

1 **Wastewater genomic surveillance tracks the spread of the SARS-CoV-2 Omicron** 2 **variant across England**

3
4 Authors:

5 Franziska S. Brunner¹, Alexander Payne², Edward Cairns¹, George Airey¹, Richard
6 Gregory¹, Natalie D. Pickwell², Myles Wilson², Matthew Carlile², Nadine Holmes², Verity Hill³,
7 Harry Child⁴, Jasmine Tomlinson⁴, Suhel Ahmed⁴, Hubert Denise⁵, William Rowe⁵, Jacob
8 Frazer⁵, Ronny van Aerle⁶, Nicholas Evens⁷, Jonathan Porter⁷, The COVID-19 Genomics UK
9 (COG-UK) Consortium⁸, Kate Templeton⁹, Aaron R. Jeffries⁴, Matt Loose², Steve Paterson¹

10
11 ¹Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool CH64 7TE,
12 UK

13 ²Deep Seq, Centre for Genetics and Genomics, The University of Nottingham, Queen's
14 Medical Centre, Nottingham, NG7 2UH, UK

15 ³Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New
16 Haven, CT, USA

17 ⁴Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter EX4
18 4QD, UK

19 ⁵Environmental Monitoring for Health Protection, UK Health Security Agency, Nobel House,
20 London SW1P 3HX, UK

21 ⁶International Centre of Excellence for Aquatic Animal Health, Cefas, Barrack Road,
22 Weymouth, DT 8UB, UK

23 ⁷Monitoring Laboratories, National Monitoring, Environment Agency EX6 8FD, UK

24 ⁸<https://www.cogconsortium.uk>, Full list of consortium names and affiliations are in the
25 appendix

26 ⁹NHS Lothian, Edinburgh, UK
27
28

29 **Abstract**

30 31 *Background*

32 Many countries have moved into a new stage of managing the SARS-CoV-2 pandemic with
33 minimal restrictions and reduced testing in the population, leading to reduced genomic
34 surveillance of virus variants in individuals. Wastewater-based epidemiology (WBE) can
35 provide an alternative means of tracking virus variants in the population but is lacking
36 verifications of its comparability to individual testing data.

37 *Methods*

38 We analysed more than 19,000 samples from 524 wastewater sites across England at least twice
39 a week between November 2021 and February 2022, capturing sewage from >70% of the English
40 population. We used amplicon-based sequencing and the phylogeny based de-mixing tool Freyja
41 to estimate SARS-CoV-2 variant frequencies and compared these to the variant dynamics
42 observed in individual testing data from clinical and community settings.

43 *Findings*

44 We show that wastewater data can reconstruct the spread of the Omicron variant across England
45 since November 2021 in close detail and aligns closely with epidemiological estimates from
46 individual testing data. We also show the temporal and spatial spread of Omicron within London.
47 Our wastewater data further reliably track the transition between Omicron subvariants BA1 and
48 BA2 in February 2022 at regional and national levels.

BA2 in February 2022 at regional and national levels. Peer review and should not be used to guide clinical practice.

49 *Interpretation*

50 Our demonstration that WBE can track the fast-paced dynamics of SARS-CoV-2 variant
51 frequencies at a national scale and closely match individual testing data in time shows that
52 WBE can reliably fill the monitoring gap left by reduced individual testing in a more affordable
53 way.

54 *Funding*

55 Department of Health and Social Care, UK, Natural Environmental Research Council, UK, COG-
56 UK

57

58 **Research in context**

59

60 *Evidence before this study*

61 Genomic monitoring of wastewater for SARS-CoV-2 variants has been introduced in several
62 countries and shown to effectively detect the spread of known variants in multiple studies.
63 However, verification of its alignment with individual testing data at a national scale has so far
64 been reported only for Austria, where sampling covered around 5.4million people. Further and
65 larger scale verifications of the reliability of wastewater-based epidemiology (WBE) are
66 needed to increase confidence in its use for public health monitoring.

67 *Added value of this study*

68 We provide evidence that WBE was able to closely track the spread of the emerging SARS-
69 CoV-2 variant Omicron, as well as its sub lineage dynamics, at a regional and national scale
70 across England. Our sampling covered >70% of the English population, equivalent to 39.4
71 million people. We thereby demonstrate the scalability of our approach to national levels. We
72 also show how WBE is able to track dynamics in different regions of the UK and at a finer
73 scale within London. Its close alignment, in estimated epidemiological timings, with results
74 from intensive individual testing in the same timeframe provides evidence that wastewater-
75 based monitoring can be a reliable alternative when large scale data from individual testing is
76 not available.

77

78 *Implications of all the available evidence*

79 Altogether, evidence is accumulating that WBE is a reliable approach for monitoring SARS-
80 CoV-2 variant dynamics and informing public health measures across spatial scales.

81

82 **Introduction**

83 As lockdowns are easing across the world, the threat from SARS-Cov-2, and especially new
84 variants, is still considerable. Therefore reliable, sustainable, and affordable ways to monitor
85 the ongoing evolution of the virus and spread of variants are needed.

86 Wastewater-based epidemiology (WBE) is a method to monitor pathogens that pose a threat
87 to public health and has been receiving much interest during the SARS-CoV-2 pandemic ¹⁻³,
88 exploiting the observation that around 50% of infected individuals shed the virus in faeces
89 ^{4,5}.Through sampling of wastewater (sewage) from catchments containing large numbers of
90 individuals, followed by detection of pathogens using molecular methods, WBE can provide
91 regular surveillance of a high percentage of a population ⁶. This provides passive sampling
92 that can encompass individuals who are asymptomatic or who have low engagement with
93 clinical testing programmes ^{1,7}. Many local and national WBE programs have been initiated
94 with the prospect of providing insight into the prevalence and diversity of the virus in different
95 communities and to detect the emergence and spread of SARS-CoV-2 variants, for example

96 in Spain⁸, Switzerland⁹, India¹⁰ and the USA^{11–13} and have informed local and national public
97 health interventions.

