Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Short Communication

Whole genome sequencing of SARS-CoV-2 from wastewater links to individual cases in catchments

Avram Levy, Jake Gazeley, Terence Lee, Andrew Jardine, Cameron Gordon, Natalie Cooper, Richard Theobald, Clare Huppatz, Sandra Sjollema, Meredith Hodge, David Speers

Department of Microbiology, PathWest Laboratory Medicine WA, Queen Elizabeth II Medical Centre, Nedlands, Western Australia, Australia
Public Health Emergency Operations Centre, Health Department of Western Australia, Western Australia, Australia
Water Corporation, Western Australia, Australia
Environmental Health Directorate, Health Department of Western Australia, Western Australia, Australia
School of Biomedical Sciences, University of Western Australia, Crawley, Western Australia, Australia
School of Medicine, University of Western Australia, Crawley, Western Australia, Australia

HIGHLIGHTS

• Understanding the resolution of wastewater SARS-CoV-2 genomics has been hampered by rapid increases in case numbers.
• SARS-CoV-2 wastewater sequences were compared to COVID-19 patient sequences when case numbers were low and cases were known.
• Wastewater COVID-19 sequences could be linked to individual active cases in a catchment size of 250,000 population.
• Potential contribution cases can be identified using genomic profiles in wastewater given enough nucleotide differences.
• Wastewater genomic epidemiology can provide very high resolution for informing outbreak management and disease surveillance.

ABSTRACT

After a limited first wave of community transmission in March 2020 and until 2022, Western Australia was largely free of COVID-19, with cases restricted to hotel quarantine, commercial vessels, and small, infrequent community clusters. Despite the low case load setting, sequencing of wastewater samples from large municipal treatment plants produced SARS-CoV-2 genomes with coverage up to 99.7% and depth to 4000×, which was sufficient to link wastewater sequences to those of active cases in the catchment at the time. This study demonstrates that ≤5 positive individuals can be enough to produce high genomic coverage (>90%) assemblies even in catchments of up to a quarter of a million people. Genomic analysis of wastewater contemporaneous with clinical cases can also be used to rule out transmission between cases in different catchments, when their SARS-CoV-2 genomes have distinguishing nucleotide polymorphisms. These findings reveal a greater potential of wastewater WGS to inform outbreak management and disease surveillance than previously recognized.
1. Introduction

In patients with COVID-19, viral shedding in stool is common and can persist after SARS-CoV-2 is no longer detectable in respiratory samples (Foladori et al., 2020; Kitajima et al., 2020; Jones et al., 2020). Diarrhea in COVID-19 patients occurs, with incidence varying from 3.8% to 80% and virus titre in stool also varying considerably (10^2 to 10^6 RNA copies/g of stool) (Foladori et al., 2020; Kitajima et al., 2020; Jones et al., 2020). In addition to shedding via stool, wastewater (WW) captures nasopharyngeal secretions and urine, which may also contribute to SARS-CoV-2 RNA levels (Jones et al., 2020; Peng et al., 2020).

Early in the COVID-19 pandemic wastewater-based epidemiology was implemented in several countries to track introduction and spread of SARS-CoV-2, primarily using RT-PCR (Ahmed et al., 2020a; Chavarria-Miró et al., 2021; La Rosa et al., 2020; Medema et al., 2020; Nemudryi et al., 2020; Randazzo et al., 2020; Wu et al., 2020; Rimoldi et al., 2020; Wurtzer et al., 2020). In Australia a network involving laboratories, wastewater utilities and public health units was established in 2020 to implement SARS-CoV-2 wastewater surveillance across the nation and integrate the results with health data. During limited community transmission, the goal was to use SARS-CoV-2 signals from WW to provide early warning of the introduction of SARS-CoV-2 into catchments with no known clinical cases, to use sub-catchment sampling to direct public health resources and community messaging to the likely locations of cases, and to monitor for clearance of virus from catchments to support the easing of restrictions post-outbreak.

