* Analytical sensitivity and efficiency comparisons of SARS-CoV-2 RT–qPCR primer–probe sets. *Nature Microbiology* **5**, 1299–1305 (2020).

13. Hadfield, J. *et al.* NextStrain: Real-time tracking of pathogen evolution. *Bioinformatics* **34**, 4121–4123 (2018).

14. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Global Challenges* **1**, 33–46 (2017).

15. 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 Research* **188**, (2021).

16. D'Aoust, P. M. *et al.* Catching a resurgence: Increase in SARS-CoV-2 viral RNA identified in wastewater 48 h before COVID-19 clinical tests and 96 h before hospitalizations. *Science of the Total Environment* **770**, (2021).

17. Lefever, S., Pattyn, F., Hellemans, J. & Vandesompele, J. Single-nucleotide polymorphisms and other mismatches reduce performance of quantitative PCR assays. *Clinical chemistry* **59**, 1470–1480 (2013).