

---

# Correlating Quantitative and Genomic SARS-CoV-2 Wastewater Data with Clinical Metrics in Metropolitan Perth, Western Australia

---

[Jake Gazeley](#)\*, Terence Lee, Daniel R Knight, Alexander Shivarev, Cameron Gordon, David Speers, Dylan D Barth, Jelena Maticevic, Meredith Hodge, Paul Armstrong, Paul Knight, Sandra Sjollem, Avram Levy

Posted Date: 5 March 2024

doi: 10.20944/preprints202403.0192.v1

Keywords: Wastewater-based epidemiology; SARS-CoV-2; COVID-19; surveillance system; correlation analysis; genomics



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

# Correlating Quantitative and Genomic SARS-CoV-2 Wastewater Data with Clinical Metrics in Metropolitan Perth, Western Australia

Jake Gazeley <sup>1,\*</sup>, Terence Lee <sup>1</sup>, Daniel R Knight <sup>1</sup>, Alex Shivarev <sup>2</sup>, Cameron Gordon <sup>3</sup>, David Speers <sup>1,4</sup>, Dylan D Barth <sup>2,5</sup>, Jelena Maticovic <sup>2</sup>, Meredith Hodge <sup>1</sup>, Paul Armstrong <sup>2</sup>, Paul Knight <sup>2</sup>, Sandra Sjollem <sup>1</sup> and Avram Levy <sup>1,6</sup>

<sup>1</sup> Department of Microbiology, PathWest Laboratory Medicine WA, Queen Elizabeth II Medical Centre, Nedlands, Western Australia, Australia

<sup>2</sup> Communicable Disease Control Directorate, Public and Aboriginal Health Division, Department of Health Western Australia, Western Australia, Australia

<sup>3</sup> Water Corporation, Western Australia, Australia

<sup>4</sup> School of Medicine, University of Western Australia, Crawley, Western Australia, Australia

<sup>5</sup> Wesfarmers Centre of Vaccines and Infectious Diseases, Telethon Kids Institute, Western Australia, Australia

<sup>6</sup> School of Biomedical Sciences, University of Western Australia, Crawley, Western Australia, Australia.

\* Correspondence: jake.gazeley@health.wa.gov.au

**Abstract:** Wastewater-based epidemiology (WBE) has emerged as a key method for continuous monitoring of COVID-19 prevalence including circulating SARS-CoV-2 lineages. WBE addresses limitations of traditional clinical COVID-19 surveillance such as clinical test availability, fluctuating testing rates and increased reliance on rapid antigen tests. Our study in Perth, Western Australia, found a significant positive correlation between SARS-CoV-2 concentrations in wastewater and clinical PCR positivity rates ( $r_s = 0.772$ ;  $p < 0.001$ ) over an 18-month period that included four successive COVID-19 waves. A strong positive correlation was apparent between the proportions of SARS-CoV-2 lineages in wastewater and clinical cases within the same region ( $r_s = 0.728$ ,  $p < 0.001$ ), including earlier detection of Omicron and recombinant lineages in wastewater before clinical case confirmation. The successful integration of WBE with healthcare data underscores its critical role in enhancing public health decision-making and pandemic management. This approach not only demonstrates the value of WBE in current global health surveillance efforts but also highlights the potential of WBE to address future public health challenges, as a comprehensive disease monitoring and response approach.

**Keywords:** Wastewater-based epidemiology; SARS-CoV-2; COVID-19; surveillance system; correlation analysis; genomics

## 1. Introduction

Communicable disease surveillance systems are essential in understanding the prevalence, and characterising the transmission of pathogens in the community. Traditional surveillance systems rely heavily on the continuous and systematic collection of clinical case data (clinical surveillance). This includes diagnostic test results, patient-reported symptoms, and information from healthcare providers, including hospitalisations[1]. Clinical surveillance is essential in tracking the spread of respiratory pathogens such as influenza, respiratory syncytial virus (RSV), and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). However, passive surveillance systems that rely solely on individuals seeking healthcare underrepresent the true disease prevalence in the community, particularly when symptoms are either mild or completely absent[2].

The effectiveness of clinical surveillance systems is dependent on the timing and completeness of the system end-to-end. Where prompt public health responses are critical, such as during the initial stages of an outbreak, delays in notifications, data collection, analysis and reporting can impact the

ability to undertake effective and timely public health action[3]. These challenges were highlighted during the COVID-19 pandemic[4], underlining the necessity for timely and representative surveillance systems. This is exacerbated with self-testing which bypasses communicable disease notification systems, such as the use of rapid antigen testing (RAT) on self-collect specimens for respiratory viruses[5].

Wastewater-based epidemiology (WBE) has emerged as a significant complementary addition to traditional clinical surveillance systems. WBE involves the analysis of raw wastewater to detect markers of disease at a population level, and offers insights independent of individual healthcare-seeking behaviours and medical practitioner testing practices[1]. WBE for SARS-CoV-2 evolved rapidly due to the requirement for improved surveillance driven by the COVID-19 pandemic[6]. Studies have demonstrated a positive correlation between SARS-CoV-2 viral concentrations in wastewater to associated clinical metrics in the corresponding region[7–9]. Recently, this has also been demonstrated for a range of other respiratory viruses including other seasonal coronaviruses, rhinovirus, human metapneumovirus, parainfluenza virus, RSV, and influenza [10–12]. This suggests that WBE is able to provide an understanding of transmission of diseases at the community-level that is less prone to biases, which often implicate clinical surveillance systems. Furthermore, there is emerging evidence of WBE's ability to act as an early warning system ahead of increases in clinical case numbers[13] or hospitalisations[14].

Recent advancements in next-generation sequencing and bioinformatic analysis have enabled the monitoring of intra-species viral lineages in wastewater samples[15,16]. This capability has the potential to provide a genomic epidemiological understanding of viral dynamics at a community level, permitting a better understanding of the progression and burden of communicable diseases cost-effectively. Given that wastewater matrices are a mixed sample from diverse individuals, the SARS-CoV-2 sequencing data generated from wastewater requires deconvolution into interpretable data. This process has been primarily applied to SARS-CoV-2 for assessing community-wide genomic epidemiology, utilizing bioinformatics tools such as Freyja[16] which recovers relative lineage abundances from mixed SARS-CoV-2 samples. Research has shown that these methods support the early detection of emerging variants within communities[9,17] and have also established a significant positive correlation between the variants of concern in wastewater and clinical cases in corresponding communities [18–20]. While SARS-CoV-2 lineage proportion data in wastewater has been publicly available in many jurisdictions since 2022[21–23], the temporal correlation analysis between Omicron and recombinant sub-lineages detected in wastewater and clinical cases has been limited [18,24].

This study demonstrates that WBE of SARS-CoV-2 correlates with clinical metrics such as COVID-19 case notifications and polymerase chain reaction (PCR) test positivity rates in Perth, Western Australia (WA) - a city with approximately 2,100,000 residents[25]. Our approach involves a detailed examination of the relative abundance of key Omicron and recombinant SARS-CoV-2 lineages in wastewater and their correlation to the percentage of clinical cases identified with these specific lineages in the corresponding region. The data produced was available in real-time to public health agencies and was used to inform disease prevention and control.

## 2. Materials and Methods

### *Wastewater Sample Collection*

Since July 2022, bi-weekly wastewater samples were collected from three metropolitan wastewater treatment plants (WWTPs) in Perth, WA as part of a SARS-CoV-2 wastewater surveillance program; these included Subiaco (approximate population 250,000), Woodman Point (approximate population 750,000) and Beenyup (approximate population 700,000). The three catchments represent approximately 79% of metropolitan Perth's population[25].