98 These programs have shown promising results, demonstrating that WBE can detect the
99 occurrence of emerging variants, often even earlier than individual testing^{9,14,15} and showing
100 differences in the timing of the spread of SARS-CoV-2 variants at a small spatial scale, such
101 as parts of a city^{16,17}. However, WBE verification studies have generally been focused on one
102 or a few cities, with only one published example of SARS-CoV-2 variant monitoring in
103 wastewater at a larger spatial scale across Austria¹⁸. The wastewater monitoring program in
104 England had expanded WBE to an unrivalled scale, covering 39.4 million people or 70% of
105 the population by mid-2021 and providing critical insights into uncertainty factors involved with
106 wastewater sampling and their mitigation¹⁹. This extensive wastewater sampling coverage,
107 combined with a comprehensive individual testing effort in the UK which can be used for
108 validation, provides an ideal case study to verify the utility of WBE for following the spread of
109 a new variant across a whole country.

110 The SARS-CoV-2 Omicron variant (B.1.1.529) was first designated as a variant of concern
111 (VOC) by The World Health Organisation on 21st November 2021, following detection in a
112 rapidly growing cluster of South African cases^{20,21}. The Omicron variant is characterised by a
113 large number of mutations, including at sites associated with functional differences on the
114 spike protein, and appears to have arisen from a distinct lineage to that of the previously
115 dominant Delta variant²². Following Omicron's emergence, it rapidly spread across the globe
116 causing increased Covid-19 case numbers in many countries and was quickly determined to
117 have substantially higher immune evasion than any previous variant, potentially combined with
118 a higher transmissibility²¹, leading to a strong selective advantage. By mid-January it had
119 already accounted for more than 50% of new infections reported to GISAID and was
120 responsible for more than 90% of new infections detected by individual testing in the UK²³.

121 Here we use the arrival and spread of Omicron in England to test the utility of WBE in tracking
122 SARS-CoV-2 variants at a national level. We use genome sequencing of SARS-CoV-2 from
123 >19,000 longitudinal samples and obtained from 524 sites, including sewage treatment works
124 and finer scale sampling from the sewer networks in cities. We use these data to test how
125 quickly WBE was able to detect the arrival of Omicron in England and its ability to then give
126 insight into Omicron's geographical spread and rise in frequency through time, verifying results
127 against data from individual testing efforts in clinical and non-clinical settings.

128

129 **Materials and Methods**

130

131 *Wastewater Sample Collection and Processing*

132 Wastewater grab samples (1 L per sample) were collected from 233 locations across the
133 sewer network in England and from 291 wastewater treatment plants between the 1st of
134 November 2021 and the 28th of February 2022, as part of the ongoing Environmental
135 Monitoring for Health Protection programme (part of NHS Test & Trace, now the UK Health
136 Security Agency) in England. Samples were transported and subsequently stored at 4 - 6°C
137 until analysis, minimising RNA degradation. Within 24h of collection, all samples were
138 centrifuged (10,000 x g, 20°C, 10 min) in sterile PPCO bottles to remove suspended solids.
139 The supernatant (150 ml) was transferred to 250 ml PPCO bottles containing 60 g of
140 ammonium sulfate (Sigma-Aldrich, Cat. No. A4915). After the ammonium sulfate had
141 dissolved, the samples were incubated at 4 °C for 1 h before further centrifugation (10,000xg,
142 4 °C, 30 min) and supernatant removal. The pellet was resuspended in 2ml of NucliSens lysis

143 buffer (BioMérieux, Marcy-l'Etoile, France, Cat No. 280134 or 200292). Concentrates were
144 stored at 4°C until nucleic acid extraction. Nucleic acids were extracted from concentrates
145 using NucliSens extraction reagent kit (BioMérieux, Cat. No. 200293) either manually (Farkas
146 et al. 2021) or using the KingFisher 24 Flex system (Thermo Scientific, Waltham, MA, USA)
147 according to the manufacturer instructions ²⁴, eluting RNA extracts with a 120 µl volume.
148 Extracts were stored at -80°C until further processing.

149 *Wastewater Sequencing*

150 Wastewater RNA extracts were purified and sequenced with a standardised EasySeq™ RC-
151 PCR SARS-CoV-2 (Nimagen) V1.0 protocol ²⁵. Briefly, samples were cleaned with Mag-Bind®
152 TotalPure NGS beads (Omega Bio-Tek) and then reverse transcribed using LunaScript® RT
153 SuperMix Kit (New England Biolabs) and the EasySeq™ RC-PCR SARS-CoV-2 (novel
154 coronavirus) Whole Genome Sequencing kit v3.0 (NimaGen). Amplicons were pooled and
155 libraries cleaned with Mag-Bind® Total Pure NGS beads (Omega Bio-Tek) or Ampure XP
156 beads (Beckman Coulter) before sequencing on an Illumina NovaSeq™ 6000 or NextSeq 500
157 platform generating 2x150bp paired end reads. Sample processing and sequencing were
158 performed in 3 different laboratories (Exeter, Liverpool and Nottingham), validated to produce
159 comparable results by sequencing of standardised synthetic samples (Fig. S1). In brief, 47
160 mixes of 4 different synthetic SARS-CoV-2 RNA Controls (Twist Bioscience) in different
161 proportions were sequenced at each of the sites. We validated reliability between sites by both
162 comparing the frequencies of individual SNPs detected and by comparing the predicted
163 lineage composition from our analysis pipeline from each of the sites with the expected lineage
164 composition based on the known mix of variants.

