

Wastewater Tiling Amplicon Sequencing Reveals Longitudinal Dynamics of SARS-CoV-2 Variants Prevalence in the Community

Yu Wang¹, Gaofeng Ni^{2*}, Wei Tian³, Haofei Wang³, Jiaying Li⁴, Phong Thai⁴, Phil M. Choi⁵, Greg Jackson⁵, Shihu Hu¹, Bicheng Yang³, Jianhua Guo^{1*}

¹Australian Centre for Water and Environmental Biotechnology, The University of Queensland, St. Lucia, Brisbane, QLD 4072, Australia

²Department of Microbiology, Biomedicine Discovery Institute, Monash University, Melbourne, Victoria, Australia

³MGI Australia Pty Ltd, 300 Herston Road, Herston, Brisbane, QLD 4006, Australia

⁴Queensland Alliance for Environmental Health Sciences (QAEHS), The University of Queensland, Brisbane, Queensland, Australia

⁵Water Unit, Health Protection Branch, Queensland Public Health and Scientific Services, Queensland Health, Brisbane, Queensland, Australia

*Corresponding author:

Jianhua Guo, Email: jianhua.guo@uq.edu.au

Gaofeng Ni, Email: gaofeng.ni@monash.edu

1 **Abstract**

2 SARS-CoV-2 continues to evolve, while the decline in clinical sequencing efforts hampers
3 public health sectors to prepare for the menace of ongoing variant emergence and future
4 COVID-19 surges.¹ Wastewater-based epidemiology (WBE) has been proposed to provide
5 complementary insights on the variants being transmitted in communities.² However, limited
6 research has been dedicated to the use sequencing methods for tracking disease prevalence and
7 variant dynamics in wastewater, particularly on a large scale. Here, we employed a tiling
8 amplicon sequencing to track the dynamics of variant of concern (VOC) in wastewater
9 collected from Queensland, Australia from 2020 to 2022. RNA concentrations in wastewater
10 measured by ATOplex showed a stronger correlation and greater consistency with the number
11 of daily new cases than a PCR-based method. The VOC dynamics observed in wastewater were
12 largely in line with clinical reports. These findings support that WBE and sensitive sequencing
13 methods can serve as a long-term approach for disease surveillance, thus aiding in disease
14 outbreak prevention, control, and management.

15 **Keywords:** Wastewater-based epidemiology; Genome sequencing; Variants of concern;
16 Variant dynamics; Prevalence

17

18

19 **1. Introduction**

20 The World Health Organization (WHO) designates evolved variants with genetic mutations
21 associated with increased transmissibility, pathogenicity and immune escape as variants of
22 concern (VOCs). Since the appearance of the first VOC, namely Alpha variant, in 2020,
23 emergent variants have caused several global surges in COVID-19.³ The monitoring of VOCs
24 and related genetic characteristics typically depends on extensive individual clinical
25 sequencing.⁴ However, clinical surveillance has an unavoidable risk of sampling bias and is
26 becoming less effective in revealing circulating variants, as the global response to COVID-19
27 strategies are adjusted.^{5,6} This situation is even worse where medical resources are scarce, as
28 large-scale clinical sequencing is laborious, expensive and time-consuming.^{4,6} The risk of new
29 variants to emerge calls for sustained monitoring strategies to inform timely preventive
30 measures.⁷

31 Wastewater-based epidemiology (WBE) enables effective identification and quantification of
32 disease prevalence at a community level.⁸ It has been implemented as a surveillance program
33 worldwide, such as in the USA, Australia, and Europe in response to COVID-19.⁹⁻¹² The use
34 of PCR-based methods has facilitated WBE in monitoring the disease burden in communities
35 and providing early warning of disease outbreaks before clinical surveillance.¹³⁻¹⁵ However,
36 mutations or degradation of the N gene, the widely-used PCR primer binding region, will
37 reduce the primer binding efficiency. This may lead to false negative PCR results and impeding
38 the timeliness of WBE.^{16,17} Moreover, despite advancements in variant-specific PCR assays
39 for identifying different VOCs in wastewater, determining transmission links and timely
40 detecting emerging variants in a community remain significant challenges.^{18,19}

41 In contrast, sequencing technologies can directly recover viral genomes from wastewater,
42 allowing the discerning of different variants and their co-occurrence.⁵ The SARS-CoV-2 RNA
43 in wastewater is typically present in low concentration and is readily fragmented, which

44 requires a sensitive sequencing method for the identification of SARS-CoV-2 and the recovery
45 of a high-quality genome.²⁰ Tiling amplicon sequencing, such as ARTIC, ATOplex and
46 Midnight, have been proposed as sensitive and cost-efficient approaches for recovering viral
47 genomes from wastewater.²¹⁻²³ By using amplicon sequencing technologies, wastewater
48 surveillance offers valuable information on disease prevalence and can potentially inform
49 transmission links in communities. For example, research has demonstrated a strong
50 correlation between the proportion of wastewater samples containing specific variant signals
51 and the corresponding percentage observed in clinical samples.^{24,25} The recovery of high-
52 coverage genomes from wastewater uncovers various VOCs introduced across different
53 regions.^{26,27} Building-level wastewater sequencing facilitates early detection of infection
54 events, enabling prompt actions to identify positive cases through clinical testing, then relocate
55 infected individuals, potentially preventing potential outbreaks.⁵ Furthermore, WBE
56 surveillance within a university campus has been shown to track the temporal profile of
57 different variants in a defined region through intensive sampling.^{5,14} However, still limited
58 effort has been made to assess the application of wastewater sequencing in tracking variant
59 dynamics and comparing it with clinical sequencing data. More importantly, it remains unclear
60 whether collecting wastewater from the strategically chosen sampling sites can effectively
61 track the variant dynamics in large-scale areas over the long term.