Sample collection involved flow-paced, continuous auto-samplers at each plant. Hourly, 400mL samples were pooled to form a 24-hour composite sample which was retained at 4°C during the sampling window. These composite samples were homogenised, and 250mL aliquots were sampled and transported on ice to the laboratory on the day of collection. Samples were stored between 2°C and 8°C until testing, which occurred within seven days of collection.

From 4 July 2022 to 31 December 2023, a total of 447 composite wastewater samples were collected from these WWTPs.

#### *Wastewater Concentration and Extraction*

Each wastewater sample was pre-treated, concentrated, and nucleic acid purified using 50mL aliquots, as per a previously described procedure[26]. Briefly, sample aliquots were pre-treated with MgCl<sub>2</sub> and centrifuged to remove solids. The supernatant was concentrated on an electronegative filter membrane and nucleic acid purification was performed on the filter membrane using the MagMAX Microbiome Ultra Nucleic Acid Isolation Kit (ThermoFisher, United States). All samples were spiked with MS2 bacteriophage, an internal process control prior to nucleic acid purification.

#### *SARS-CoV-2 Real-Time PCR*

SARS-CoV-2 PCR analysis was performed on all wastewater samples collected between 4 July 2022, and 31 December 2023 (N = 447). Quantitative real-time PCR (qPCR) was employed for SARS-CoV-2 quantitation, using the PerkinElmer SARS-CoV-2 Real-time RT-PCR assay (PerkinElmer, United States). This assay targets the nucleocapsid and ORF1ab genomic regions of SARS-CoV-2. Additionally, MS2 bacteriophage is targeted as an internal process control. Following initial verification, the volume of reagents used in each qPCR reaction was halved. Each wastewater sample, alongside negative template controls, were assayed in duplicate technical replicates. Purified SARS-CoV-2 RNA was included as a positive amplification control. Pre-established expected cycle threshold (Ct) values and their 95% confidence intervals were used as benchmarks for the positive control, with repeat batch analysis when the control results did not fall within these confidence intervals.

The limits of detection (LoD) and quantification (LoQ) were established at 5 copies/reaction (equivalent to 200 copies/50mL) and 12.5 copies/reaction (equivalent to 500 copies/50mL), respectively for both SARS-CoV-2 targets.

#### *Molecular Inhibition Assessment*

MS2 Ct values served as indicators of extraction issues and to monitor for PCR inhibition. The 95% confidence limits for these values in negative template controls were established. Samples that yielded MS2 Ct values outside these established limits were repeated. In the repeat analysis, both 5µL and 2µL aliquots of purified nucleic acid were used per reaction. Sample results with MS2 Ct values within the 95% confidence limits were considered valid. Those that consistently fell outside these limits were excluded from further quantification and the result was reported as indeterminate.

#### *Wastewater SARS-CoV-2 Viral Quantification*

Commercially acquired, quantified SARS-CoV-2 RNA (Twist Bioscience, United States) was used to generate standard curves in the PerkinElmer SARS-CoV-2 Real-time RT-PCR assay. Upon receipt, the RNA stock, transported on dry ice, was promptly stored at -80°C. For assay preparation, the RNA was thawed and serially diluted to create a range of concentrations from 10<sup>-2</sup> to 10<sup>-5</sup> (equivalent to 10 to 10,000 copies/µL). These dilutions were then dispensed into 20µL volumes, maintained on ice during the process, and subsequently stored at -80°C. Individual aliquots were thawed for each analytical batch to ensure consistency across different runs.

TriPLICATE 5µL template volumes of four serial dilution levels (50 to 50,000 copies/reaction) were used. Each PCR batch used fresh aliquots of RNA for standard curve production.

To express the sample quantitation results as genome copies per 50mL of wastewater, the concentrations deduced from the standard curve were multiplied by the concentration factor of 40. qPCR assays were executed using the CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories, United States), with quantification calculated through CFX Maestro Software (Bio-Rad Laboratories, United States). The final quantified outcome for each specimen, denoted as genome copies/50mL, represented the mean of the values from dual replicates across both the ORF1ab and

nucleocapsid targets. Results from templates diluted below 5 $\mu$ L were manually recalibrated by the corresponding dilution factor.

All samples collected after the 9<sup>th</sup> of October 2023 were quantified exclusively using the average results from the ORF1ab target due to a mutation identified in all BA.2.86 sub-lineages, which impacted the nucleocapsid assay performance (data not shown).

Results below the LoQ were considered in the correlation analysis.

#### *SARS-CoV-2 Genome Sequencing*

SARS-CoV-2 genome sequencing (GS) was performed weekly on one wastewater sample from each WWTP for the complete study period (N = 239).

Upon completion of the SARS-CoV-2 quantification, cDNA synthesis was performed on each purified nucleic acid extract using SuperScript VILO mastermix (ThermoFisher, United States) or LunaScript RT SuperMix (New England Biolabs, United States). These reverse transcription chemistries were validated to ensure comparable results (data not shown). The genome of SARS-CoV-2 was amplified using a modified ARTIC V3 primer set (400 bp) in combination with Q5 Hot Start DNA Polymerase (New England BioLabs, United States), following the manufacturers recommended protocol. Weekly sequencing included a positive control of a known lineage, a non-template control that underwent parallel nucleic acid purification and a non-template control of Ultra-Pure H<sub>2</sub>O.

Libraries for sequencing were prepared using the Illumina Nextera XT kit (Illumina, United States), adhering to the manufacturer's recommendations but with a half-volume of reagents and employing unique dual indexing. Sequencing was conducted on Illumina platforms (iSeq, 150 bp reads; MiniSeq 150 bp reads; and MiSeq, 300 bp reads) (Illumina, United States). Quality control of the sequencing runs was carried out using a custom in-house pipeline involving mapping of trimmed reads to ref strain Wuhan-Hu-1 (accession NC\_045512.2) and SARS-CoV-2 lineage assignment with Pangolin v4.3.1 (<https://github.com/cov-lineages/pangolin>) in clinical cases.

Before January 24, 2023, a custom in-house pipeline was employed for the routine reporting of SARS-CoV-2 lineage abundance in wastewater. On January 24, 2023, all historical wastewater sequencing data were reanalysed using Freyja v1.3.9[16], applying the most recent lineage-determining mutational "barcodes" available at that time which are derived from the Ultrafast Sample placement on Existing tRees (USHER) global phylogeny (<https://github.com/andersen-lab/Freyja/>). From this date, Freyja was used as detailed in Supplementary Material S1, Table S1. SARS-CoV-2 genome coverage was calculated for each sample using a 10x depth. Samples with a genome coverage of <60% were not reported or used for subsequent correlation analysis. Sub-lineages were condensed into their respective parent lineages to facilitate result presentation and data analysis as per Supplementary Material S1, Table S4.

#### *Clinical Data*

PathWest Laboratory Medicine WA (PathWest) is the sole public pathology provider in WA, offering clinical and environmental microbiology testing. For this study, we calculated the clinical PCR test positivity rate, defined as the number of positive molecular clinical tests divided by the total number of molecular tests conducted for SARS-CoV-2 at PathWest. The tests used to calculate the PCR positivity were restricted to residents from metropolitan Perth, enabling comparison with SARS-CoV-2 wastewater concentrations from the three metropolitan WWTPs.

PathWest is the sole laboratory in WA undertaking clinical whole-genome sequencing (WGS) for SARS-CoV-2. This study incorporates data from all sequenced SARS-CoV-2 cases in the metropolitan Perth metropolitan area throughout the duration of the analysis including the assigned Pangolin designation[27]. Sub-lineages designated by Pangolin were condensed into their respective parent lineages to facilitate result presentation and data analysis as per Supplementary Material S1, Table S5.