165 *Wastewater Bioinformatics*

166 Following sequencing, the wastewater data were demultiplexed and processed using a
167 custom Nextflow v21.04.0 ²⁶ pipeline (https://github.com/LooseLab/ww_nf_minimal). Briefly,
168 reads were filtered using fastp v0.20.1 ²⁷ and mapped to the SARS-CoV-2 reference
169 (accession: NC_045512.2) using BWA (v0.7.17; <http://arxiv.org/abs/1303.3997>). Following
170 alignment, primer sequences were trimmed using iVar v1.3.1 ²⁸ and a bed file containing the
171 amplicon primer positions. Bam files containing the trimmed alignments then underwent
172 variant (SNP) calling using VarScan v2.4.4 ²⁹ and relative SARS-CoV-2 lineage abundance
173 estimation using Freyja (v1.3.1; <https://github.com/andersen-lab/Freyja>; curated lineage file
174 and UShER global phylogenetic tree downloaded on 2022-07-12). Variant calls were
175 aggregated and subsequent analyses were conducted using custom scripts in R and Python
176 (available on github). The specific SARS-CoV-2 variants (lineages) predicted by Freyja to be
177 present in the samples were summarised into major lineages, i.e. defined variants of concern
178 (VOC) and variants under investigation (VUI) according to the Pango nomenclature
179 (<https://cov-lineages.org>).

180

181 *Sample quality control and method verification*

182 To control for variant detection issues in low quality samples, we excluded wastewater
183 samples from further analyses if coverage depth was ≤ 100x for amplicon 121, which contains
184 7 of the defining mutations of the Omicron variant, or if less than 100 out of 154 amplicons
185 had at least 20x coverage depth. This reduced our dataset from 19,911 samples to 9,883
186 samples (Fig S2b).

187 We further checked these quality thresholds against the SARS-CoV-2 RNA concentration
188 found in the wastewater samples initially by qPCR (Fig.S2c&d).

189 We used a presence/absence call on characteristic Omicron mutations to validate the results
190 obtained with Freyja for first detection and general detection of Omicron presence in a given
191 area (Fig S3), using the following definitions by Public Health England: The B.1.1.529-BA.1
192 lineage has 1 synonymous and 16 non-synonymous signature SNPs of which 11 are unique
193 amongst known VOC and VUI, detection confirmed when 11 of 16 signature SNPs and more
194 than 6 of 11 unique SNPs detected and co-occurrence of SNPs on the same amplicon
195 detected, or when more than 11 signature SNPs and more than 6 unique SNPs are present.
196 The B.1.1.529-BA.2 lineage has 4 synonymous and 16 non-synonymous signature SNPs of
197 which 12 are unique amongst known VOC and VUI, detection confirmed when more than 13 of
198 20 signature SNPs and 6 of 12 unique SNPs detected and co-occurrence on the same
199 amplicon detected, or when more than 13 signature SNPs and more than 6 unique SNPs are
200 present.

201 We also used Varscan v2.4.4²⁹ to call SNPs, averaging frequencies across SNPs matching
202 the Omicron profile to validate frequency estimates obtained by Freyja (Fig S4).

203 *Individual infection sequence metadata*

204 Sequence metadata was obtained from the master COG-UK dataset on 29th March 2022,
205 generated each day using a custom pipeline, datapipe (<https://github.com/COG-UK/datapipe>).
206 As part of this pipeline, SARS-CoV-2 variants were assigned using Scorpio
207 (<https://github.com/cov-lineages/scorpio>), and only sequences with sampling dates after the
208 1st November 2021 and published less than three weeks after the sequencing date were used
209 for Omicron sequences to avoid including incorrectly assigned sequences. The same sample
210 date cut-off as for wastewater samples of 28th February 2021 was used. Geographical
211 locations were also cleaned as part of the analysis pipeline, and details of geographical
212 metadata cleaning can be found here (https://github.com/COG-UK/geography_cleaning).
213 Individual infection data came from two pillars of testing employed in England – pillar 1 data
214 coming from hospital patients and health and care workers, pillar 2 data coming from swab
215 tests in the wider population. We tested the alignment of data from pillar 1 and pillar 2 with
216 wastewater data separately and as pooled data.

217 *Statistical Analyses*

218 We used R version 4.1.2³⁰ and RStudio v. 1.4.1717³¹ and daily data for all statistical analyses
219 unless stated otherwise. For geographical visualisations of the data, we calculated weekly
220 averages of variant frequencies for each site. For increased visibility on the maps, we then
221 aggregated data by county for both the individual infection data and wastewater data. We
222 plotted maps with complete colouring of the respective counties for better legibility, maps of
223 only the specific areas covered by the sampled wastewater network (covering the main
224 population aggregations) can be found in the Supplement (Fig S5). Shapefiles for mapping
225 were obtained from gadm.org (county delineations) and the maps package in R³². We
226 generated maps of sampling area outlines in Python, using the libraries geopandas v0.11.0,
227 pyplot v3.5.2 and contextily v1.2.0. These outlines are depicted as approximate, not exact
228 outlines due to data sharing restrictions.

229 We modelled the rise of the Omicron variant in each of the 9 regions of England separately
230 for wastewater and individual infection data by fitting a two-parameter log-logistic function with
231 lower limit 0 and upper limit 1 by least squares estimation using the drm function from the drc
232 package v3.0-1³³.

$$233 \quad f(x) = \frac{1}{1 + \exp(b(\log(x) - \log(e)))}$$

234 To compare the estimated rates of Omicron spreading in each region and according to
235 different data types, we extracted from each model the date when Omicron was predicted to
236 reach dominance (inflection point, e) and the growth rate during the exponential growth phase
237 which we transformed to a more intuitive growth rate of the function from 5 days before to 5
238 days past the inflection point as $g = \frac{f(e+5)-f(e-5)}{10}$.

239 To assess the time match between Omicron frequency estimates in wastewater and individual
240 infection data, we ran Spearman's correlation tests in the range of time delays of 5 day lags
241 of wastewater data to 3 day lags of individual infection data, using county level daily average
242 frequency data.