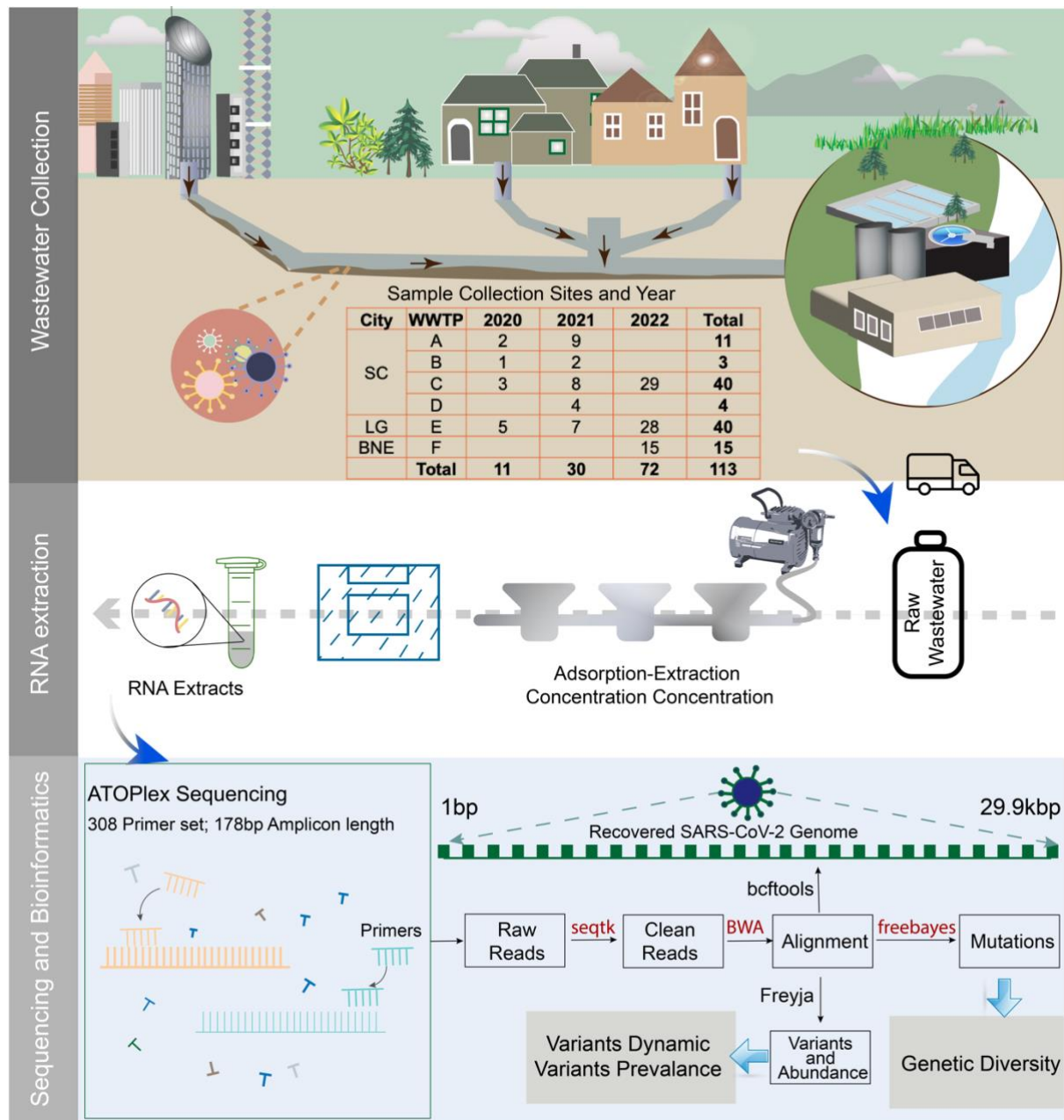
62 This study conducted a longitudinal investigation, collecting wastewater samples over a span
63 of three years (from mid-2020 to late-2022), from three municipal sites in Queensland,
64 Australia. The objective of this study was to evaluate the effectiveness of long-term wastewater
65 sequencing surveillance at sentinel sites in estimating disease prevalence and monitoring the
66 dynamics of SARS-CoV-2 variants at the state level. By using ATOplex sequencing, we
67 successfully tracked the circulating dynamics of SARS-CoV-2 variants and have the potential
68 to detect emerging variants at an early stage in wastewater samples. Furthermore, wastewater

69 sequencing exhibited stronger correlation and consistency in estimating disease prevalence
70 trends compared to a PCR-based method in different sampling sites. Overall, this study offers
71 valuable insights into the use of WBE for monitoring disease dynamics in the community,
72 particularly in the context of easing public health measures.

73 **2. Methods**

74 ***2.1 Study area and sample collection***

75 Wastewater samples were collected from six wastewater treatment plants (WWTPs) located in
76 three municipalities within Queensland, Australia (Figure S1). A total of 113 samples that
77 tested positive for SARS-CoV-2 via RT-qPCR testing were obtained, including WWTPs A
78 (n=11), B (n=3), C (n=40), D (n=4), E (n=40), and F (n=15), respectively (Figure 1). Samples
79 from WWTPs C and E were collected between July 2020 and September 2022, while WWTPs
80 A, B and D ceased collection in 2022. To further investigate the presence of emerging variants
81 that have not been locally endemic in Queensland, 15 grab samples were collected in
82 November-December 2022 from the largest WWTP in Queensland, WWTP F, which serves
83 the busiest locations of Queensland, including the capital city Brisbane (BNE) and an
84 international airport (Figure 1). The sampling sites and the served population of each WWTP
85 included in the study are depicted in Figure S1. In 2020 and for the majority of 2021,
86 Queensland implemented stringent public health measures, such as tight border restrictions,
87 during which 14 days of quarantine was required to enter the state. Consequently, the collected
88 samples had a low RNA concentration. Moreover, the RNA in wastewater was subject to
89 degradation during storage, which further contributed to negative RT-qPCR results at the time
90 of sequencing in September 2022.¹² Therefore, only a few samples from 2020/2021 were
91 included to test the sensitivity of ATOPlex sequencing.



92

93 Figure 1. Overview of the wastewater surveillance program. Top panel: Concept of the WBE
 94 and collection sites of raw wastewater samples. Middle panel: Wastewater concentration and
 95 RNA extraction. Bottom panel: ATOPlex sequencing and bioinformatics analysis.

96 2.2 Wastewater sample preparation, concentration, and RNA extraction

97 Samples were transported on ice and stored at -20 °C until further study. The frozen samples
 98 were first thawed at 4 °C followed by wastewater concentration and RNA extraction as
 99 described in previous studies (Figure 1).^{28,29} Briefly, 0.5 mL of 25 mM dissolved MgCl₂ was

100 added to the aliquots of 50 mL raw wastewater samples and incubated at room temperature for
101 a few minutes. Samples were then filtered through a 0.45- μ m pore-size, 47-mm diameter
102 electronegative HA membrane (HAWP04700; Merck Millipore Ltd, Sydney, Australia) using
103 a multichannel filter funnel. All equipment was autoclaved at 121 °C for 20 minutes prior to
104 use. The membrane was removed from the filtration apparatus, rolled, shredded and placed into
105 a 2 ml beat tube from RNeasy PowerMicrobiome Kit (Qiagen) using two sterile tweezers. RNA
106 was then extracted following the manufacturer's instructions with the following modifications:
107 (1) 100 μ L of phenol–chloroform–isoamyl alcohol (25:24:1) was added before the lysis step to
108 protect RNA integrity; (2) Beat-beat for 60s with Minilys® homogenizer at medium speed; (3)
109 IRS solution was increased to 200 μ L to remove more inhibitors, and (4) RNA eluted with 50
110 μ L Nuclease free water (Biolab) and pass-through spin column twice to ensure maximum RNA
111 yield. RNA extracts were stored at -80 °C before sequencing.

112 ***2.3 Positive control preparation***

113 A commercially available RNA positive control was used to prepare a serial dilution with 5
114 orders of gradient, as shown in Figure S1 (b). Briefly, 10 μ L RNA was reverse transcribed to
115 cDNA using the high-capacity cDNA reverse transcription kit (Catalog No. 4368814, Applied
116 Biosystems™) and serially diluted by 5 orders of magnitude with nuclease-free water.²⁶ The
117 concentration of each positive control gradient was measured by droplet digital Polymerase
118 Chain Reaction (dd-PCR) assay, as described in a previous study,²⁸ with the concentration
119 ranging from 6235.58 to 0.63 copies/ μ L.

120 ***2.4 ATOPlex sequencing***

121 A total of 113 wastewater RNA extracts and 5 positive control gradients (in triplicate) were
122 sequenced using the ATOPlex SARS-CoV-2 full-length genome panel (MGI, China) as
123 described previously. Briefly, sample RNA was reverse transcribed to cDNA with random
124 hexamers (5'-NNNNNN-3'). Then, 200 copies of Lambda phage DNA were dosed into each

125 sample as an internal control to set a baseline for result quantification. The resulting product
126 undergoes two steps of PCR amplification followed by purification, with the first PCR of 13
127 cycles for target enrichment, and the second PCR of 27 cycles for barcoding and product
128 amplification. The process was completed using the automation sample processing system
129 MGISP-960. The purified final PCR products were quantified using Qubit Quant-iT™ dsDNA
130 Assay Kits (ThermoFisher, Q33120). Subsequently, PCR products of 96 samples were pooled
131 into two pools, with each pool containing 48 samples and 400 ng of library. The pooled dsDNA
132 libraries were denatured to generate linear ssDNA, circularized using a splint oligo that is
133 complementary to the start and end of linear ssDNA, followed by digestion reaction and
134 purification to remove non-circularized DNA debris. The process was completed manually
135 using MGIEasy Dual Barcode Circularization Kit (MGI, China) to produce purified single-
136 strand circularized products. The ssDNA circularized product was then used as a template to
137 generate DNA nanoball (DNBs) by performing a rolling circle amplification reaction. The
138 DNBs were introduced into flow cell containing patterned arrays of DNB binding sites for
139 sequencing. The sequencing was done on a DNBSeg-G400 sequencer running for 24 hours to
140 get 8M reads for each sample at the MGI Australia Demo Lab.