The use of genomic sequencing for SARS-CoV-2 wastewater surveillance has been described in regions with considerable SARS-CoV-2 community transmission (Bar-Or et al., 2021; Crits-Christoph et al., 2021; Izquierdo-Lara et al., 2021; Fontenele et al., 2021; Jahn et al., 2021). Sequences from wastewater were compared to lineages circulating in the community at the time (Bar-Or et al., 2021; Crits-Christoph et al., 2021; Izquierdo-Lara et al., 2021; Fontenele et al., 2021). Wastewater genomics was used to detect single nucleotide polymorphisms (SNPs) and low-frequency variants not reported in clinical cases residing within the region in order to provide an early warning of emerging new variants (Crits-Christoph et al., 2021; Izquierdo-Lara et al., 2021; Fontenele et al., 2021; Jahn et al., 2021). Where clinical testing was minimal, wastewater SARS-CoV-2 WGS identified the proliferation of variant of concern sequences in communities (Bar-Or et al., 2021).

Western Australia (WA) is the largest Australian jurisdiction covering approximately 2.5 million km². The population of 2.7 million reside mostly in the geographically isolated Southwest corner. Oceans, remote land borders and a single arrival point for commercial international flights facilitated a considerable buffer to undetected COVID-19 introductions. The first COVID-19 case in Western Australia (WA) was notified on the 21st of February 2020, with case numbers increasing before widespread restrictions took effect (Spatial Services Unit EB et al., n.d.). Strict regulations on movements, closure of non-essential businesses, closed state and international borders to non-residents and mandatory supervised 14-day hotel quarantine of returning residents ended community transmission in WA within several months. As of the 11th of November 2021, 1110 confirmed COVID-19 cases had been reported in WA with 89% of cases acquired in another country or while at sea, 1% acquired in another Australian jurisdiction, and the remaining 10% of infections acquired from a known close contact or unknown source of infection within WA (Spatial Services Unit EB et al., n.d.). Samples from all COVID-19 cases in WA were referred for WGS (prior to sustained local transmission in 2022) to determine SARS-CoV-2 lineages and variants of concern. Genomic analyses were used to identify, confirm or disprove transmission clusters in hotel quarantine, in the community or other settings. Weekly monitoring of SARS-CoV-2 at five metropolitan WW treatment plants commenced in November 2020, and then expanded early in 2021 to include a total of eighteen treatment plants spanning all geopolitical subdivisions with results subsequently made publicly available each week. Only one catchment regularly returned positive results, the catchment encompassing quarantine hotels where active COVID-19 cases are isolated. Targeted sub-catchment sampling has also been initiated on an ad hoc basis in response to small community clusters.

The aim of this study was to provide insights into the sensitivity and utility of WW genomic epidemiology for tracking and monitoring SARS-CoV-2 spread. Established and sensitive methods for generating whole genomes from low viral load clinical cases were applied to WW samples at a time when the population caseload was extremely low (<10 cases in catchments), clinical genomes were available for most cases and the catchments to which the cases contributed were known. Sensitivity of WW genomic epidemiology was assessed, and WW genomic data was compared to the active clinical cases in the catchments at the time.

2. Methods

In Western Australia all samples from clinical COVID-19 cases are submitted for WGS as part of the public health response, unless viral load is deemed insufficient. Wastewater sampling also occurs as part of the public health response to the COVID-19 pandemic with five municipal treatment plants ranging in catchment size from approximately 90,000 to 740,000 sampled weekly from April 2020. In 2021 the program was expanded to include a total of eighteen metropolitan and regional sites. Composite samples were collected using ISCO auto-sampler devices, flow paced and continual with 400 mL sampled each hour from six metropolitan wastewater treatment plants. 250 mL aliquots were collected and shipped to the laboratory on ice, arriving on the same day and tested within 24 h. All quarantine hotels are located in the Perth CBD and contribute to a wastewater treatment plant catchment with total population size of approximately 250,000. Between 1st April 2020 and 24th May 2021, a total of 485 raw wastewater samples were tested for the presence of SARS-CoV-2 throughout Western Australia. A total of 44 samples were presumptive positive or positive, all these samples were collected from the wastewater treatment plant servicing the quarantine hotels.