243

244 **Results**

245 *Detection limits*

246 By comparing synthetic mixes of SARS-CoV-2 variants in each of our sequencing facilities,
247 we could confirm that variant levels as low as 1% were detected as being present in all
248 samples by our sequencing approaches and Freyja based analyses (Fig.S1). Samples with
249 SARS-CoV-2 concentrations above 100,000 genome copies/L consistently yielded
250 sequencing data above our quality criteria, however many samples with lower RNA
251 concentration also yielded sufficient sequencing reads, including many samples where SARS-
252 CoV-2 RNA was not detected by qPCR initially (Fig.S2).

253 *Detection and frequency estimation of Omicron*

254 We first detected the Omicron variant in wastewater samples from the South East and the
255 East Midlands on the 26th of November 2021, verified with two different approaches using a
256 minimum number of characteristic SNPs and the phylogeny based de-mixing tool Freyja
257 (Fig.S3). The first individual cases from clinical and community testing (referred to as individual
258 testing from here onwards) of Omicron were confirmed by sequencing on the 20th and 22nd
259 of November 2021 in Essex and Greater London, respectively. The following rapid sweep of
260 the variant through England throughout December 2021 (Fig.1) was first evident in wastewater
261 in the London area, where Omicron frequencies reached >25% in the first week of December
262 (Fig.1d) and log-logistic growth models fitted to the regional data indicated that the Omicron
263 variant represented the majority of virus particles in wastewater by the 12th of December
264 (Fig.2a). By contrast, we did not detect Omicron frequencies over 25% in the South West of
265 England until mid December (Fig.1) and dominance of the variant was only predicted from the
266 21st of December (Fig.2a). We also found a matching decline in the Delta variant across the
267 country, validating our results (Fig.S6). Unfortunately, reduced availability of samples over the
268 Christmas period meant that some areas of the country had no frequency estimates from
269 wastewater samples during the week of the 20th of December. Variant frequency estimates
270 with Freyja were corroborated by estimates based on Varscan, with a maximum discrepancy
271 of 4 days in the estimated date when Omicron reached dominance in any given region
272 (Fig.S4).

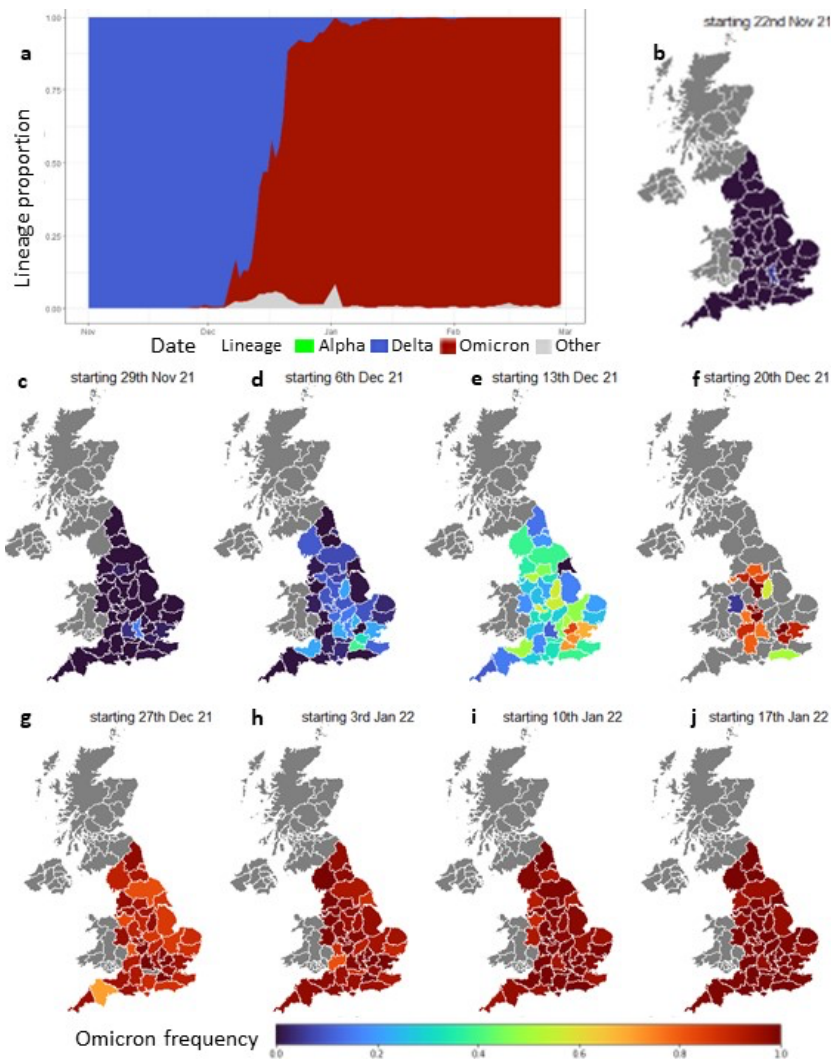
273 A closer look at the London area also revealed finer scale differences in the spread of Omicron
274 (Fig.3). The first notable levels of Omicron appeared in the East of the city in the week of the
275 29th of November, but subsequently the central parts of the city saw a quicker rise in Omicron
276 prevalence to over 25% in the week of the 6th of December and also earlier full dominance
277 (>50% frequency) of the variant in the following week.