141 ***2.5 Data analysis and interpretation***

142 As part of the surveillance program, wastewater monitoring of SARS-CoV-2 using the RT-
143 qPCR method was applied, and the result was obtained from Queensland Public Health and
144 Scientific Services. The sequencing raw data were analysed with the bioinformatic pipeline for
145 multiplex-PCR data developed by MGI ([https://github.com/MGI-tech-bioinformatics/SARS-](https://github.com/MGI-tech-bioinformatics/SARS-CoV-2_Multi-PCR_v1.0)
146 [CoV-2_Multi-PCR_v1.0](https://github.com/MGI-tech-bioinformatics/SARS-CoV-2_Multi-PCR_v1.0)). The established pipeline is used to analyse raw sequencing data,
147 including genome alignment, mutation calling, consensus genome generation and
148 identification (Figure 1).^{28,30} Viral concentration in wastewater was calculated based on
149 normalized SARS-CoV-2 sequencing reads. The higher coverage and depth underpin a greater

150 likelihood of accurate sequencing and variant detection at that position.²⁵ To ensure data
151 precision, only samples identified as positive and genomes with coverage (10X) of at least 20%
152 were included for variant identification and investigation of genetic diversity. Freyja (V1.3.11),
153 a well-established pipeline bioinformatics tool, was employed to determine co-circulating
154 variants and corresponding prevalence in wastewater samples.⁵ COVID-19 case data in the
155 study areas were sourced from the Queensland Government Open Data Portal
156 (<https://www.data.qld.gov.au>). Clinical genomic data from Queensland were sourced from the
157 GISAID database.³¹ The clinical genomes were then grouped by collection date, and mutations
158 in the groups were obtained using snp-sites (V2.5.1),³² the mutation frequency was calculated
159 using a custom script. Genetic diversity of the wastewater samples and clinical samples was
160 measured by the richness (count of the number of mutations present in the sample) and Shannon
161 Entropy ($H(x) = - \sum p(x) \log_2(p(x))$, where $p(x)$ is the allele frequency at the position x), the
162 distribution of each mutation and degree of genetic diversity in the sample).

163 ***2.6 Data availability***

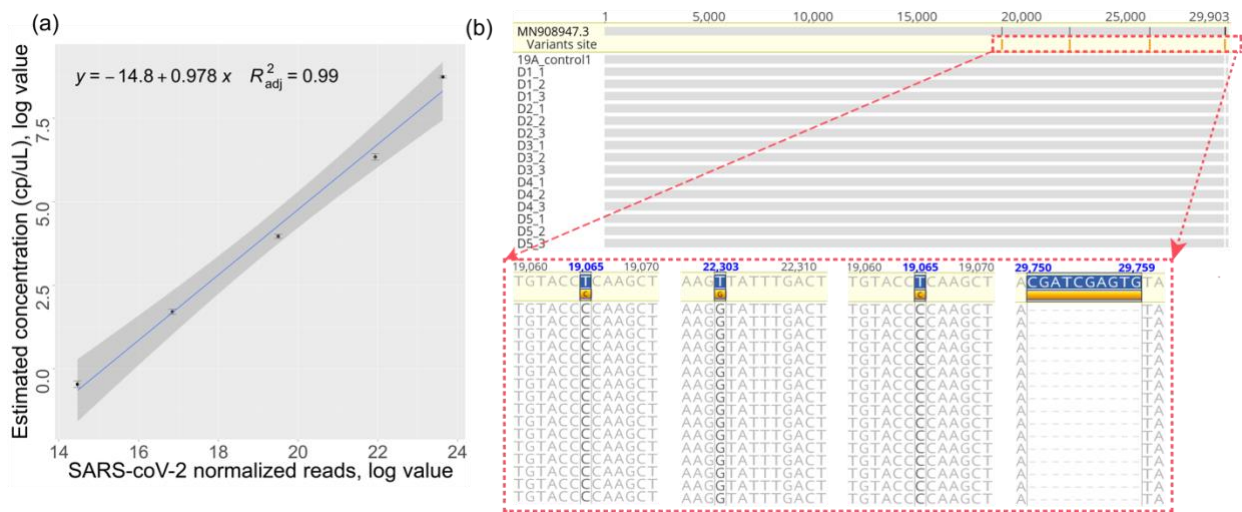
164 All the raw sequencing data are available at the National Center for Biotechnology Information
165 (NCBI) Sequence Read Archive (SRA) under BioProject number PRJNA945486.

166 **3. Results**

167 ***3.1 Accurate viral RNA concentration estimation and genome recovery***

168 To verify whether the chosen sequencing method can accurately quantify the virus
169 concentration, a linear regression analysis is used to predict the relationship between the RNA
170 concentration measured by dd-PCR with the sequencing reads. The result revealed a linear
171 regression correlation ($R^2_{adj}=0.99$) between the normalized SARS-CoV-2 sequencing reads and
172 absolute RNA concentration measured by dd-PCR from concentration gradients D1 (6235.58
173 cp/ μ L) to D5 (0.63 cp/ μ L), as shown in Figure 2-a. The result is consistent with our past study

174 and further verifies the stability of the ATOplex sequencing method.²⁸ Furthermore, ATOplex
 175 sequencing can accurately recover the SARS-CoV-2 genome (median 10X coverage of
 176 98.19%) even with mutations in comparison to the reference genome (MN908947.3) and at a
 177 concentration as low as 0.63 cp/μL (Figure 2-b). The accuracy estimation of viral concentration
 178 and derived genomes facilitated the subsequent analysis for qualitative and quantitative
 179 analysis of the SARS-CoV-2 virus.