In WA, COVID-19 is a notifiable infectious disease under the *Public Health Act 2016*[28], mandating all pathology laboratories and medical or nursing practitioners to report any detections

of COVID-19 to the WA Department of Health. From 7 February 2022, under the provisions of the *Emergency Management Act 2005*[29], individuals with a positive RAT for SARS-CoV-2 were mandated to self-report their results to the WA Department of Health via an online portal. The portal to report RATs to the WA Department of Health ceased operations on 9 October 2023. This study utilises notification data of only PCR-confirmed COVID-19 clinical cases with a residential address in metropolitan Perth, notified to the WA Department of Health to allow for comparison with wastewater concentration levels from the three metropolitan WWTPs.

#### *Data Processing, Correlation and Statistical Significance.*

Each week, the average SARS-CoV-2 concentrations in wastewater were calculated by aggregating quantified results from samples across all three WWTPs, excluding indeterminate results. Concurrently, the corresponding rate of clinical PCR positivity and the total number of weekly PCR-positive clinical notifications were collated (Supplementary Material S2, Table S1).

The weekly relative abundance of SARS-CoV-2 parent lineages in both clinical and wastewater samples was determined (Supplementary Material S2, Table S2 and Table S3). Information on the number of wastewater samples used each week for calculating the average concentrations and relative lineage abundances is available in Supplementary Material S2, Table S4. The quarterly number of clinical cases undergoing WGS decreased during the study period (Table 1).

**Table 1.** The number of clinical SARS-CoV-2 sequences that were assigned a lineage and the number of wastewater samples successfully sequenced (genome coverage >60%). The associated Spearman correlation rank ( $r_s$ ) and statistical significance ( $p$ ) between clinical and wastewater proportions of all parent lineages from week ending in Q3 2022 to Q4 2023.

Quarter-Year	Number of Clinical Cases Sequenced	Number of Wastewater Samples Sequenced	$r_s$	$p$
Q3-2022	1411	43	0.806	<0.001
Q4-2022	1737	39	0.747	<0.001
Q1-2023	1613	26	0.773	<0.001
Q2-2023	1736	38	0.703	<0.001
Q3-2023	708	36	0.771	<0.001
Q4-2023	210	40	0.670	<0.001

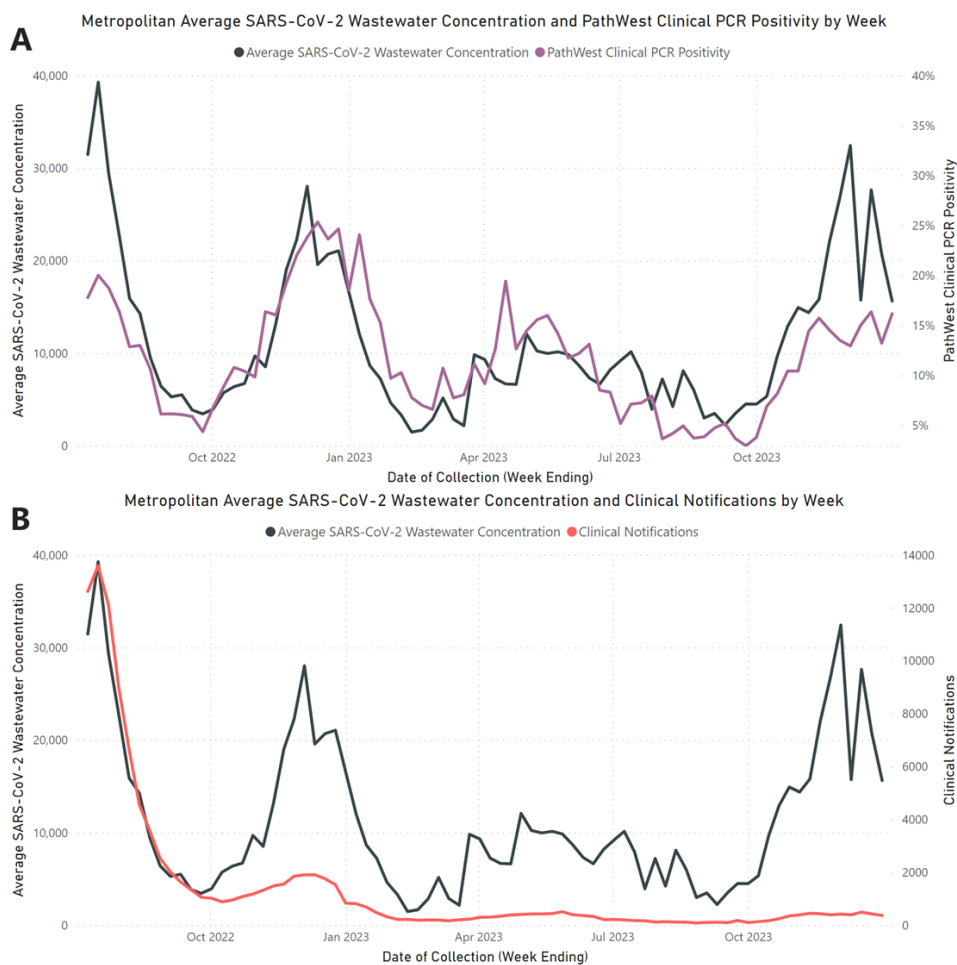
To analyse the relationship between the wastewater and clinical datasets, Spearman's rank correlation coefficient ( $r_s$ ) was employed. Additionally, the statistical significance of the correlation was determined ( $p$ ) [30].

### 3. Results

#### *Quantitative Correlation of Respiratory Clinical Metrics to Wastewater Concentrations*

From 4 July 2022 to 31 December 2023, 447 raw wastewater samples were analysed for SARS-CoV-2 using qPCR, with a detection rate of 100%.

Throughout the study period, analysis of weekly average wastewater concentrations from the three metropolitan WWTPs identified four distinct waves of increased SARS-CoV-2 RNA levels, aligning with peaks in clinical counts (Figure 1). These wastewater waves peaked on the following dates: Wave 1 on 17 July 2022; Wave 2 on 4 December 2022; Wave 3 on 30 April 2023; and Wave 4 on 3 December 2023. A strong and statistically significant positive correlation was established between the metropolitan average weekly SARS-CoV-2 wastewater RNA concentrations and clinical PCR positivity rates reported by PathWest ( $r_s = 0.772$ ;  $p < 0.001$ ). A significant but weaker positive correlation was noted between average weekly wastewater RNA concentrations and metropolitan PCR-confirmed COVID-19 case notifications ( $r_s = 0.577$ ;  $p < 0.001$ ). Detailed values of the weekly average SARS-CoV-2 wastewater concentrations, PathWest clinical PCR positivity rates, and metropolitan PCR confirmed clinical notifications are documented in Supplementary Material S2, Table S1.



**Figure 1.** Assessment of SARS-CoV-2 Trends in Metropolitan Perth Wastewater. (A) Presents a comparative view of the weekly average SARS-CoV-2 concentrations in wastewater (black) against the PathWest clinical positivity rate per week (purple). (B) Presents a comparative view of the weekly average SARS-CoV-2 concentrations in wastewater (black) against weekly PCR-confirmed metropolitan notifications of SARS-CoV-2 (red).

A total of 11 out of 447 samples (2.46%) did not yield quantifiable results as their MS2 Ct values were outside the 95% confidence limits. Seven samples from Subiaco WWTP exhibited MS2 Ct values lower than anticipated, suggesting an intrinsic presence of MS2 that surpassed the spiked reference. Three samples from Woodman Point WWTP consistently showed high Ct values that were not resolved by dilution. One sample had a quantification level below the established LoQ, with less than 500 genome copies per 50mL.