Wastewater samples of 50 mL were pre-treated with MgCl₂ to a final concentration of 25 mM (Ahmed et al., 2020b) then centrifuged to separate solids at 4600 rpm for 30 min at 4°C. The supernatant was concentrated via vacuum filtration onto 0.45 μm electropositive membrane filter (Ahmed et al., 2020b) in a Microcheck® beverage monitor (Pall, United States). Each filter membrane was added to 800 μL of lysis buffer in a MagMAX™ Microbiome Bead Tube (ThermoFisher, United States) and homogenized on a TissueLyser II (Qiagen, Germany) for 10 min at 10,000 rpm then centrifuged at 14,000 rpm for 2 min. RNA was purified from 400μL aliquots using the MagMAX™ Microbiome Ultra Nucleic Acid Isolation Kit (ThermoFisher, United States) on the MagMAX™ Express-96 (ThermoFisher, United States) according to manufacturer's protocol. MS2 bacteriophage was added to the sample plate as an extraction and amplification control. Final nucleic acid elution occurred in 100 μL of supplied elution buffer. Isolated nucleic acid was assayed immediately on the PerkinElmer® SARS-CoV-2 Real-time RT-PCR assay (PerkinElmer, United States) in two technical replicates using half volume of the mastermix. Nucleic acid was diluted four-fold to minimize PCR inhibition (10 μL Mastermix, 15 μL Ultra-Pure H₂O, 5 μL nucleic acid). An initial validation exercise ascertained the limit of detection of the PCR to be one copy per μL of extract (data not shown). Standard curves were not regularly used to derive quantitative results.

For WGS, cDNA synthesis occurred immediately upon completion of screening PCR using SuperScript™ VILO™ mastermix (ThermoFisher, United States). Tiled amplicons covering the full genome of SARS-CoV-2 were amplified using both the ARTIC V2 primer set modified by removing each second primer pair to achieve 1000 bp amplicons and the ARTIC V3 primer set (unmodified; 400 bp) in separate reactions in conjunction with Q5 Hot Start DNA Polymerase (New England BioLabs, https://www.waternia.com.au/project-details/264).
United States) as per recommended protocol. Each wastewater sample comprised of two to four technical replicates of the ARTIC V2 and V3 amplicons. Sequences were uploaded to the NCBI SRA database under accession number PRJNA785351.

Illumina Nextera XT (Illumina, United States) libraries were created as per manufacturer recommendations, but using half volume of reagents and unique dual indexing. Sequencing was by Illumina iSeq (Illumina, United States) using 75 bp paired-end chemistry. Run quality control and sequence analysis was performed by custom in-house pipeline and Geneious R11 (Biomatters, New Zealand). Fastq files from the replicates were analysed independently and also as merged BAM files. COVID-19 cases in hotel quarantine and in the community were identified by review of WA Health Department records. Collection dates, RT-qPCR Ct values and unique dual indexing. Sequencing was by Illumina iSeq (Illumina, per manufacturer recommendations, but using half volume of reagents and obtained from (Suo et al., 2020).

Separate 50 mL aliquots of SARS-CoV2 PCR positive wastewater samples were collected for WGS as the China CDC set (Water Research Australia, 2021) and obtained from (Suo et al., 2020).

Consensus whole genome sequences are available from GISAID (Appendix Table A). Primer and probe sequences of the Perkin Elmer kit were identified as the China CDC set (Water Research Australia, 2021) and obtained from (Suo et al., 2020).