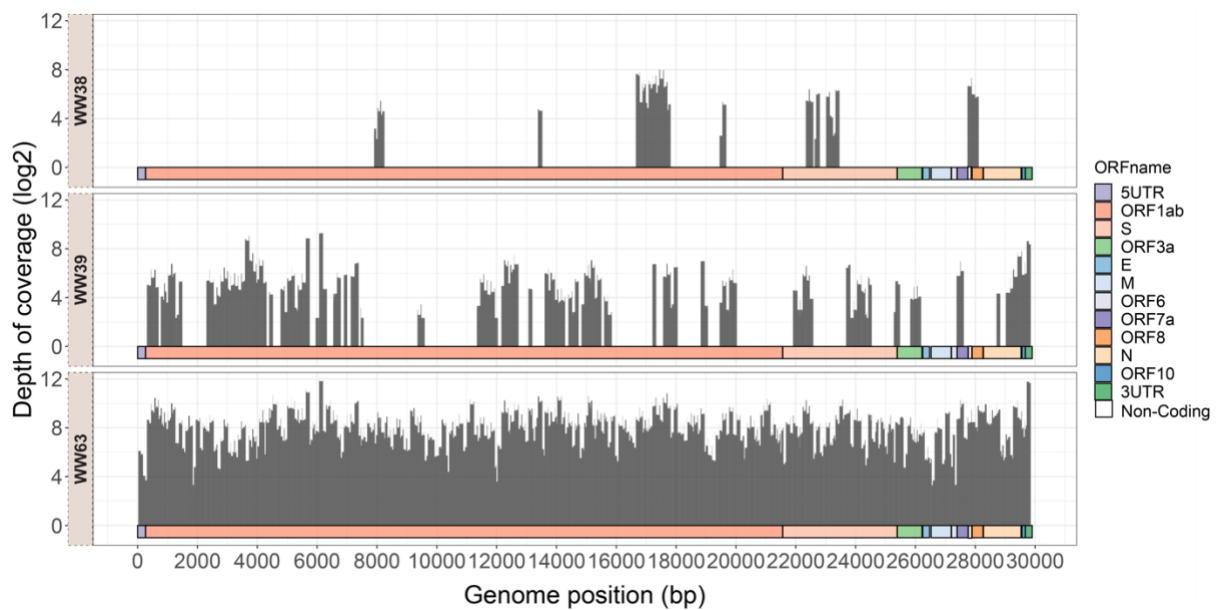


180
 181 Figure 2. (a) Linear relationship between normalized ATOplex SARS-CoV-2 reads and viral
 182 concentration measured by dd-PCR at log value. (b) Sequence alignments of the reference
 183 genome (MN908947.3) and ATOplex recovered genome across serial dilution D1-D5 align to
 184 the positive control 19A genome. Nucleotides that differ from the positive control genome are
 185 highlighted in black and have detailed alignment information enlarged.

186 3.2 Identification of the SARS-CoV-2 in wastewater samples

187 ATOplex classify samples as positive, negative, or indeterminate by normalizing the number
 188 of SARS-CoV-2 reads to the spike-in lambda phage DNA control reads and assessing genome
 189 coverage.^{28,30} In summary, out of the 113 wastewater samples 109 tested positive, while only
 190 4 samples were categorized as indeterminate due to low genome coverage. Wherein some
 191 samples might have become highly degraded and fragmented during long-time storage. To gain

192 further insights into the completeness of the recovered viral genome, the genome coverage was
193 visualized (Figure S2). Median SARS-CoV-2 genome coverage (10X) was 52.66% (range
194 0.69–99.88%). Figure 3 illustrates three examples of the distribution of sequencing reads with
195 a different genome coverage (10X). It is apparent that ATOPlex still can identify the SARS-
196 CoV-2 virus even in instances where the primer regions N1 and N2 commonly used in PCR-
197 based methods were degraded (Figure 3). ATOPlex sequencing can provide high-quality
198 genomes, except in several cases where the genome region has undergone significant
199 degradation.



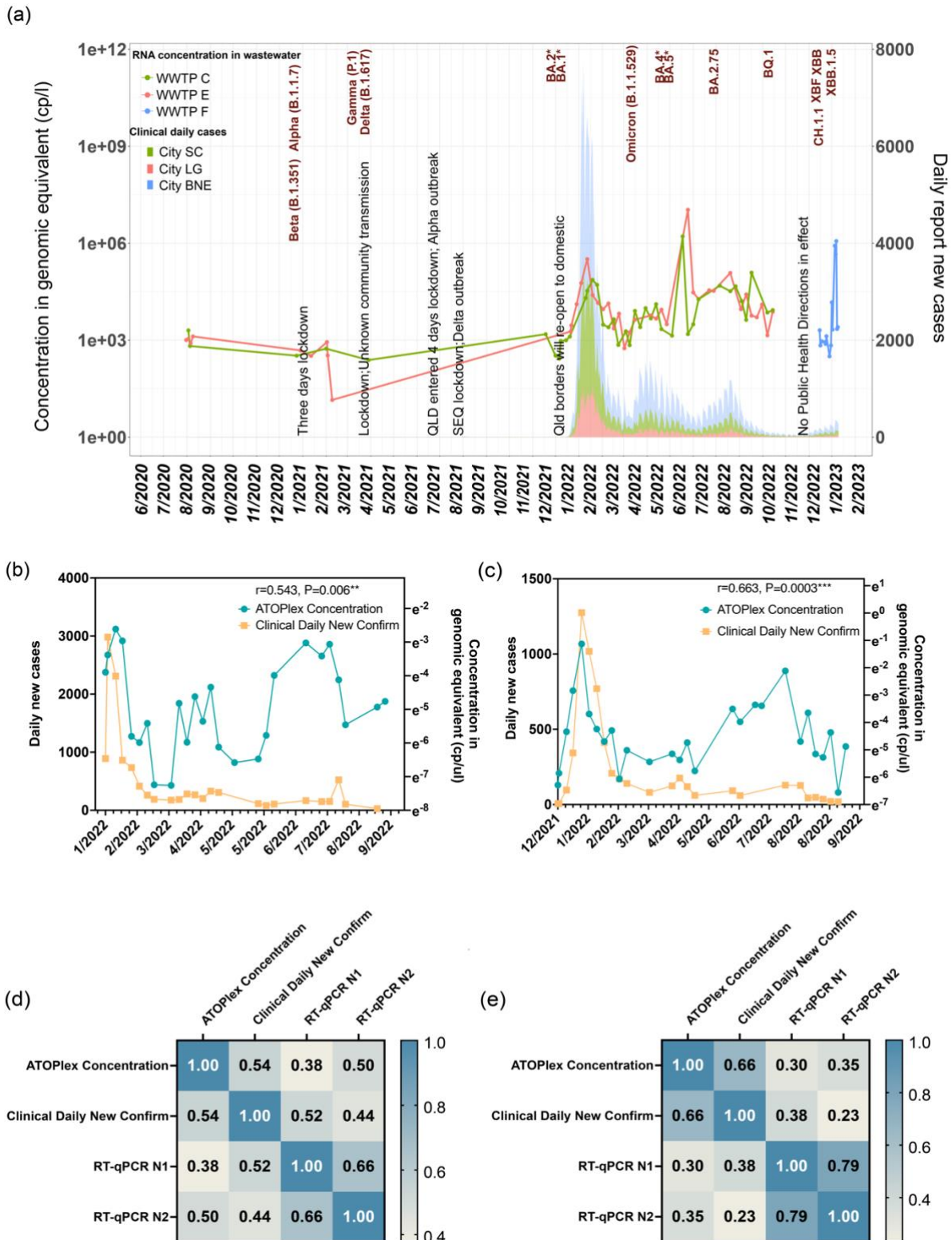
200
201 Figure 3. Example of genome coverage (10X) of the wastewater samples. The 10X genome
202 coverage for WW38, WW39 and WW63 are 14.63%, 50.62 and 99.88% respectively. The x-
203 axis represents the coverage of every single nucleotide position with respect to the reference
204 genome (MN908947.3) and the y-axis represents the sequencing depth at the log2 value at each
205 position.

206 *3.3 Wastewater sequencing reveals the viral load and prevalence trend*

207 To assess the consistency of the disease prevalence between wastewater surveillance and
208 clinical surveillance, we compared RNA concentration estimated by wastewater sequencing

209 with clinical daily new case numbers. In 2020 and 2021, Queensland was marked by strict
210 disease control measures, such as border closures, gathering restrictions, and targeted
211 mandatory vaccinations, thus minimising local transmission (Table S1). However, following
212 the opening of the Queensland border in December 2021, there is a dramatic increase in weekly
213 total cases, with a nearly 150% increase observed in the week after the reopening compared to
214 the week prior (Figure 4-a). Concurrently, a consistent upward trend in RNA concentration was
215 observed in wastewater samples (Figure 4-a). In addition, COVID-19 exhibited four main
216 surges during the study period in terms of both clinical and wastewater surveillance, which is
217 thought to be associated with the emergence of new variants (Figure 4-a). For example, BA.1
218 was first detected in clinical testing in December 2021 and rapidly spread after that. Along with
219 low levels of other variants, BA.1 was observed in all areas inspected in this study, resulting in
220 the first COVID-19 epidemic wave in Queensland (Figure 4-a, Figure S3). Interestingly, the
221 RNA concentration in wastewater samples in February 2022 exceeded the clinical case line on
222 a monthly scale (Figure 4-b, c), indicating a higher estimated prevalence of COVID-19 through
223 wastewater surveillance than clinically confirmed cases.