278 *Matching wastewater base results and individual testing results*

279 Individual testing data from clinical and community PCR testing closely matched our
280 wastewater results, with the Omicron variant becoming the majority of detections in the
281 London area by mid-December and the rest of England following suit about a week later
282 (Fig.2). When modelling the logistic growth of Omicron frequencies at the region level, the time
283 point estimates for Omicron becoming the dominant variant in wastewater varied between a 1
284 day lead on individual testing estimates (North East) to a 6 day lag (South East) (Fig.2a).
285 Frequency growth rate estimates for Omicron were higher in wastewater data in five of nine
286 regions, with rates of 0.052-0.065 frequency change/day in individual testing and 0.047-0.097
287 frequency change/day in wastewater (Fig.2b). Growth rate estimates were highest for
288 wastewater data from the North East, however this was also the data subset with lowest
289 sample numbers around the transition phase in December, making the rate estimate less
290 reliable. Daily Omicron frequencies aggregated at county level correlated strongly between
291 individual testing and wastewater data across a range of investigated timing lags (Table S1),
292 with the strongest correlation observed when no lag was assumed between wastewater and
293 individual testing data (Spearman's $\rho = 0.905$, Fig.2c). Examination of the different types of
294 individual testing data from Pillar 1 data (collected from patients and staff in hospitals) and
295 Pillar 2 data (collected in the community outside hospitals) revealed a good alignment with
296 wastewater estimates in both data types (Fig.S7). Pillar 1 data yielded slightly earlier estimates
297 of Omicron dominance, with a predicted lead time of 6 days or more on wastewater data in six
298 out of nine regions, whereas Pillar 2 data was predicted to lead on wastewater data by 6 days
299 in only one region (Fig.S7a). Combining data for all of England, we found the strongest
300 correlation for a 2 day lag of wastewater based frequencies behind Pillar 1 data (Spearman's
301 $\rho = 0.871$, Fig.S7b). We found greater variability between regions in the estimates for Pillar 1
302 data, which is likely related to much lower sample numbers and thereby increased
303 stochasticity.

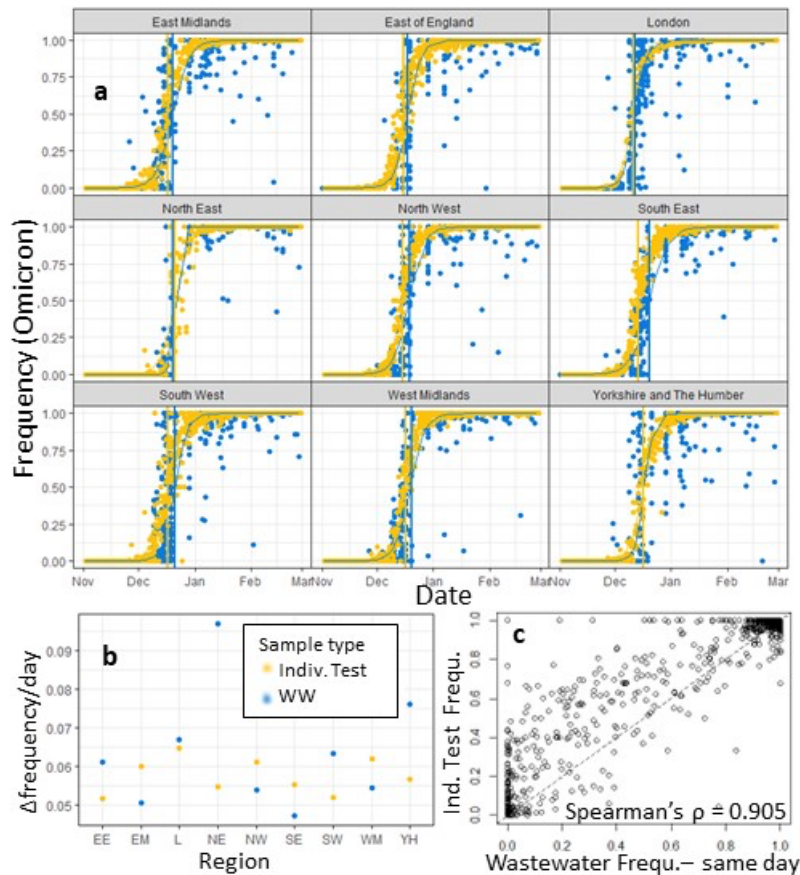
304 *Sublineage patterns*

305 We further tested the sensitivity of our lineage detection in wastewater by reconstructing the
306 shift from subvariant BA1 to subvariant BA2 of the Omicron lineage and found a clear shifting
307 pattern from BA1 to BA2 in wastewater samples for each of the nine regions through February
308 (Fig.4).