### Table 1

| Sample | Catchment | Date collected | Catchment size | Number of active cases in catchment | N gene Ct<sup>a</sup> | ORF1ab Ct<sup>a</sup> | Total reads<sup>b</sup> | % GC<sup>b</sup> | Genome coverage<sup>b</sup> | Average depth<sup>b</sup> | Replicates |
|--------|-----------|----------------|----------------|-------------------------------------|----------------------|-----------------------|--------------------|-----------------|-----------------|-------------------|--------|
| A      | 1<sup>c</sup> | 25/05/2021 | 247,777        | 3                                   | ND                   | 36.5                  | 1,012,339          | 39.1            | 99.7            | 4095              | 2 V2   |
| B      | 1         | 15/06/2021 | 247,777        | 2                                   | ND                   | 36.5                  | 188,212            | 39.3            | 92.0            | 763               | 2 V3   |
| C      | 1         | 29/06/2021 | 247,777        | 5                                   | 35.9                 | 36.5                  | 993,278            | 39.4            | 95.7            | 4161              | 1 V2   |
| D      | 2         | 29/06/2021 | 87,483         | 1                                   | 38.7                 | 39.2                  | 66,178             | 39.9            | 54.2            | 279               | 1 V2   |
| E      | 1         | 1/07/2021  | 247,777        | 2                                   | 38.0                 | 38.3                  | 203,928            | 39.7            | 90.5            | 868.1             | 2 V2   |
| F      | 3         | 2/07/2021  | 698,299        | 3                                   | 37.60                | 39.7                  | 1220               | 39.1            | 4.9             | 1 V2              | 2 V3   |
| G      | 3f        | 3/07/2021  | 99,191         | 2                                   | 36.5                 | 37.8                  | 141,655            | 38.9            | 93.6            | 588.7             | 2 V2   |
| H      | 1         | 7/07/2021  | 247,777        | 5                                   | 37.3                 | 38.8                  | 66,178             | 39.9            | 95.7            | 4161              | 2 V3   |
| I      | 2         | 8/07/2021  | 87,483         | 2                                   | 38.7                 | ND                    | 1803               | 40.3            | 71.2            | 8                 | 2 V2   |

ND = not detected.

<sup>a</sup> Average of two replicates.

<sup>b</sup> Sum of replicates.

<sup>c</sup> Catchment 1 is the hotel quarantine catchment.

In May and June 2021 three WW samples from the quarantine hotel catchment in Perth produced SARS-CoV-2 average ORF1ab PCR Ct values of 36.5 (Table 1), which was lower than previously seen (ORF1ab mean 39.2; ORF1ab range 37.4 to 43.2), with the first two samples also negative for the N gene target (ORF1ab positive only) (nucleocapsid mean 37.8; nucleocapsid range 35.2 to 39.3). There were between two and five active cases in the catchment throughout the time of sampling. Sequencing of all three WW samples produced reads covering 99.7 %, 92.0 % and 95.7 % of the reference genome with variable depth (Table 1; samples A-C). The first two WW samples failed to react in the N gene target of the PerkinElmer SARS-CoV-2 screening assay despite average Cts of 36.5 in the ORF1ab target. Comparison of the China CDC primer and probe binding regions (Water Research Australia, 2021; Suo et al., 2020) with the WW composite genomes identified a 3 bp substitution GGG to AAC spanning across 2 amino acids causing mutations N:R202K and N:G203R at the 5′ end of the primer binding region of the N gene (Appendix Fig. B). A 3 bp (del:28896:3) deletion was identified in the 3′ end of the primer binding region of the same primer (Appendix Fig. B). The substitution and deletion were identified in 100 % of combined WW reads covering that genomic site and subsequently identified in GISAID and WA sequences of lineage B.1.1.318, including those from cases in hotel quarantine at the time of thefirst two WW samples. Subsequent testing of the WA B.1.318 lineage samples (n = 10) by the PCR kit used for wastewater confirmed the N target drop-out occurred in a concentration-dependent manner.