224 Moreover, in terms of correlation with different clinical epidemiology factors, WBE has a
225 higher performance in estimating daily new cases compared to active or cumulative clinical
226 cases.³³ To evaluate the correlation between RT-qPCR and ATOplex in estimating the the new
227 cases reported each day, Pearson correlation coefficient analysis was employed. The
228 sequencing method exhibited a higher correlation with the daily new cases (Figure 4-d, e), with
229 a $|r|$ equal to 0.54 ($p=0.006$) and 0.663 ($p=0.0003$)) in the SC and LG regions of Queensland,
230 respectively. In contrast, a weaker correlation between the daily new cases with RT-qPCR Ct
231 value targets N1 or N2 ($|r|$ ranging from 0.52~0.23, Figure 4-d,e) was observed. These results
232 suggest the sequencing method provides a more robust and consistent estimate of the
233 prevalence trend compared to RT-qPCR across different regions.



234

235 Figure 4. Epidemiological curves depicting disease prevalence in the study regions. (a) The

236 line graph represents the wastewater epidemiological curve and the area plot shows the clinical

237 daily new confirmed cases. The major public health measures (shown in black), as well as the

238 first detection date of each emerging variant (shown in red) identified through clinical
239 sequencing during the study period are labelled. (b) and (c) zoom in on the curve measured by
240 wastewater sequencing (in blue) and clinical tests (in yellow) in SC and LG regions,
241 respectively. (d) and (e) shows The Pearson correlation coefficient ($|r|$) matrix for each
242 measuring method with the daily new cases in SC and LG.

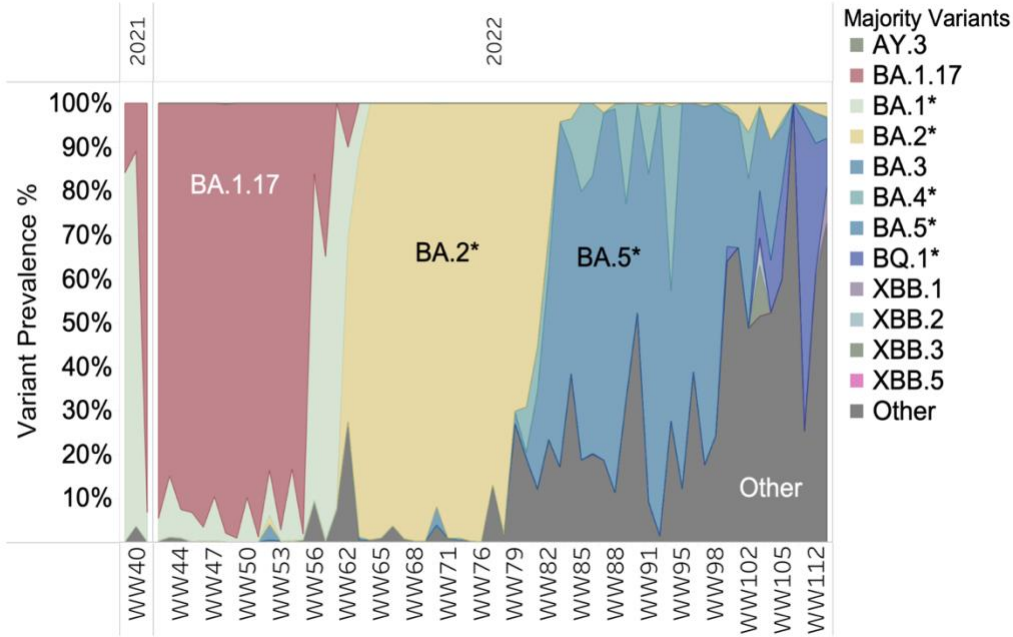
243 ***3.4 Early detection and transition patterns of emerging VOCs***

244 The emergence of SARS-CoV-2 variants with a growth advantage could lead to a greater
245 infectivity or transmissibility and antibody escape, thus likely increasing the incidence rate and
246 supplanting previous circulating variants.³ It is thought that public health measures effectively
247 suppressed the local spread of these emerging variants in 2020/2021, evidenced by the fact that
248 the low number of confirmed cases and the low RNA concentration in wastewater. For
249 example, Brisbane entered three days of lockdown in response to the circulating Alpha variant
250 (B.1.1.7, table S1) in January and late June 2021. Southeast Queensland implemented a 3-day
251 lockdown and further extended it for an additional 5 days following the outbreak of the Delta
252 variant (B.1.617) in August 2021. Correspondingly, these emerging variants were confined to
253 a small cluster due to the timely implementation of measures. However, after the Queensland
254 border reopened, the Omicron and its sublineages (BA*) transmitted rapidly and became
255 dominant in Queensland (Figure S3).

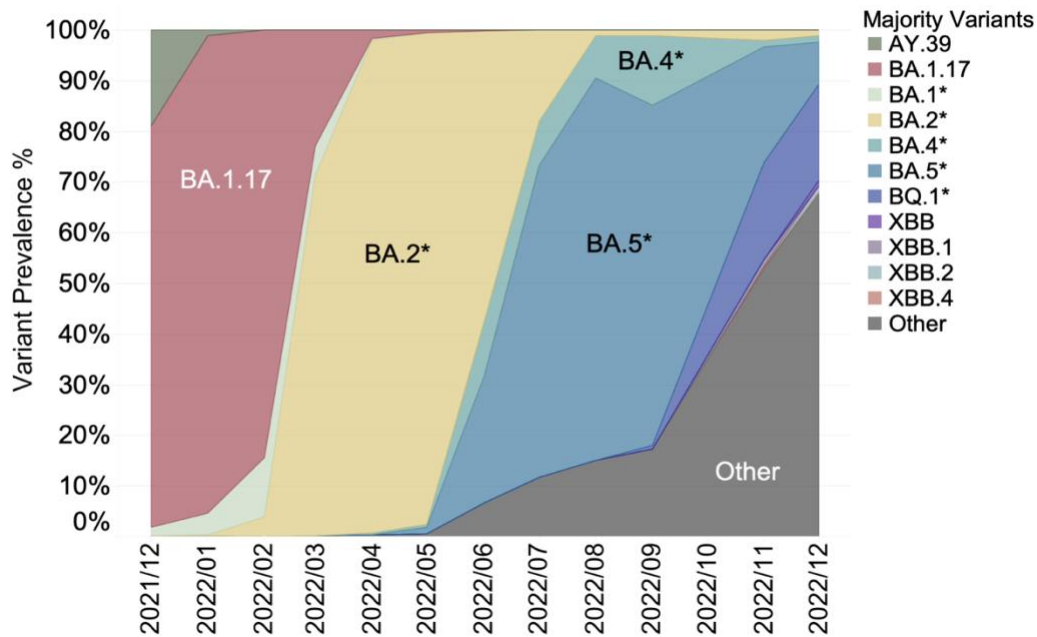
256 In response to the limitation of extensive clinical surveillance, we explored the potential of
257 sentinel sites wastewater sequencing as an alternative strategy to estimate the prevalence of
258 circulating variants at the state level. To accomplish this, we compared the variant dynamics
259 obtained in wastewater samples with that obtained in state-level clinical samples over the study
260 period. Notably, variants detected in wastewater samples showed the main circulating variants
261 were Omicron and its sublineages, with other co-occurring variants (e.g. Gamma and Delta)
262 present in low abundance after the border reopened (Figure 5-a). The prevalence of the BA.2

263 variant replaced the BA.1.17 variant in mid-March 2022, and then later in July the BA.5*
264 became dominant (Figure 5-a). This finding is consistent with the variant prevalence trend
265 shown by clinical sequencing data (Figure 5-b). For example, clinical data showed that BA.2
266 replaced the BA.1.17 variant as the predominant circulating variant in Queensland in mid-
267 March (Figure 5-b). More importantly, the detection of new variants in wastewater samples
268 preceded their identification in clinical settings (Table S2). For instance, the parent Omicron
269 variant (B.1.1.529) was detected in wastewater in February, 2022, but it was reported from
270 clinical sequencing in March, 2022. Further, signals of the BA.2.75 variant were found in
271 wastewater in April, 2022, preceding its local spread which was reported clinically in July,
272 2022. Moreover, the BA.3* variant was identified in wastewater but not in clinical sequencing,
273 revealing the potential cryptic transmission of these variants in Queensland. Similarly, BQ.1
274 was detected in wastewater in late June, while it was not reported clinically until mid-
275 September. It should be noted that we also observed the BA.2 variant was present in wastewater
276 in early February 2022 but it had been observed in mid-January 2022 in confirmed clinical
277 cases (Figure 5-a, b).