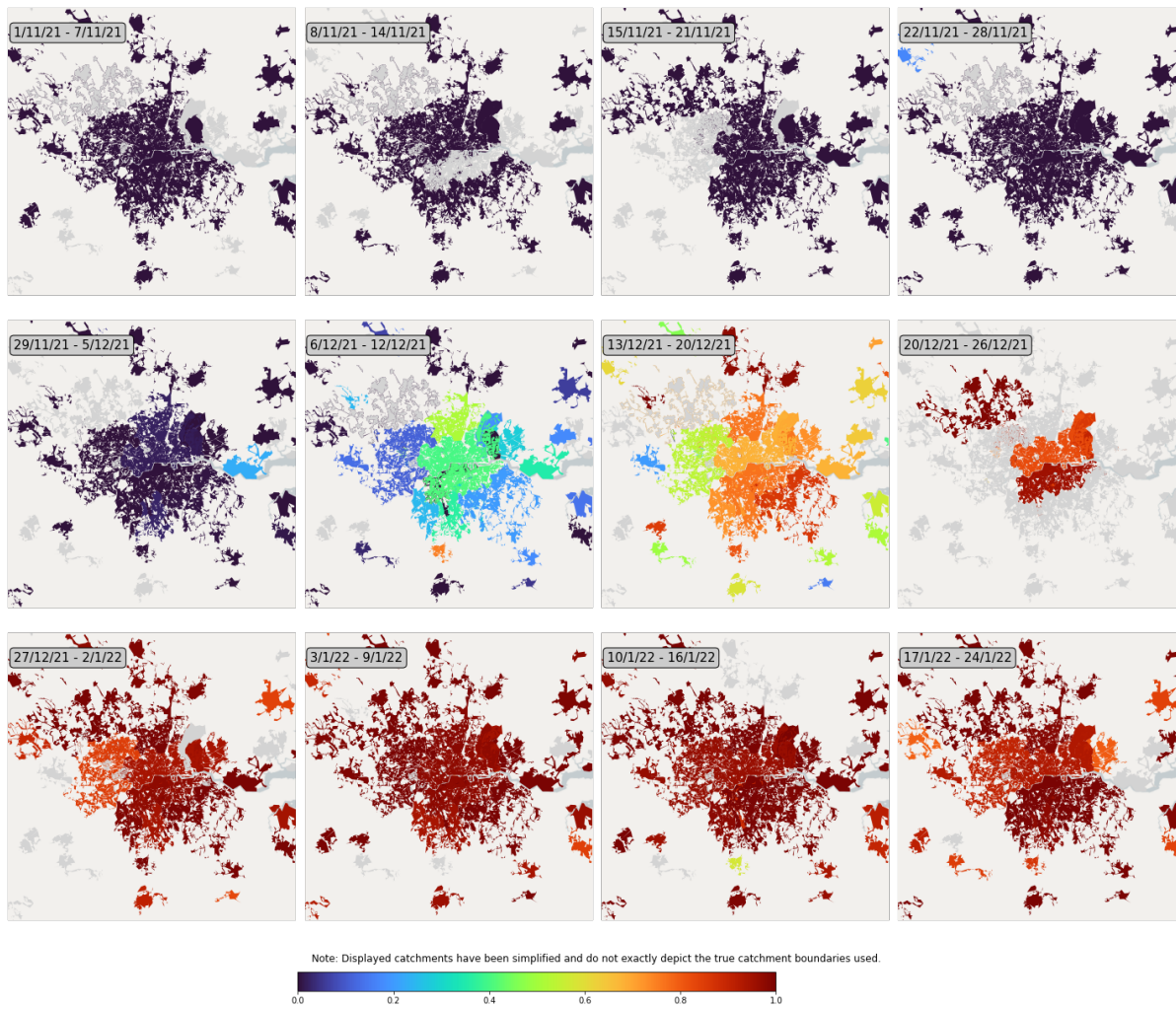
309
310
311
312
313
314
315
316

Figure 1: **a** Proportions of major lineages detected in wastewater across England between
wastewater detection of Omicron variant frequency, averaged
by county and calendar week, between late November 2021 and mid-January 2022. Grey
areas indicate missing data, due to low sample quality or lack of sampling.

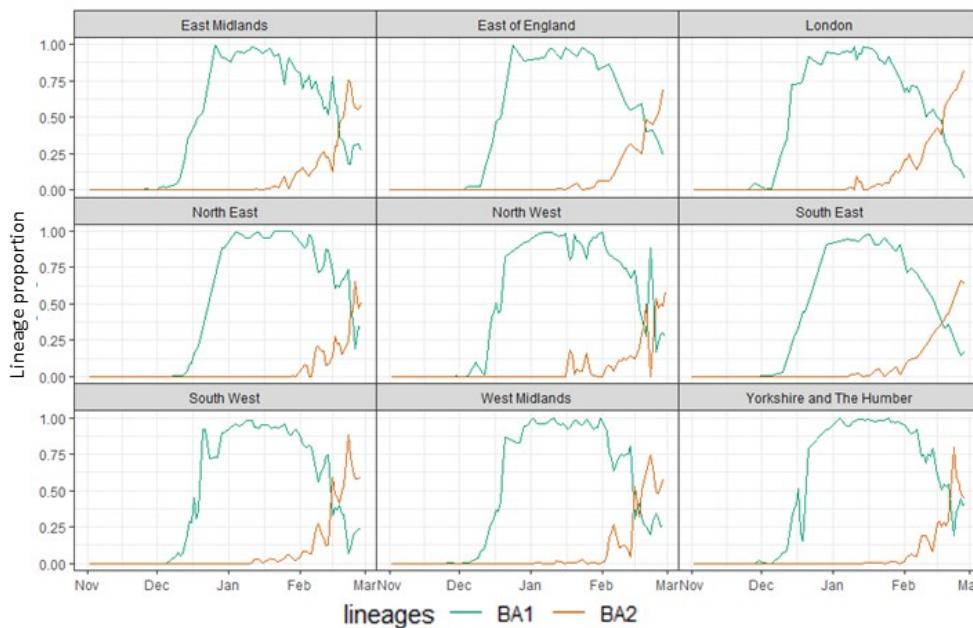


317
 318 Figure 2: **a** Log-logistic growth models of the Omicron frequencies in 9 English regions.
 319 Vertical lines indicate the inflection points of the models which estimate the timepoints when
 320 Omicron frequencies reach 50% in wastewater (blue) and individual testing (yellow) samples.
 321 **b** growth rates of the log-logistic growth models in each region in wastewater (blue) and
 322 individual testing (yellow) samples, calculated as predicted frequency change between 5 days
 323 before and 5 days after inflection point. **c** Spearman's correlation tests show strongest
 324 correlation between county level daily Omicron frequency in individual testing and wastewater
 325 with no time lag between frequency estimates.

326
 327
 328
 329



330
 331 Figure 3: Weekly Omicron frequencies within London. Depicted are approximate outlines of
 332 the wastewater catchment area captured by sampling. Grey areas indicate missing data, due
 333 to low sample quality or lack of sampling.



334
 335 Figure 4: Frequencies of the subvariants BA1 and BA2 of the Omicron variant between

336 November 2021 and February 2022, shown as 3 day rolling averages of frequencies out of all
337 detected variants in wastewater samples.

338

339 Discussion

340 We show that tracking of the SARS-CoV-2 Omicron variant by WBE closely mirrors that seen
341 by mass community surveillance (i.e., the sequencing of PCR positive individuals) and
342 mapped the rise of Omicron through time, with the earliest rise seen in London followed by
343 the rest of England. Estimated Omicron frequencies from the two approaches are highly
344 correlated, with no overall lag time between WBE and community surveillance by sample date,
345 despite small differences in estimated growth rates of Omicron and transition timings from
346 Delta to Omicron when looking at different regions within England. If we consider only samples
347 taken in a hospital setting, we see an even earlier rise of the Omicron variant, however this
348 comparison is weakened by the much more sparse data from the hospital environment
349 compared to wider community testing. Furthermore, we show that even sublineage dynamics
350 such as the transition from Omicron subvariant BA1 to BA2 during February 2022³⁴ can readily
351 be detected with our WBE approach.