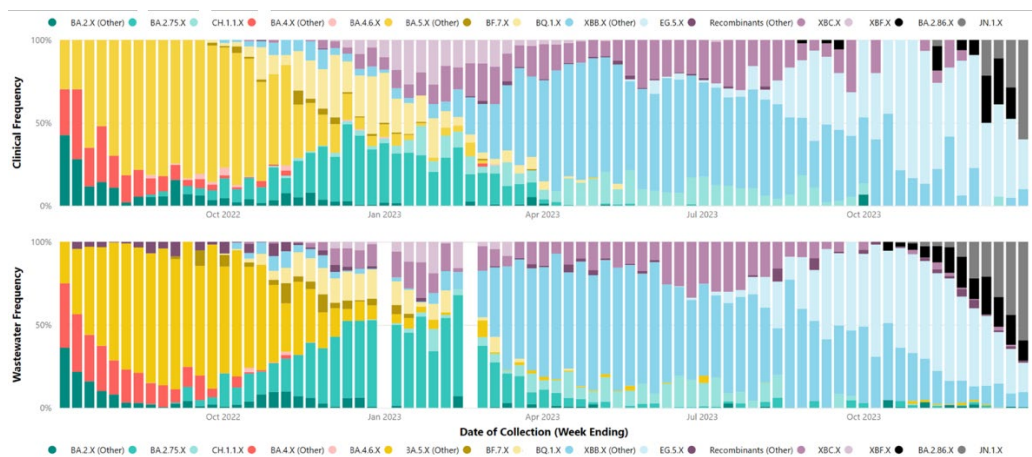
#### *SARS-CoV-2 Wastewater Genomic Coverage and Correlation to Quantitation*

The average genome coverage across the 239 wastewater samples that underwent GS was 87.01% (median 91.30%, range 39.40% to 99.31%). Seventeen samples (7.11%) did not meet the quality control standards for relative lineage abundance assessment, as their genome coverage fell below 60% (Supplementary Material S1, Table S1).

A significant negative correlation was observed between genome coverage and the average MS2 Ct value ( $r_s = -0.293$ ,  $p < 0.001$ ). A positive and significant correlation was noted between the concentration of SARS-CoV-2 in wastewater and genome coverage ( $r_s = 0.596$ ,  $p < 0.001$ ). The paired data for each of these correlations are available in Supplementary Material S1, Table S2 and Table S3.

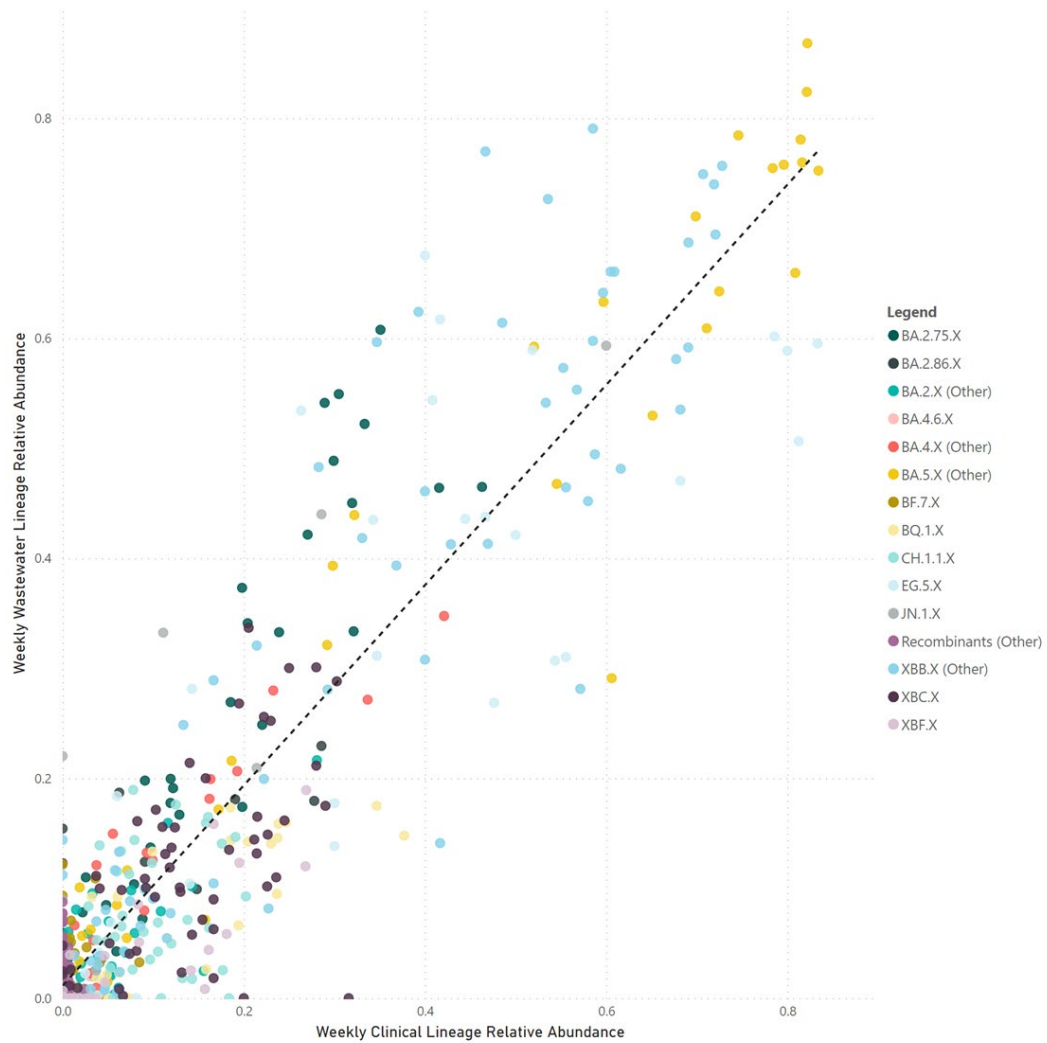
### SARS-CoV-2 Relative Lineage Abundance Correlation

The analysis of relative weekly parent lineage abundance in wastewater accurately captured the dominance of specific parent lineages during the identified waves, consistent with clinical data. Wave 1 was predominantly characterized by BA.4.X and BA.5.X sub-lineages. Wave 2 was mainly dominated by BA.2.75.X sub-lineages. In Wave 3, XBB.X sub-lineages were prevalent, and Wave 4 initially saw the dominance of the XBB.X sub-lineage EG.5.X, later transitioning to the BA.2.86.X sub-lineages taking precedence, including JN.1.X (Figure 2). Parent lineage grouping for each identified sub-lineage in each wastewater and clinical samples are recorded in Supplementary Material S1, Table S4 and Table S5, respectively.