(a)



(b)

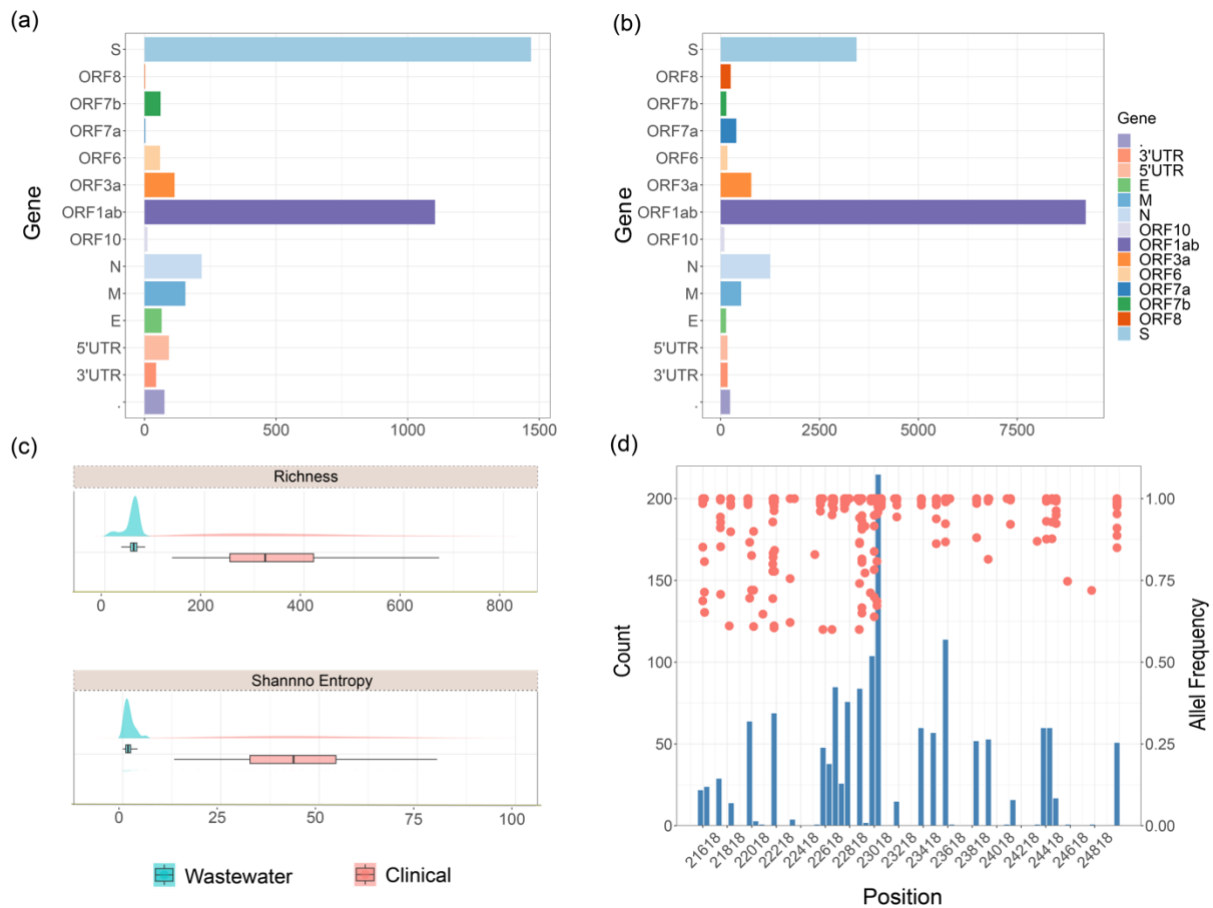


278

279 Figure 5. Prevalence of the SARS-CoV-2 variants identified in wastewater samples (a, n=64
 280 wastewater-derived genomes with 10X coverage >20) and clinical samples (b, n=4878 GISAID
 281 sourced complete clinical genomes) during the experimental period expressed as a percentage.
 282 The different variants detected in wastewater are shown in different colours. The details of
 283 wastewater and clinical epidemiological data can be found in supplementary table S2.

284 **3.5 SARS-CoV-2 genetic diversity in wastewater**

285 The identification of SARS-CoV-2 variants relies heavily on the single nucleotide variants
286 (SNVs) associated with each variant, thus genetic information is vital for monitoring the
287 emergence of new variants.¹ To determine the potential of wastewater sequencing in detecting
288 the evolution of the virus, we calculated the genetic diversity in both wastewater and clinical
289 samples by mapping the sequencing data against the reference genome (MN908947.3). To sum
290 up, the mutation analysis of 65 wastewater samples and 4878 clinical genomes yielded 3473
291 and 17017 mutations, respectively, with a majority of mutations occurring in the S gene and
292 ORF1ab gene (Figure 6-a, b). Genetic diversity metrics of richness and Shannon Entropy
293 suggest the genetic diversity of SARS-CoV-2 in the clinical samples is greater than that in the
294 wastewater samples (Figure 6-c). Figure 6-d depicts the 1469 mutations identified in the spike
295 protein within wastewater samples. Among these mutations, 1155 are present in the S1 gene
296 of the spike protein, which normally poses a high risk of instigating the emergence of novel
297 variants (Figure S4). Notably, despite the higher genetic diversity observed in clinical samples,
298 our analysis identified 28 unique spike protein mutations in wastewater samples, whereas they
299 were absent in the clinical counterparts (Table S3).



300

301 Figure 6. Summary of genetic variations in SARS-CoV-2 genes in wastewater and clinical
 302 samples. Count of mutations that occur on the different SARS-CoV-2 genes in wastewater
 303 samples (a) and clinical samples (b). Genetic diversity of SARS-CoV-2 in wastewater and
 304 clinical samples (c), represented by richness (top) and Shannon Entropy (bottom). Mutations
 305 detected in the Spike protein of SARS-CoV-2 in wastewater samples are shown in (d). The
 306 histogram (bin width = 50 bp) is the count of the number of each mutation that occurs on the
 307 genome position that is aligned with the left y-axis, while the scatter plot represents the allele
 308 frequency of each mutation in the corresponding position aligned with the right y-axis.

309 4. Discussion

310 This study highlights the capability of wastewater sequencing to accurately reflect the disease
 311 prevalence trend and the potential to provide early information on the circulating variant
 312 dynamics. These findings further support the WBE setting as an alternative surveillance

313 program providing long-term, large-scale, and sustainable monitoring of SARS-CoV-2 across
314 a wide geographic area. Moreover, the implementation of WBE as a surveillance tool holds
315 significant promise in improving our preparedness and response strategies for future outbreaks
316 and pandemics beyond SARS-CoV-2.