352 The ability to detect the presence of an emerging variant – regardless of frequency – may be
353 of particular utility for local public health teams. WBE has the advantage that it is able to
354 monitor infections in areas exhibiting low engagement with community testing programmes,
355 or where not all PCR-positive infections are genome sequenced or from asymptomatic
356 infections⁷. On the other hand, WBE has disadvantages in a lack of individually identifiable
357 information, which limits the potential for public health teams to investigate where cases live
358 or work or whether they are recent travellers to an area. Here we find that initial detections
359 differ between individual testing and WBE. For example, the first instances of detecting the
360 Omicron variant in November 2021 in wastewater were in the East Midlands and South East
361 whereas the first individual infections with the variant were confirmed in Essex and Greater
362 London. This is perhaps unsurprising given that detecting the first cases of a new variant in
363 an area, either by individual testing or WBE, is inherently stochastic. Using both techniques
364 would therefore maximise the likelihood of capturing these early events. During late 2021, the
365 UK also tested travellers entering the country, with the aim to detect and prevent the spread
366 of imported variants. However, in March 2022 all travel restrictions into the UK were removed
367 and individual testing in the wider community (Pillar 2 approach in England) reduced to
368 minimal levels, in an approach similar to that of many other countries. Our data indicate that
369 WBE could be used to detect the initial spread of variants from around 1% prevalence. The
370 use of WBE could therefore represent a viable alternative for genomic surveillance with
371 minimal bias, low costs and demonstrated sensitivity.

372 Our demonstration of a close alignment in nation-wide WBE and individual testing data also
373 shows that some inherent differences between these two types of samples may be less
374 relevant than previously thought. Wastewater samples capture mostly faecal shedding of virus
375 particles and a meta-analysis has shown that only an average of 43% of Covid-19 patients
376 shed detectable levels of faecal SARS-CoV-2 RNA and that the duration of particle shedding
377 in faeces exceeds that in respiratory samples by weeks⁵. If this impacted the proportion
378 between SARS-CoV-2 variants, we would expect to see different variant dynamics between
379 individual sampling and wastewater, especially in the transition phase when a new variant
380 spreads across a country. However, our data show a close alignment in variant dynamics
381 nonetheless, suggesting that no major bias in shedding exists between variants at least and
382 that WBE provides a robust measure of variant proliferation at local and national scales.

383 The comparative logistic requirements of community sampling and WBE merit discussion. The
384 data presented here represent around 1k-2.3k samples per week for wastewater sampling
385 and around 30k-48k samples per week for community sampling, which come from RNA left
386 over from PCR testing. In the context of the UK testing programme at the time, 1-3.5m PCR
387 tests were conducted each week, of which around 10-32% were positive. Testing at this
388 national scale is a substantial effort that is likely impossible to sustain indefinitely as SARS-
389 CoV-2 becomes endemic and has already been abolished in many countries. Sequencing
390 from only clinical samples, e.g. hospitalised patients, is an alternative, but this may bias
391 sampling against younger age-groups; who are least likely to be hospitalised but among whom
392 transmission may be highest. Furthermore, reduced individual surveillance leads to later
393 detection of newly emerging or imported variants in a country. Wastewater sampling has a
394 significant set up cost, in developing agreements with water companies, installation of
395 autosamplers and logistics to transport samples to laboratories and then in overcoming
396 technical challenges of extracting SARS-Cov-2 RNA from wastewater at scale. On the other
397 hand, it can provide high coverage of a population using fewer resources than, and without
398 the bias associated with, community testing. Around 70% of England was covered by the WBE
399 analysis presented here involving sampling over 200 sewage network site and close to 300
400 sewage treatment plants. Optimisation of sampling site distribution could potentially provide
401 similar high coverage of the population while reducing the number of sites sampled. As shown
402 in samples from the London sewer network, it also provides insights into the emergence or
403 rise of new variants and how this varies across a city. This means that wastewater monitoring
404 of SARS-CoV-2 evolution is a feasible and cost-effective option for keeping track of emerging
405 local hotspots of the disease and the spread of new variants, even when large-scale individual
406 testing is wound down. The wastewater samples used, and the general approach of targeted
407 pathogen sequencing can also be used for ongoing surveillance of other pathogens to help
408 protect public health, as evidenced already for polio and monkeypox^{35,36}.
409 Overall, we have demonstrated the usefulness and reliability of WBE for tracking an emerging
410 variant from its first detection to its spread across a whole country. This method can therefore
411 be considered a reliable tool for monitoring large scale genetic dynamics in wide-spread
412 diseases.

413

414 **Acknowledgements**

415 We thank B. Jones, R. Crompton, T. Foster and N. Cadu for assistance with laboratory work.
416 We thank partners across the Environmental Monitoring for Health Protection programme
417 (EMHP) in England for the coordination and collection of wastewater samples. This includes
418 the UK Health Security Agency (UKHSA), water companies in England, the Environment
419 Agency (EA), Centre for Environment, Fisheries and Aquaculture Science (Cefas) and the
420 Department for Environment, Food and Rural Affairs (Defra).
421 Funding was provided by NERC (NE/V003860/1) and DHSC UK (2020_097). This report is
422 independent research funded by the Department of Health and Social Care. COG-UK is
423 supported by funding from the Medical Research Council (MRC) part of UK Research &
424 Innovation (UKRI), the National Institute of Health Research (NIHR) [grant code:
425 MC_PC_19027], and Genome Research Limited, operating as the Wellcome Sanger Institute.
426 The authors acknowledge use of data generated through the COVID-19 Genomics
427 Programme funded by the Department of Health and Social Care. The views expressed are
428 those of the authors and not necessarily those of the Department of Health and Social Care
429 or UKHSA.