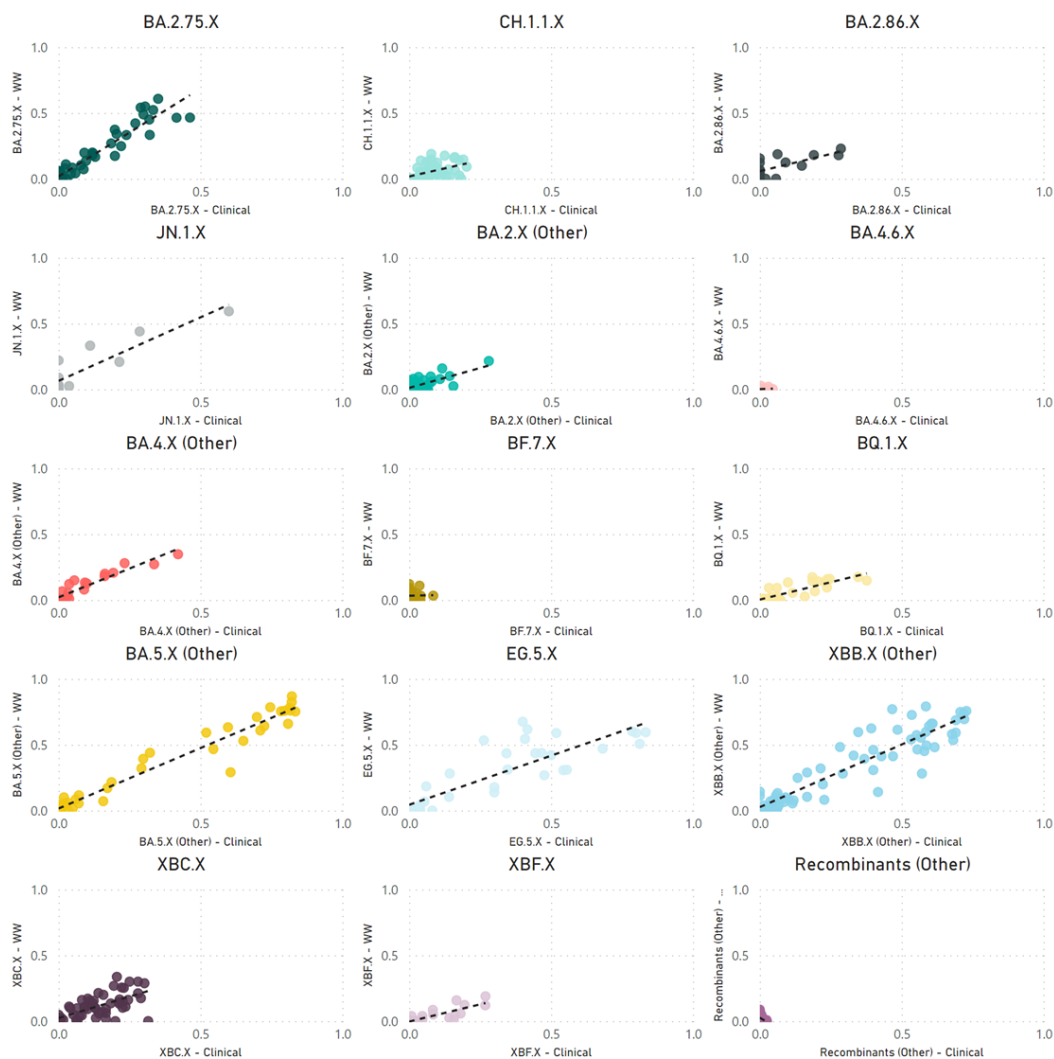


**Figure 2.** Weekly proportion of SARS-CoV-2 parent lineages in metropolitan Perth, WA – July 2022 to December 2023. Clinical cases (top) compared to weekly metropolitan wastewater relative abundance bottom (bottom). The X following the lineage name indicates the inclusion of all corresponding sub-lineages unless otherwise separated (e.g., BA.2.75.X is not inclusive of CH.1.1.X).

Paired analyses of weekly wastewater relative parent-lineage abundance and the proportion of clinical cases in the metropolitan area showed a strong positive correlation ( $r_s = 0.728$ ,  $p < 0.001$ ) (Figure 3). The correlation coefficients varied across separate parent lineages. Notably, stronger positive correlations were observed for parent lineages with moderate to high abundance in clinical patients. These included BA.4.X (Other) ( $r_s = 0.916$ ,  $p < 0.001$ ), BA.5.X (Other) ( $r_s = 0.908$ ,  $p < 0.001$ ), XBB.X (Other) ( $r_s = 0.883$ ,  $p < 0.001$ ), EG.5.X ( $r_s = 0.845$ ,  $p < 0.001$ ), BA.2.75.X ( $r_s = 0.823$ ,  $p < 0.001$ ), BQ.1.X ( $r_s = 0.788$ ,  $p < 0.001$ ), XBF.X ( $r_s = 0.770$ ,  $p < 0.001$ ), JN.1.X ( $r_s = 0.717$ ,  $p = 0.009$ ), BA.2.86.X ( $r_s = 0.597$ ,  $p = 0.019$ ), CH.1.1.X ( $r_s = 0.555$ ,  $p < 0.001$ ), and BA.2.X (Other) ( $r_s = 0.491$ ,  $p < 0.001$ ). In contrast, parent lineages that were sporadically detected in both clinical cases and wastewater showed no or negative correlation, such as BF.7.X ( $r_s = 0.045$ ,  $p = 0.838$ ), BA.4.6.X ( $r_s = -0.040$ ,  $p = 0.870$ ), and recombinants (Other) ( $r_s = -0.456$ ,  $p < 0.001$ ) (Figure 4).



**Figure 3.** Scatter plot of the combined weekly proportion of SARS-CoV-2 parent lineages in metropolitan Perth, WA, wastewater (y-axis) and clinical cases (x-axis). Data that is not paired were excluded (i.e., weeks where clinical data are available and wastewater data are not). Dashed line indicates the line of best fit.



**Figure 4.** Scatter plots of the combined weekly proportion of SARS-CoV-2 parent lineages in metropolitan Perth, WA, wastewater (y-axis) and clinical cases (x-axis). Data that is not paired were excluded (i.e., weeks where clinical data are available, and wastewater data are not). Dashed line indicates the line of best fit.

Throughout the study, there was a decrease in the weekly number of clinical cases sequenced due to changes in funding of clinical GS. From Quarter 3 (Q3) 2022 to Quarter 4 (Q4) 2023, a significant positive correlation was consistently observed between the proportions of parent lineages in clinical and wastewater samples. However, the correlation in Q4 2023 was comparatively lower ( $r_s = 0.670$ ,  $p < 0.001$ ) (Table 1).

#### *Initial Detection of Lineage Groups – Wastewater vs Clinical*

The initial detection dates for each parent lineage group were compared between clinical and wastewater samples. Sub-lineages including recombinants (Other), BA.2.X, BA.4.X, BA.5, and BA.2.75.X were excluded from this comparison, as their primary parent lineage had been assigned a Pangolin designation more than a day prior to the commencement of the study. Among the analysed parent lineages, 5 out of 10 were identified in wastewater prior to their detection in clinical samples. The range of this lead time varied, with the earliest being 65 days prior to clinical detection for BF.7.X and the latest being 52 days post-clinical detection for CH.1.1.X (Table 2).

**Table 2.** Earliest collection date of positive samples for select parent lineage groups between wastewater and clinical specimen types.

Lineage	Pangolin Designation Date	Wastewater (Earliest Date of Collection)	Clinical (Earliest Date of Collection)	Difference (Number of Days)
BA.4.6.X	03/07/2022	05/07/2022	15/08/2022	-41
BF.7.X	16/07/2022	19/07/2022	22/09/2022	-65
BQ.1.X	03/09/2022	29/09/2022	25/09/2022	+4
XBB.X (Other)	18/09/2022	06/10/2022	30/09/2022	+6
XBC.X	25/09/2022	13/10/2022	13/10/2022	0
CH.1.1.X	12/10/2022	21/12/2022	30/10/2022	+52
XBF.X	04/11/2022	17/11/2022	01/12/2022	-14
JN.1.X	13/01/2023	19/10/2023	12/11/2023	-24
EG.5.X	23/04/2023	20/04/2023	16/05/2023	-26
BA.2.86.X	17/08/2023	12/10/2023	22/08/2023	+51

#### 4. Discussion

Our 18-month investigation establishes a strong positive correlation between average wastewater viral loads and PathWest clinical PCR positivity rates ( $r_s = 0.772$ ,  $p < 0.001$ ). However, this correlation was weaker when compared to confirmed PCR case notifications ( $r_s = 0.577$ ,  $p < 0.001$ ) which can be expected due to reporting bias and shifts in testing modality. This finding supports the globally recognised observation that SARS-CoV-2 concentrations in wastewater are indicative of the prevalence of clinical cases within the associated community[31] and reinforces the role of WBE as a complementary surveillance tool to clinical epidemiology in enhancing public health surveillance efforts, especially when clinical data may be limited.