317 Although the PCR-based methods can provide timely and cost-effective estimation of disease
318 prevalence, it is incapable of targeting emerging variants and providing genetic information to
319 track virus evolution.^{12,15,34–36} In contrast, the ATOplex sequencing method not only enables
320 accurate quantification of the RNA concentration in wastewater but is also capable of
321 identifying the co-occurrence of different VOCs, thus demonstrating wastewater as a valuable
322 tool to monitor disease trends and variant dynamics in the community. This study found a
323 stronger Pearson's correlation between ATOplex sequencing and number of daily new cases
324 compared to the PCR-based method, particularly in the LG region (The correlation coefficients
325 ($|r|$) for ATOplex, N1, and N2 in relation to the daily new cases were found to be 0.66, 0.38,
326 and 0.23, respectively. (Figure 4)). The higher dilution status in the SC region may have
327 resulted in a decreased RNA signal, thus increasing the likelihood of RT-qPCR missing the
328 target.¹² In addition, an increasing disparity between the estimated viral concentrations in
329 wastewater and the daily new cases reported in clinics has been observed. This can partially be
330 explained by the gradual easing of public health measures, especially the widespread use of
331 rapid antigen test (RAT) kits, which has led to a significant decline in the PCR-confirmed cases
332 reported to clinics. Specifically, from mid-January to mid-February 2022, the PCR test number
333 administered per 1,000 individuals per week in Queensland dropped from over 50 to
334 approximately 15, which may account for the decrease in PCR-confirmed cases from 1,334.8
335 to 258.6 per 100,000 population.³⁷ However, the correlation coefficient ($|r|$) between RNA
336 concentration in wastewater and clinical parameters (e.g. daily or active total cases) varied in
337 different studies, ranging from 0.14 to 1.³³ This might be attributed to various evaluation

338 conditions, such as environmental factors, sampling strategies and epidemiological factors.³³
339 In addition, the Pearson correlation coefficient may be biased due to increased biases in the
340 clinical data. Even though ATOPlex sequencing exhibits a higher $|r|$ compared to the PCR
341 method, further research is required to assess the potential impact of various factors on the
342 strength of the correlation coefficient. Additionally, there is a need to optimize the prediction
343 model to better understand the disease prevalence in the community.

344 Moreover, with millions of people still getting infected and thousands of deaths globally to
345 date, ongoing investigations into the impact of COVID-19 variants and intervention policies
346 (e.g., booster vaccination), are crucial due to the continuously changing circumstances. This
347 longitudinal study demonstrates the potential of wastewater sequencing to track the transition
348 of different variants, which aligns with clinical epidemiological reports in Queensland,
349 Australia (Figure 5-a, b). Additionally, monitoring mutations in the SARS-CoV-2 genome is
350 vital for informing health surveillance and response strategies, particularly in identifying and
351 monitoring emerging variants.²⁵ WBE as a means of monitoring mutations occurring in the
352 SARS-CoV-2 genome is practical where large-scale clinical sequencing is unfeasible. Similar
353 to clinical sequencing, the majority of mutations detected in wastewater were found in the
354 receptor-binding domain (RBD) region of the spike protein, which has been identified as a key
355 factor in the emergence of new variants (Figure 6-a,d).^{38,39} For example, the L452R mutation
356 enhances the immune escape fitness and virulence of the Delta variant, while its absence in the
357 Omicron variant reduces its pathogenicity.³⁸ Furthermore, unique mutations were observed in
358 multiple wastewater samples, revealing the hidden cryptic circulating variants in the
359 community. These findings require careful interpretation to better understand the
360 characteristics of SARS-CoV-2.

361 While obtaining high-quality genomes from wastewater samples is still challenging due to
362 factors such as RNA degradation or low concentration, ATOPlex sequencing can recover

363 complete or partial genomes from the wastewater samples, thus providing valuable
364 retrospective epidemiological insights. However, certain regions with low sequencing depth
365 were observed in the recently collected wastewater samples (Figure S2), underscoring the
366 necessity to optimize the ATOplex primer set used for amplicon sequencing to enable the
367 investigation of emerging variants.²⁵ Additionally, while retrospective analysis demonstrated
368 the feasibility of early detection of emerging variants, it was observed that the genetic diversity
369 obtained from regional wastewater was lower compared to the comprehensive clinical
370 sequencing conducted across the entire state. This finding highlights the significance of
371 selecting sampling sentinel sites with precision, particularly in areas characterized by high
372 traffic or population density, as this is essential to ensure a comprehensive understanding of
373 the circulating variants in the entire state. Future studies should focus on identifying
374 appropriate sampling sites and establishing long-term routine wastewater sequencing
375 monitoring, aiding effective public health decision-making and outbreak prevention.

376 **5. Conclusions**

377 This study aims to investigate the performance of using sentinel wastewater surveillance to
378 track the variant dynamics at the state level. In this regard, we established a high-throughput,
379 multiplex, amplicon-based sequencing technology, namely ATOplex, to discern multiple
380 variants from a longitudinal wastewater surveillance program in Queensland, Australia. The
381 key conclusions are as follows:

- 382 • Comparison with a qPCR-based method showed that ATOplex sequencing provides a more
383 reliable snapshot of disease prevalence in a community. With the decrease in clinical tests,
384 this method can assist public health sectors in comprehending the prevailing trends.

385 • Wastewater sequencing can track the variant dynamics over time and the results are
386 comparable to the clinical sequencing results. It also has the potential to detect new variants
387 before their proliferation, thus supporting the prevention of pandemics.

388 • Mutations can be identified in the recovered genome showing the ability to track virus
389 evolution on a large scale.

390 These findings collectively suggest that sequencing-based WBE is a cost-effective and timely
391 alternative strategy for monitoring diseases within a community.

392 **Acknowledgement**

393 JG would like to acknowledge the Australian Research Council for funding support through
394 Discovery Project (DP220101526). GN would like to acknowledge support from the Advance
395 Queensland Industry Research Fellowship (RM2019002600) and an Early Career Postdoctoral
396 Fellowship (ECPF23-8566329039) at the Faculty of Medicine, Nursing and Health Sciences
397 of Monash University. We thank MGI Australia for sequencing support. The authors thank
398 City of Gold Coast, Urban Utilities and Logan Water for their help in collecting samples. We
399 thank Queensland Health for their help with data retrieval.

400 **Notes**

401 The authors declare no competing financial interest.

402

403 **References**

404 1. Sapoval, N. *et al.* Enabling accurate and early detection of recently emerged SARS-CoV-
405 2 variants of concern in wastewater. *Nat. Commun.* **14**, 2834 (2023).