Of the active cases at the time of the first WW sample (Table 1; sample A), only one case was assigned the B.1.1.318 lineage. A SNP comparison showed three SNPs from the WW sample were found in all active cases. The active B.1.1.318 case (patient A) shared 30 out of 33 SNPs identified in sample A while remaining active cases only shared 3–13 SNPs with the WW sample. Although we could not rule out partial contributions by the other remaining active cases, we were able to determine that the B.1.1.318 case was contributing to the positive result. This B.1.1.318 active case had an unusual pattern of respiratory sample PCR positivity, with Cts in the range 24.5–41.8 for approximately three weeks.

---

1. https://www.nebtlabs.com.au/protocols/2019/07/02/pcr-using-q5u-hot-start-high-fidelity-dna-polymerase-occ-m5515.
2. https://www.nebtlabs.com.au/protocols/2019/07/02/pcr-using-q5u-hot-start-high-fidelity-dna-polymerase-occ-m5515.
3. https://github.com/cov-lineages/pangolin.
Two weeks later a second WW sample with the same pattern of N gene drop-out produced a WGS with 92.0% coverage and 763× depth (Table 1; sample B). This second sample coincided with a subsequent B.1.1.318 case (patient G) in hotel quarantine. This later case could be differentiated from the first B.1.1.318 case (patient A) by 17 distinct SNPs. The aligned reads of the second WW sample showed 21 SNPs to the reference sequence, all of which were present in 100% of aligned reads. Of the 21 SNPs, all were also present in the consensus genome of the second B.1.1.318 case (patient G) and 17 SNPs were present in the consensus sequence of patient A, the first B.1.1.318 case (Table 2). The first B.1.1.318 case may have still been shedding virus into the catchment but none of the 17 SNPs distinguishing this case from the subsequent case were evident in the second WW sample.

The remaining clinical cases were of other lineages and had no >7 of the 21 SNPs found in the wastewater.

One week later a third WW sample from the same catchment serviced by quarantine hotels produced average Cts of 35.9 and 36.5 in N and ORF respectively (Table 1; sample C). By this time there was a cluster of unrelated community cases in a separate catchment of WW from suburbs to the north of Perth. Several cases in hotel quarantine and the northern suburbs cluster were assigned the B.1.617.2 lineage (Delta variant), but could be distinguished from each other by a number of SNPs. This third WW sample had a total of 31 SNPs to the reference sequence and was highly similar (30 and 29 SNPs) to two cases in hotel quarantine, a family cluster that had recorded extremely strong Cts (as low as 13) from respiratory samples at the time of the

### Table 2
SNPs identified from wastewater sample A and known active cases from the same catchment at the time the wastewater sample was collected. Empty cells indicate a reference base. Green highlighted rows indicate common SNPs to WW as well as all known active cases while yellow highlighted rows indicate SNPs between 10% and 100% frequency of the mapped wastewater reads. * indicates read depth of <10×.