As our study progressed, there was an increasing divergence between wastewater SARS-CoV-2 concentrations and PCR COVID-19 clinical case notifications, as illustrated in Figure 1B. The divergence can be primarily attributed to decreased COVID-19 clinical testing, with a notable shift from higher rates of PCR testing in Q3-2022 to the increased adoption of self-conducted RATs as the predominant testing modality throughout the remaining quarters[32]. This shift was contributed to by state and workplace-led programs that offered free RATs[33], making them more accessible and affordable compared to PCR testing. The attenuation in PCR testing was further influenced by several factors, including the rescindment of testing mandates for clinical cases and close contacts, the closure of independent and free COVID-19 testing clinics (which reduced access to PCR testing centres)[34], and changes in community attitudes towards COVID-19, leading to decreased testing and reduced medical attendance for COVID-19 symptoms. Over the period of July 2022 to December 2023, the decrease in PCR testing and notification for COVID-19 highlights major limitations in traditional clinical surveillance systems in understanding the true community burden of disease, due to the factors discussed above. In the context of our study, WBE continued to provide a strong understanding of the population-wide burden of COVID-19 despite reduced PCR testing, demonstrating a significant benefit of WBE over traditional clinical surveillance systems in its representativeness and ability to provide additional surveillance information.

Our correlation between SARS-CoV-2 parent lineage proportions in wastewater and clinical cases over an 18-month period ( $r_s = 0.728$ ,  $p < 0.001$ ) underscores the effectiveness of wastewater GS in accurately measuring the prevalence of Omicron and recombinant lineages within a community. By condensing lineages into their respective parent lineages, we navigated the challenges posed by identifying numerous low-frequency sub-lineages. This approach made the data more accessible to audiences with limited genomic epidemiology expertise, facilitating a clearer understanding of virus

dynamics in the population. Table 1 demonstrates a trend in decreasing number of clinical samples undergoing GS from Q3 2022 to Q4 2023, primarily attributed to decreased funding towards clinical WGS of PCR samples. The reduction in clinical WGS of SARS-CoV-2 has also led to reduced strength of correlation between the lineages identified in clinical WGS and wastewater GS. As the denominator for clinical WGS (number of samples sequenced) reduces, the overall scale for comparison is also reduced. This means that smaller changes in the numerator (each individual clinical WGS result) have a greater and more pronounced effect on the overall frequency or proportion identified. This numerator-denominator bias does not affect the results from wastewater GS. This is because the wastewater GS results rely on the proportion of lineages identified in a sample, which is inherently representative of the population. The wastewater GS denominator encompasses all lineages identified in that specific sample, mitigating potential biases associated with variations in the sample size.

Despite efforts, challenges were encountered in accurately correlating wastewater data with clinical metrics for closely related sub-lineages within the same parent lineage, such as XBB.1.5, XBB.1.9, and XBB.1.16. These difficulties likely stem from a small number (<5) of single nucleotide polymorphisms between these highly related sub-lineages. Additionally, reduced sequencing depth in important genomic regions and the limitations of short-read sequencing technologies may hinder bioinformatics tools' ability to distinguish between these lineages at the read level. Long-read sequencing holds the potential to improve the differentiation between closely related sub-lineages, due to its capacity to identify a greater number of mutations within a single read. However, the fragmentation of SARS-CoV-2 RNA in wastewater could impede the success of long-read sequencing.

The representative identification of SARS-CoV-2 lineages in wastewater provides a foundation for applying similar genomic techniques to characterise other pathogens present in wastewater, as has been recently demonstrated for polioviruses[15], RSV[35] and influenza[36]. This methodological approach not only reinforces the utility of WBE for SARS-CoV-2 surveillance but also underscores its potential for broader pathogen detection and characterisation[37], offering a scalable and non-invasive tool for public health monitoring and response planning.

Our research highlights the potential of WBE to preemptively identify SARS-CoV-2 variants in the community, with wastewater surveillance occasionally detecting variants before clinical confirmation (Table 2). In WA, clinical PCR testing predominantly targets symptomatic individuals, with GS performed on a select group of these samples. Conversely, our wastewater surveillance sampling involved bi-weekly collections of two 24-hour composite samples from three WWTPs, each representing the same weekly timeframe. This bi-weekly sampling schedule, while systematic, limits the timeliness of WBE as an early detection tool due to the frequency of data points. Enhancing wastewater sampling frequency or adopting continuous surveillance could markedly improve early variant detection capabilities.

In our correlation analyses, we aligned the week of sample collection for both clinical and wastewater specimens to ensure comparability. However, the inherent delay in laboratory analysis and result reporting presents a barrier to the immediate utility of WBE as an early warning system. Although samples are collected systematically, GS on both clinical and wastewater specimens was conducted weekly in WA, with wastewater GS results reported approximately seven days post-collection. Addressing these reporting delays is critical for maximizing WBE's effectiveness as a proactive tool in public health surveillance. Reducing lag times, while mindful of logistical and economic limitations, could significantly enhance public health response agility by offering timely insights into emergent trends, thereby optimising early detection and intervention strategies for infectious diseases. Additional correlation analyses, incorporating reporting dates, are required to gain a clearer understanding of the utility of WBE as an early predictor in WA.

Despite the advantages of WBE, the importance of maintaining clinical sequencing alongside wastewater surveillance should not be overstated. Ultimately, WBE provides information at the population level. Clinical sequencing, however, offers insights into the clinical implications of new lineages[38], including their implications on disease severity, antiviral suitability, and vaccine efficacy[39]. Moreover, it allows confirmation of the presence and effects of specific sub-lineages in

individual patients, a dimension beyond the scope of wastewater analysis. Clinical sequencing also validates and complements wastewater findings, ensuring a comprehensive understanding of the virus's behaviour and evolution. Thus, integrating both wastewater and clinical sequencing approaches is fundamental to a robust, holistic surveillance strategy in the context of COVID-19. The strong and significant correlations noted throughout this study indicate that the relative abundances of lineages in wastewater may continue to reflect community clinical trends, despite the evolving dynamics of clinical surveillance strategies.

The implementation of WBE provides significant benefits for public health monitoring, yet it faces numerous challenges. Accurate quantification of viral concentrations in wastewater is complicated by its heterogeneity and PCR inhibition from various compounds[40]. The interpretation of WBE data is further complicated by limitations in understanding pathogen faecal shedding kinetics and the critical role of sampling site and method selection in influencing the accuracy of epidemiological analyses[41]. Although external factors like stormwater ingress can dilute viral concentrations, necessitating normalisation techniques for accurate measurement[41], such methods were not performed in this study. The reason for this omission is attributed to the moderate to strong positive correlation observed between clinical outcomes and WBE data throughout the study period, suggesting that, despite these challenges, WBE can still provide valuable insights into disease prevalence in WA. Moreover, the separation of stormwater and wastewater systems in WA minimises the impact of rainfall on the wastewater analysis, though normalization markers like pepper mild mottle virus (ppMoV) could enhance quantification data interpretation by accounting for population dynamics and changes in water use.