- 406 2. Bonanno Ferraro, G. *et al.* A State-of-the-Art Scoping Review on SARS-CoV-2 in
407 Sewage Focusing on the Potential of Wastewater Surveillance for the Monitoring of the
408 COVID-19 Pandemic. *Food Environ. Virol.* (2021) doi:10.1007/s12560-021-09498-6.
- 409 3. Brito, A. F. *et al.* Global disparities in SARS-CoV-2 genomic surveillance. *Nat.*
410 *Commun.* **13**, 7003 (2022).
- 411 4. Levy, J. I., Andersen, K. G., Knight, R. & Karthikeyan, S. Wastewater surveillance for
412 public health. *Science* **379**, 26–27 (2023).
- 413 5. Karthikeyan, S. *et al.* Wastewater sequencing reveals early cryptic SARS-CoV-2 variant
414 transmission. *Nature* **609**, 101–108 (2022).
- 415 7. Wolfe, M. *et al.* Detection of SARS-CoV-2 Variants Mu, Beta, Gamma, Lambda, Delta,
416 Alpha, and Omicron in Wastewater Settled Solids Using Mutation-Specific Assays Is
417 Associated with Regional Detection of Variants in Clinical Samples. *Appl. Environ.*
418 *Microbiol.* e0004522 (2022) doi:10.1128/aem.00045-22.
- 419 8. Schmitz, B. W. *et al.* Enumerating asymptomatic COVID-19 cases and estimating SARS-
420 CoV-2 fecal shedding rates via wastewater-based epidemiology. *Sci. Total Environ.* **801**,
421 149794 (2021).
- 422 9. Ahmed, W. *et al.* First confirmed detection of SARS-CoV-2 in untreated wastewater in
423 Australia: A proof of concept for the wastewater surveillance of COVID-19 in the
424 community. *Sci. Total Environ.* **728**, 138764 (2020).
- 425 10. Sherchan, S. P. *et al.* First detection of SARS-CoV-2 RNA in wastewater in North
426 America: A study in Louisiana, USA. *Sci. Total Environ.* **743**, 140621 (2020).
- 427 11. La Rosa, G. *et al.* First detection of SARS-CoV-2 in untreated wastewaters in Italy. *Sci.*
428 *Total Environ.* **736**, 139652 (2020).
- 429 12. Li, J. *et al.* Monitoring of SARS-CoV-2 in sewersheds with low COVID-19 cases using a
430 passive sampling technique. *Water Res.* **218**, 118481 (2022).

- 431 13. Sellers, S. C. *et al.* Building-level wastewater surveillance of SARS-CoV-2 is associated
432 with transmission and variant trends in a university setting. *Environ. Res.* 114277 (2022)
433 doi:10.1016/j.envres.2022.114277.
- 434 14. Corchis-Scott, R. *et al.* Averting an Outbreak of SARS-CoV-2 in a University Residence
435 Hall through Wastewater Surveillance. *Microbiol. Spectr.* **9**, e0079221 (2021).
- 436 15. Medema, G., Heijnen, L., Elsinga, G., Italiaander, R. & Brouwer, A. Presence of SARS-
437 Coronavirus-2 RNA in Sewage and Correlation with Reported COVID-19 Prevalence in
438 the Early Stage of the Epidemic in The Netherlands. *Environ. Sci. Technol. Lett.* **7**, 511–
439 516 (2020).
- 440 16. Le, T. T. B. *et al.* SARS-CoV-2 Omicron and its current known unknowns: A narrative
441 review. *Rev. Med. Virol.* **33**, e2398 (2023).
- 442 17. Laine, P. *et al.* SARS-CoV-2 variant with mutations in N gene affecting detection by
443 widely used PCR primers. *J. Med. Virol.* **94**, 1227–1231 (2022).
- 444 18. Yaniv, K. *et al.* Direct RT-qPCR assay for SARS-CoV-2 variants of concern (Alpha,
445 B.1.1.7 and Beta, B.1.351) detection and quantification in wastewater. *Environ. Res.* **201**,
446 111653 (2021).
- 447 19. Oloye, F. F. *et al.* Rapid transition between SARS-CoV-2 variants of concern Delta and
448 Omicron detected by monitoring municipal wastewater from three Canadian cities. *Sci.*
449 *Total Environ.* **841**, 156741 (2022).
- 450 20. Li, J. *et al.* Impact of sewer biofilms on fate of SARS-CoV-2 RNA and wastewater
451 surveillance. *Nat. Water* **1**, 272–280 (2023).
- 452 21. Freed, N. E., Vlková, M., Faisal, M. B. & Silander, O. K. Rapid and inexpensive whole-
453 genome sequencing of SARS-CoV-2 using 1200 bp tiled amplicons and Oxford
454 Nanopore Rapid Barcoding. *Biol. Methods Protoc.* **5**, bpaa014 (2020).

- 455 22. Wang, Y. *et al.* Detection of SARS-CoV-2 Variants of Concern with Tiling Amplicon
456 Sequencing from Wastewater. *ACS EST Water* (2022) doi:10.1021/acsestwater.2c00083.
- 457 23. Jahn, K. *et al.* Early detection and surveillance of SARS-CoV-2 genomic variants in
458 wastewater using COJAC. *Nat. Microbiol.* **7**, 1151–1160 (2022).
- 459 24. Kaya, D. *et al.* Correlation between Clinical and Wastewater SARS-CoV-2 Genomic
460 Surveillance, Oregon, USA. *Emerg. Infect. Dis.* **28**, 1906–1908 (2022).
- 461 25. Agrawal, S. *et al.* Prevalence and circulation patterns of SARS-CoV-2 variants in
462 European sewage mirror clinical data of 54 European cities. *Water Res.* **214**, 118162
463 (2022).
- 464 26. Bar-Or, I. *et al.* Detection of SARS-CoV-2 variants by genomic analysis of wastewater
465 samples in Israel. *Sci. Total Environ.* **789**, 148002 (2021).
- 466 27. Crits-Christoph, A. *et al.* Genome Sequencing of Sewage Detects Regionally Prevalent
467 SARS-CoV-2 Variants. *mBio* **12**, e02703-20 (2021).
- 468 28. Ni, G. *et al.* Novel Multiplexed Amplicon-Based Sequencing to Quantify SARS-CoV-2
469 RNA from Wastewater. *Environ. Sci. Technol. Lett.* **8**, 683–690 (2021).
- 470 29. Pastorino, B., Touret, F., Gilles, M., de Lamballerie, X. & Charrel, R. N. MANUAL FOR
471 DETECTION OF SARS-CoV-2 RNA IN WASTEWATER. *Viruses* **12**, 735 (2020).
- 472 30. Ahmed, W. *et al.* RT-qPCR and ATOPlex sequencing for the sensitive detection of
473 SARS-CoV-2 RNA for wastewater surveillance. *Water Res.* **220**, 118621 (2022).
- 474 31. Khare, S. *et al.* GISAID’s Role in Pandemic Response. *China CDC Wkly.* **3**, 1049–1051
475 (2021).
- 476 32. Page, A. J. *et al.* SNP-sites: rapid efficient extraction of SNPs from multi- FASTA
477 alignments. *Microb. Genomics.*

- 478 33. Li, X. *et al.* Correlation between SARS-CoV-2 RNA concentration in wastewater and
479 COVID-19 cases in community: A systematic review and meta-analysis. *J. Hazard.*
480 *Mater.* **441**, 129848 (2023).
- 481 34. Li, L. *et al.* Longitudinal monitoring of SARS-CoV-2 in wastewater using viral genetic
482 markers and the estimation of unconfirmed COVID-19 cases. *Sci. Total Environ.* **817**,
483 152958 (2022).
- 484 35. Li, L. *et al.* Detecting SARS-CoV-2 Variants in Wastewater and Their Correlation With
485 Circulating Variants in the Communities. *Res. Sq.* rs.3.rs-1435729 (2022)
486 doi:10.21203/rs.3.rs-1435729/v1.
- 487 36. Zheng, X. *et al.* Quantification of SARS-CoV-2 RNA in wastewater treatment plants
488 mirrors the pandemic trend in Hong Kong. *Sci. Total Environ.* **844**, 157121 (2022).
- 489 37. COVID-19 National Incident Room Surveillance Team. COVID-19 Australia:
490 Epidemiology Report 59 Reporting period ending 13 March 2022. *Commun. Dis. Intell.*
491 **46**, (2022).
- 492 38. Zhang, Y. *et al.* SARS-CoV-2 spike L452R mutation increases Omicron variant
493 fusogenicity and infectivity as well as host glycolysis. *Signal Transduct. Target. Ther.* **7**,
494 1–3 (2022).
- 495 39. Barut, G. T. *et al.* The spike gene is a major determinant for the SARS-CoV-2 Omicron-
496 BA.1 phenotype. *Nat. Commun.* **13**, 5929 (2022).
- 497