| Ref base | Wastewater | Known active cases |
|----------|------------|--------------------|
| Location | Sample A   | Frequency | Patient A | Patient B | Patient C | Patient D | Patient E | Patient F |
| 241      | C          | T    | 1 | T | T | T | T | T | T |
| 14408    | C          | T    | 1 | T | T | T | T | T | T |
| 23403    | A          | G    | 1 | G | G | G | G | G | G |
| 23604    | C          | A    | 1 | A | A | G | G | G | G |
| 1230     | A          | G    | 1 | G |   |   |   |   |   |
| 3037     | C          | T    | 1 | T | T | T | T | T | T |
| 3271     | T          | C    | 1 | C |   |   |   |   |   |
| 3852     | A          | T    | 1 | T |   |   |   |   |   |
| 3961     | C          | T    | 1 | T |   |   |   |   |   |
| 7798     | G          | T    | 1 | T |   |   |   |   |   |
| 9891     | C          | T    | 1 | T* | T* |   |   |   |   |
| 10116    | C          | T    | 1 | T |   |   |   |   |   |
| 11291    | G          | A    | 1 | A |   |   |   |   |   |
| 11296    | T          | G    | 1 | G | G |   |   |   |   |
| 20578    | G          | A    | 1 | A |   |   |   |   |   |
| 21846    | C          | T    | 1 | T |   |   |   |   |   |
| 23012    | G          | A    | 1 | A |   |   |   |   |   |
| 23287    | T          | C    | 1 | C |   |   |   |   |   |
| 23948    | G          | C    | 1 | C |   |   |   |   |   |
| 24382    | C          | T    | 1 | T |   |   |   |   |   |
| 25186    | G          | T    | 1 | T |   |   |   |   |   |
| 25276    | C          | A    | 1 | A |   |   |   |   |   |
| 26767    | T          | C    | 1 | C | C | C |   |   |   |
| 27597    | T          | A    | 1 | A |   |   |   |   |   |
| 28209    | G          | T    | 1 | T |   |   |   |   |   |
| 28271    | A          | G    | 1 | G |   |   |   |   |   |
| 28881    | G          | A    | 1 | A | A | T | T |   |   |
| 28882    | G          | A    | 1 | A | A |   |   |   |   |
| 28883    | G          | C    | 1 | C | C |   |   |   |   |
| 29769    | C          | T    | 1 | T |   |   |   |   |   |
| 1406     | C          | T    | 0.23 | T |   |   |   |   |   |
| 8619     | C          | T    | 0.36 | T | T |   |   |   |   |
| 28896    | C          | G    | 0.28 | G |   |   |   |   |   |
Table 3
SNPs identified from wastewater sample B and known active cases from the same catchment at the time the wastewater sample was collected. Empty cells indicate a reference base. Green highlighted rows indicate common SNPs to WW as well as all known active cases. * indicates read depth of <10 ×.

| Location | Ref base | Wastewater Sample B | Known active cases |
|----------|----------|---------------------|--------------------|
|          | Location | Frequency | Patient G | Patient H | Patient J | Patient K | Patient A |
| 241      | C        | T         | T         | T         | T         | T         | T         |
| 3037     | C        | T         | T         | T         | T         | T         | T         |
| 14408    | C        | T         | T         | T         | T         | T         | T         |
| 23403    | A        | G         | G         | G         | G         | G         | G         |
| 23604    | C        | A         | A         | G         | G         | G         | G         |
| 26767    | T        | C         | C         | C         | C         | C         | C         |
| 680      | G        | A         | A         | T         | T         | T         | T         |
| 7798     | G        | T         | T         | T         | T         | T         | T         |
| 8950     | C        | T         | T         | T         | T         | T         | T         |
| 9072     | C        | T         | T         | T         | T         | T         | T         |
| 9891     | C        | T         | T         | T         | T         | T         | T         |
| 10116    | C        | T         | T         | T         | T         | T         | T         |
| 21004    | G        | T         | T         | T         | T         | T         | T         |
| 21846    | C        | T         | T         | T         | T         | T         | T         |
| 23287    | T        | C         | C         | T         | T         | T         | T         |
| 24382    | C        | T         | T         | T         | T         | T         | T         |
| 25276    | C        | A         | A         | A         | A         | A         | A         |
| 26615    | G        | T         | T         | T         | T         | T         | T         |
| 28209    | G        | T         | T         | T         | T         | T         | T         |
| 28271    | A        | G         | G         | G         | T         | T         | T         |
| 29769    | C        | T         | T         | T         | T         | T         | T         |

WW collection. Seven SNPs present in the WW sample were found exclusively in the two cases and not in the northern suburbs catchment (Table 3).

Sub catchment sampling was initiated in response to the northern suburbs cluster, with one sub catchment of population approximately 100,000 within the very large northern suburb catchment (~700,000) being positive, consistent with the location of the known positive cases quarantining in their homes. A WW sample from this sub catchment produced WGS sequence with coverage of 93.6 % but at a relatively low depth of 588 ×. Comparison of this WW sample to sequences from the cluster active at the time showed only one heterozygous nucleotide location where the alternate base identified in 60 % of reads (G13332C) could not be accounted for by the cluster sequences.