Our study encountered specific challenges, notably during the week ending 19 February 2023, when all wastewater samples failed GS due to SARS-CoV-2 genome coverage below 60%, attributed to low SARS-CoV-2 wastewater concentrations. However, these instances were relatively rare, mitigated by our testing protocols designed for low viral concentrations in response to stringent border controls in WA[42] and the need to detect and characterise virus importation events[26]. Collaboration with the Australian WBE consortium (ColoSSoS, WaterRA) enabled the selection of a qPCR kit resistant to common wastewater inhibitors, enhancing the reliability of SARS-CoV-2 RNA quantification. This approach effectively addressed most challenges associated with WBE, demonstrating the critical dependency of GS efficacy on sufficient viral load and highlighting the importance of overcoming molecular assay inhibitors to generate reliable results.

Maximising the effectiveness of WBE necessitates the facilitation of data exchange and integration between wastewater surveillance networks and healthcare systems. The ability to efficiently share data is vital for aligning WBE insights with clinical data. However, this process can be hindered by logistical and regulatory barriers, which may result in delays in data comparison and introduce gaps in the overall epidemiological understanding. Overcoming these data-sharing obstacles is essential for improving the precision and applicability of WBE in public health monitoring and intervention strategies. In WA, the presence of a well-established environmental microbiology unit within the state public health laboratory, PathWest, facilitated simple and rapid data exchange between the laboratory and state public health authorities. This arrangement helped to overcome numerous data-sharing challenges and regulatory constraints, attributable to the laboratory's status as a government health service provider.

The proactive use of this WBE surveillance program by the WA Department of Health exemplifies a forward-thinking approach to pandemic and public health management. The department used wastewater data to facilitate a better understanding of COVID-19 prevalence and trends within the community, guiding public health decisions and strategies. The decision to make this data publicly accessible via an online dashboard[23], updated weekly, ensured transparent communication of the latest trends to the public and the media. This initiative not only facilitated timely information dissemination but also enabled individuals to make informed personal risk assessments, contributing to a more informed public discourse during the pandemic. This strategy, blending advanced genomic surveillance with traditional epidemiological methods, significantly

bolstered the overall response to COVID-19, fostering an environment of informed decision-making crucial for effectively managing a public health challenge of this scale.

## 5. Conclusions

Our study demonstrates the benefits of wastewater viral quantification and lineage determination in understanding community-wide trends of SARS-CoV-2. This approach, integrating total viral RNA concentration measurement with prevalent lineage identification, offers vital insights into the intensity and patterns of SARS-CoV-2 transmission waves and provides a window into genetic shifts and lineage predominance. These insights are invaluable for understanding the true burden of COVID-19 in the community and variant distribution, and ultimately aided in shaping informed public health responses throughout the pandemic and beyond. Our study demonstrates that wastewater surveillance reflects community-wide virus activity and serves as a timely and efficient tool for public health management, with the benefits of WBE having the potential to be applied to other communicable diseases.

**Supplementary Materials:** The following supporting information can be downloaded at Preprints.org. S1 – Table S1 – Wastewater samples that underwent WGS, the associated SARS-CoV-2 genome coverage and date of Freyja bioinformatics processing. S1 – Table S2 – Paired wastewater data including SARS-CoV-2 genome coverage and MS2 Ct value used for correlation. S1 – Table S3 – Paired wastewater data including SARS-CoV-2 genome coverage and SARS-CoV-2 wastewater concentration used for correlation. S1 – Table S4 – SARS-CoV-2 sub-lineages detected in wastewater and the associated parent lineage grouping. S1 – Table S5 – SARS-CoV-2 sub-lineages detected in clinical cases and the associated parent lineage grouping. S2 – Table S1 – Weekly average SARS-CoV-2 wastewater concentrations, PathWest SARS-CoV-2 PCR Clinical PCR positivity and total metropolitan PCR confirmed notifications. S2 – Table S2 – Weekly average SARS-CoV-2 grouped lineage abundances in wastewater. S2 – Table S3 – Weekly average SARS-CoV-2 grouped lineages abundances in clinical cases. S2 – Table S4 – Number of samples used each week to calculate the average for SARS-CoV-2 concentration and grouped lineage abundances.

**Author Contributions:** Conceptualization, Jake Gazeley, Terence Lee, and Avram Levy. Methodology & Validation, Jake Gazeley, Terence Lee, Daniel R Knight, and Avram Levy. Bioinformatics Software Development and Implementation, Terence Lee, and Daniel R Knight. Data Provision, Dylan Barth, Alex Shivarev, Jelena Maticcevic, Paul Knight, and Paul Armstrong. Data Analysis & Visualization, Jake Gazeley, Dylan D Barth, and Paul Knight. Writing – Original Draft preparation, Jake Gazeley. Writing – Review and Editing, Jake Gazeley, Terence Lee, Avram Levy, Daniel R Knight, Meredith Hodge, Sandra Sjollem, David Speers, Dylan D Barth, Alex Shivarev, Jelena Maticcevic, Paul Knight, Paul Armstrong. Sample Collection Coordination, Cameron Gordon. Supervision, Avram Levy, Sandra Sjollem, Meredith Hodge, David Speers, Cameron Gordon, Jelena Maticcevic, and Paul Armstrong. Project Administration, Avram Levy, Sandra Sjollem, Meredith Hodge, David Speers, Cameron Gordon, Jelena Maticcevic, and Paul Armstrong.

**Funding:** This work was funded by the Western Australian Department of Health and PathWest Laboratory Medicine WA. This surveillance program received no external funding.

**Data Availability Statement:** The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials. Additional data that support the findings of this study are available from the corresponding author, JG, upon reasonable request and approval by the Western Australia Wastewater Advisory Committee.

**Acknowledgments:** We appreciate and thank all the extended staff involved with this work at WA Department of Health, PathWest Laboratory Medicine WA, and Water Corporation, including sample collectors, laboratory scientists and technicians and staff involved in administration and logistics.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. World Health Organization. Environmental surveillance for SARS-CoV-2 to complement other public health surveillance. Geneva; 2023.
2. Galanti M, Comito D, Ligon C, Lane B, Matienzo N, Ibrahim S, et al. Active surveillance documents rates of clinical care seeking due to respiratory illness. *Influenza Other Respir Viruses*. 2020;14(5):499-506.
3. Bonačić Marinović A, Swaan C, van Steenberg J, Kretzschmar M. Quantifying reporting timeliness to improve outbreak control. *Emerg Infect Dis*. 2015;21(2):209-16.

4. Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wijert JHHM, Bonten MJM. Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study. *The Lancet Public Health*. 2020;5(8):e452-e9.
5. Australian Government COVID-19 rapid antigen tests: Guidance on performance requirements and risk mitigation strategies. In: Department of Health and Aged Care - Therapeutic Goods Administration, editor. Version 3.0 ed. Australia2023.
6. Lodder W, de Roda Husman AM. SARS-CoV-2 in wastewater: potential health risk, but also data source. *Lancet Gastroenterol Hepatol*. 2020;5(6):533-4.
7. Feng S, Roguet A, McClary-Gutierrez JS, Newton RJ, Kloczko N, Meiman JG, et al. Evaluation of Sampling, Analysis, and Normalization Methods for SARS-CoV-2 Concentrations in Wastewater to Assess COVID-19 Burdens in Wisconsin Communities. *ACS ES&T Water*. 2021;1(8):1955-65.
8. Weidhaas J, Aanderud ZT, Roper DK, VanDerslice J, Gaddis EB, Ostermiller J, et al. Correlation of SARS-CoV-2 RNA in wastewater with COVID-19 disease burden in sewersheds. *Science of The Total Environment*. 2021;775:145790.
9. Bagutti C, Alt Hug M, Heim P, Maurer Pekerman L, Ilg Hampe E, Hübner P, et al. Wastewater monitoring of SARS-CoV-2 shows high correlation with COVID-19 case numbers and allowed early detection of the first confirmed B.1.1.529 infection in Switzerland: results of an observational surveillance study. *Swiss Medical Weekly*. 2022;152(2526):w30202.
10. Boehm AB, Hughes B, Duong D, Chan-Herur V, Buchman A, Wolfe MK, et al. Wastewater concentrations of human influenza, metapneumovirus, parainfluenza, respiratory syncytial virus, rhinovirus, and seasonal coronavirus nucleic-acids during the COVID-19 pandemic: a surveillance study. *The Lancet Microbe*. 2023;4(5):e340-e8.
11. Hughes B, Duong D, White BJ, Wigginton KR, Chan EMG, Wolfe MK, et al. Respiratory Syncytial Virus (RSV) RNA in Wastewater Settled Solids Reflects RSV Clinical Positivity Rates. *Environmental Science & Technology Letters*. 2022;9(2):173-8.
12. de Melo T, Islam G, Simmons DBD, Desaulniers JP, Kirkwood AE. An alternative method for monitoring and interpreting influenza A in communities using wastewater surveillance. *Front Public Health*. 2023;11:1141136.
13. Assoum M, Lau CL, Thai PK, Ahmed W, Mueller JF, Thomas KV, et al. Wastewater Surveillance Can Function as an Early Warning System for COVID-19 in Low-Incidence Settings. *Trop Med Infect Dis*. 2023;8(4).
14. Li X, Liu H, Gao L, Sherchan SP, Zhou T, Khan SJ, et al. Wastewater-based epidemiology predicts COVID-19-induced weekly new hospital admissions in over 150 USA counties. *Nature Communications*. 2023;14(1):4548.
15. Shaw AG, Mampuela TK, Lofiko EL, Pratt C, Troman C, Bujaki E, et al. Sensitive poliovirus detection using nested PCR and nanopore sequencing: a prospective validation study. *Nature Microbiology*. 2023;8(9):1634-40.
16. Karthikeyan S, Levy JI, De Hoff P, Humphrey G, Birmingham A, Jepsen K, et al. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. *Nature*. 2022;609(7925):101-8.
17. Gupta P, Liao S, Ezekiel M, Novak N, Rossi A, LaCross N, et al. Wastewater Genomic Surveillance Captures Early Detection of Omicron in Utah. *Microbiol Spectr*. 2023;11(3):e0039123.
18. Yousif M, Rachida S, Taukobong S, Ndlovu N, Iwu-Jaja C, Howard W, et al. SARS-CoV-2 genomic surveillance in wastewater as a model for monitoring evolution of endemic viruses. *Nature Communications*. 2023;14(1):6325.
19. Li L, Uppal T, Hartley PD, Gorzalski A, Pandori M, Picker MA, et al. Detecting SARS-CoV-2 variants in wastewater and their correlation with circulating variants in the communities. *Scientific Reports*. 2022;12(1):16141.
20. Kaya D, Falender R, Radniecki T, Geniza M, Cieslak P, Kelly C, et al. Correlation between Clinical and Wastewater SARS-CoV-2 Genomic Surveillance, Oregon, USA. *Emerg Infect Dis*. 2022;28(9):1906-8.
21. Norwegian Institute of Public Health. Weekly reports on COVID-19, Influenza and other respiratory viruses.
22. Centre for Disease Control & Prevention USA. COVID-19 Variants in Wastewater [Available from: <https://www.cdc.gov/nwss/rv/COVID19-variants.html>].
23. COVID-19 wastewater surveillance [Internet]. 2024 [cited 25/02/2024]. Available from: [https://www.health.wa.gov.au/articles/a\\_e/coronavirus/covid19-wastewater-surveillance](https://www.health.wa.gov.au/articles/a_e/coronavirus/covid19-wastewater-surveillance).
24. Rajput V, Pramanik R, Malik V, Yadav R, Samson R, Kadam P, et al. Genomic surveillance reveals early detection and transition of delta to omicron lineages of SARS-CoV-2 variants in wastewater treatment plants of Pune, India. *Environmental Science and Pollution Research*. 2023;30(56):118976-88.
25. Australian Bureau of Statistics. Greater Perth, 2021 Census All persons QuickStats. 2021.
26. Levy A, Gazeley J, Lee T, Jardine A, Gordon C, Cooper N, et al. Whole genome sequencing of SARS-CoV-2 from wastewater links to individual cases in catchments. *Sci Total Environ*. 2022;851(Pt 2):158266.

27. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nature Microbiology*. 2020;5(11):1403-7.
28. Government of Western Australia. Public Health Act. 2016.
29. Government of Western Australia. Emergency Management Act. 2005.
30. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*: Elsevier Science; 2013.
31. Shah S, Gwee SXW, Ng JQX, Lau N, Koh J, Pang J. Wastewater surveillance to infer COVID-19 transmission: A systematic review. *Science of The Total Environment*. 2022;804:150060.
32. Government of Western Australia - Department of Health. COVID-19 weekly surveillance report 1 – 7 January 2024.
33. Department of Health - Government of Western Australia. Free RAT program ends 2023 25/02/2024. Available from: <https://www.health.wa.gov.au/News/2023/Free-RAT-program-ends>.
34. Government of Western Australia. Announcements - Western Australian government news stories, media releases and community updates [
35. Allen DM, Reyne MI, Allingham P, Levickas A, Bell SH, Lock J, et al. Genomic Analysis and Surveillance of Respiratory Syncytial Virus (RSV) Using Wastewater-Based Epidemiology (WBE). medRxiv. 2023:2023.07.21.23293016.
36. Vo V, Harrington A, Chang CL, Baker H, Moshi MA, Ghani N, et al. Identification and genome sequencing of an influenza H3N2 variant in wastewater from elementary schools during a surge of influenza A cases in Las Vegas, Nevada. *Sci Total Environ*. 2023;872:162058.
37. Tisza M, Javornik Cregeen S, Avadhanula V, Zhang P, Ayvaz T, Feliz K, et al. Wastewater sequencing reveals community and variant dynamics of the collective human virome. *Nature Communications*. 2023;14(1):6878.
38. Berno G, Fabeni L, Matusali G, Gruber CEM, Rueca M, Giombini E, et al. SARS-CoV-2 Variants Identification: Overview of Molecular Existing Methods. *Pathogens*. 2022;11(9).
39. Goswami C, Sheldon M, Bixby C, Keddache M, Bogdanowicz A, Wang Y, et al. Identification of SARS-CoV-2 variants using viral sequencing for the Centers for Disease Control and Prevention genomic surveillance program. *BMC Infectious Diseases*. 2022;22(1):404.
40. Mao K, Zhang K, Du W, Ali W, Feng X, Zhang H. The potential of wastewater-based epidemiology as surveillance and early warning of infectious disease outbreaks. *Curr Opin Environ Sci Health*. 2020;17:1-7.
41. Kumblathan T, Liu Y, Uppal GK, Hrudey SE, Li XF. Wastewater-Based Epidemiology for Community Monitoring of SARS-CoV-2: Progress and Challenges. *ACS Environ Au*. 2021;1(1):18-31.
42. Government of Western Australia. Review of Western Australia's COVID-19 Management and Response - July 2023. 2023.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.