None of the wastewater samples produced observable cytopathic effect with cell culture.

4. Discussion

This wastewater genomic epidemiology study provides several examples where it was possible to trace back and link composite SARS-CoV-2 genomes obtained from WW to individuals in hotel quarantine. Furthermore, during an episode of limited community transmission, a comparative analysis of SNP profiles in the composite genome from WW with the contemporaneous clinical sample genomic sequences ruled out the cases in hotel quarantine being the source of the community transmission observed in a separate catchment. The source of the community transmission was subsequently identified to be a returned traveller from another Australian jurisdiction. The very small number of COVID-19 cases in WA prior to 2022 (Spatial Services Unit EB et al., n.d.) along with a rigorous WGS approach has created opportunities to link detections in wastewater to corresponding known clinical cases isolating in hotels.

Due to the low prevalence of SARS-CoV-2 in WA in 2021, only the treatment plant servicing the quarantine hotels regularly returned positive RT-qPCR results. Detections of SARS-CoV-2 in WW were weak despite the highly sensitive PCR assay used, with Cts from auto-samplers at treatment plants not <36 cycles. The observation of stronger than usual Cts in May and July was therefore unprecedented in our experience, with several factors potentially contributing. Cooler ambient wastewater temperatures coinciding with the seasonal drop in temperature in Perth may have improved persistence of RNA in the WW systems. Temperature has been shown to affect SARS-CoV-2 RNA detections in spiking experiments (Wurtzer et al., 2021).

Two of the three situations where near complete composite genomes were obtained involved cases of B.1.1.318 lineage. This lineage is uncommon in Australian sequences submitted to the GISAID repository (Elbe and Buckland-Merrett, 2017), but had been circulating globally, peaking in April to May 2021 (CoV-Lineages4) and responsible for a large proportion of cases in some regions (Tegally et al., 2021). Known cases of lineage B.1.1.318 in WA were reviewed and were found to coincide with a slight drop in WW Cts at the time. It is plausible that there are virological determinants of viral shedding with higher viral load in stool associated with infections by particular lineages, however there was no published evidence to support this. The two cases of lineage B.1.1.318 which led to the stronger than usual WW signals were both partially vaccinated. It is unknown if the faecal shedding of SARS-CoV-2 is altered in post-vaccination infection.

SARS-CoV-2 RNA in WW is widely regarded as present in viral fragments and there are no reports of virus culture from wastewater despite documented efforts (Rimoldi et al., 2020; Wang et al., 2020). Our attempts were similarly unsuccessful, although the method was not developed or

https://cov-lineages.org/lineage.html?lineage=B.1.1.318.
optimized for SARS-CoV-2 in WW. The ability to distinguish between inactive but intact virion and fragmented virus has been described (Rimoldi et al., 2020; Wurtzer et al., 2021) but was not attempted in this study. It is possible that the high genomic coverage results, as well as the greater than expected PCR signal strengths suggest intact viral RNA may have been present in the WW samples. However, this theory was not specifically assessed.

Highly sensitive detection and genomic characterisation of SARS-CoV-2 relies on PCR amplification which introduces an intrinsic bias towards higher copy number template in the specimen. WW is likely to be dominated by RNA from cases with a higher copy number when creating composite genomes. Therefore, sequence diversity in composite genomes from WW is likely to underestimate the true sequence diversity in the WW, instead representing the most common fragments only. The use of multiple grab samples alongside 24-hour collections by auto or passive sampling devices in certain settings may provide additional sequence as it may increase the representation from individuals over different time points.

For two of the wastewater samples in this study, single contributing cases were identified as the most likely sources and in both instances the cases were infected with SARS-CoV-2 of lineage B.1.1.318. ORF1ab targets of the commercial SARS-CoV-2 PCR either reacted poorly or were negative with WW. WGS results identified a deletion in the forward primer binding region, which was also present in a large proportion of GISAID sequences of lineage B.1.1.318 (Elbe and Buckland-Merrett, 2017). Previous WA sequences of B.1.1.318 had the same deletion and when the samples were retrieved and tested by the WW PCR assay showed a concentration dependent signal delay and drop-out of the ORF1ab target.

Although each WW sample could be linked to specific clinical sequences through multiple unique SNPs, there was one heterozygous base identified in the WW results which did not correspond to any known active or historical cases. A broader search for sequences with the alternate base at that location perform on the GISAID and NCBI database also did not yield any results, suggesting that the SNPs from the WW samples were template misreads introduced during the PCR stage.

Wastewater-based epidemiology has become a valuable tool for screening large populations for the presence of SARS-CoV-2 viral fragments whilst individuals within the population remain anonymous. This study demonstrates the previously unrecognized ability of WW genomic surveillance to identify the individual contributors to positive WW samples by comparing the results to clinical sequence data. This is only possible in the uncommon situation where communities are free of more widespread COVID-19, and the location of all active cases in regards to WW catchment are likely known and have their genomic sequences from respiratory samples available for comparison. In communities with multiple contributors of SARS-CoV-2 to WW, all with very similar or identical genomes contributing to composite WW genomes, it has not been possible to understand the discriminatory power of genomic methods targeting WW RNA

The application of sensitive genomic approaches in combination with microbial surveillance data has the potential to provide early warnings of the presence of new and emerging pathogen variants, as well as their persistence and eradication. As a result of COVID-19, the scope of WW microbial surveillance has evolved from focusing on gastrointestinal viruses to informing public health responses to an evolving respiratory pathogen. The introduction of genomics creates a much more powerful tool for interrogation of PCR detections from WW. These technologies are likely to be applicable to an expanded range of pathogens of public health significance and for monitoring antimicrobial resistance.

CRediT authorship contribution statement

Avram Levy Conceptualization, Writing - Original Draft, Writing - Review & Editing, Resources, Project administration.
Jake Gazley Investigation, Formal analysis, Writing - Original Draft.
Terence Lee Formal analysis, Investigation, Methodology, Writing - Original Draft.
Andrew Jardine Resources, Investigation, Writing - Original Draft.
Cameron Gordon Resources, Investigation, Writing - Original Draft.
Natalie Cooper Investigation.
Richard Theobald Writing & Review & Editing.
Clare Huppatz Resources.
Sandra Sjollema Writing - Review & Editing, Resources.
Meredith Hodge Conceptualization, Writing - Original Draft, Writing - Review & Editing, Supervision.
David Speers Conceptualization, Resources, Writing - Original Draft, Writing - Review & Editing, Supervision.

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.

Acknowledgements

We appreciate the advice on cell culture and virus isolation methods provided by Dr. Chisha Sikazwe and thank the WGS scientists and COVID wastewater team at PathWest. This work was funded by WA Department of Health and PathWest Laboratory Medicine WA.

Appendix A

| Table A |
|---|---|
| List of patient and GISAID IDs. |

| Patient ID | GISAID ID |
|---|---|
| Patient A | hcov-19/Australia/WA726/2021 |
| Patient B | hcov-19/Australia/WA724/2021 |
| Patient C | hcov-19/Australia/WA700/2021 |
| Patient D | hcov-19/Australia/WA723/2021 |
| Patient E | hcov-19/Australia/WA663/2021 |
| Patient F | hcov-19/Australia/WA678/2021 |
| Patient G | hcov-19/Australia/WA727/2021 |
| Patient H | hcov-19/Australia/WA730/2021 |
| Patient I | hcov-19/Australia/WA732/2021 |
| Patient J | hcov-19/Australia/WA736/2021 |
| Patient K | hcov-19/Australia/WA733/2021 |