430

431 **CRedit Author contributions**

432 **Franziska S Brunner** - conceptualisation, data curation, formal analysis, investigation,
433 methodology, project administration, validation, visualisation, writing – original draft, writing –
434 review & editing

435 **Alexander Payne** - conceptualisation, data curation, investigation, methodology, resources,
436 software, validation, writing – review & editing

437 **Edward Cairns** - investigation, methodology, project administration

438 **George Airey** - investigation, methodology, project administration

439 **Richard Gregory** - project administration, resources, software

440 **Natalie D Pickwell** - investigation, methodology, project administration

441 **Nadine Holmes** – investigation, methodology

442 **Myles Wilson** - investigation, methodology

443 **Matthew Carlile** - investigation, methodology

444 **Verity Hill** - investigation, methodology, software, writing – review & editing

445 **Harry Child** - investigation, methodology, project administration

446 **Jasmine Tomlinson** - investigation, methodology

447 **Suhel Ahmed** - investigation, methodology

448 **Hubert Denise** - investigation, methodology, writing – review & editing

449 **William Rowe** - investigation, methodology

450 **Jacob Frazer** - investigation, methodology

451 **Ronny van Aerle** - investigation, methodology, writing – review & editing

452 **Nicholas Evens** - investigation, methodology, supervision, writing – review & editing

453 **Jonathan Porter** - investigation, methodology, supervision, writing – review & editing

454 **The COVID-19 Genomics UK (COG-UK) Consortium** - funding acquisition, investigation,
455 methodology, project administration, resources, software

456 **Kate Templeton** - conceptualisation, funding acquisition, supervision

457 **Aaron R. Jeffries** - conceptualisation, funding acquisition, investigation, methodology,
458 supervision, writing – review & editing

459 **Matt Loose** - conceptualisation, funding acquisition, investigation, methodology, supervision,
460 writing – review & editing

461 **Steve Paterson** - conceptualisation, funding acquisition, investigation, methodology,
462 supervision, writing – review & editing

463

464 **Conflict of interest statement**

465 The authors declare no conflicts of interest.

466

467 **Data accessibility statement**

468 Wastewater sequencing data will be made publicly available on the European Nucleotide
469 Archive under Study ID PRJEB55313. The clinical case data used in this study are visualised
470 at <https://www.cogconsortium.uk/tools-analysis/public-data-analysis-2/>. A filtered, privacy
471 conserving version of the lineage-LTLA-week dataset is publicly available online
472 (<https://covid19.sanger.ac.uk/downloads>) and gives access to almost all used data, despite a
473 small number of cells having been suppressed to conserve patient privacy.

474 **Ethics Statement**

475 Use of surplus nucleic acid derived from routine diagnostics and associated patient data was
476 approved through the COG-UK consortium by the Public Health England Research Ethics and
477 Governance Group (R&D NR0195).

478

479 **References**

- 480 1 Farkas K, Hillary LS, Malham SK, McDonald JE, Jones DL. Wastewater and public health: the
481 potential of wastewater surveillance for monitoring COVID-19. *Curr Opin Environ Sci Health*
482 2020; **17**: 14–20.
- 483 2 Hillary LS, Farkas K, Maher KH, *et al.* Monitoring SARS-CoV-2 in municipal wastewater to
484 evaluate the success of lockdown measures for controlling COVID-19 in the UK. *Water Res*
485 2021; **200**: 117214–117214.
- 486 3 Peccia J, Zulli A, Brackney DE, *et al.* Measurement of SARS-CoV-2 RNA in wastewater tracks
487 community infection dynamics. *Nat Biotechnol* 2020; **38**: 1164–7.
- 488 4 Foladori P, Cutrupi F, Segata N, *et al.* SARS-CoV-2 from faeces to wastewater treatment: What
489 do we know? A review. *Sci Total Environ* 2020; **743**: 140444.
- 490 5 Zhang Y, Cen M, Hu M, *et al.* Prevalence and Persistent Shedding of Fecal SARS-CoV-2 RNA in
491 Patients With COVID-19 Infection: A Systematic Review and Meta-analysis. *Clin Transl*
492 *Gastroenterol* 2021; **12**: e00343.
- 493 6 Morvan M, Jacomo AL, Souque C, *et al.* An analysis of 45 large-scale wastewater sites in
494 England to estimate SARS-CoV-2 community prevalence. *Nat Commun* 2022; **13**: 4313.
- 495 7 Polo D, Quintela-Baluja M, Corbishley A, *et al.* Making waves: Wastewater-based
496 epidemiology for COVID-19 – approaches and challenges for surveillance and prediction.
497 *Water Res* 2020; **186**: 116404.
- 498 8 Carcereny A, Martínez-Velázquez A, Bosch A, *et al.* Monitoring Emergence of the SARS-CoV-2
499 B.1.1.7 Variant through the Spanish National SARS-CoV-2 Wastewater Surveillance System
500 (VATar COVID-19). *Environ Sci Technol* 2021; **55**: 11756–66.
- 501 9 Jahn K, Dreifuss D, Topolsky I, *et al.* Detection and surveillance of SARS-CoV-2 genomic
502 variants in wastewater. *medRxiv* 2021. DOI:10.1101/2021.01.08.21249379.
- 503 10 Dharmadhikari T, Rajput V, Yadav R, *et al.* High throughput sequencing based direct detection
504 of SARS-CoV-2 fragments in wastewater of Pune, West India. *Sci Total Environ* 2022; **807**:
505 151038.
- 506 11 Crits-Christoph A, Kantor RS, Olm MR, *et al.* Genome Sequencing of Sewage Detects
507 Regionally Prevalent SARS-CoV-2 Variants. 2021; **12**: 9.
- 508 12 Fontenele RS, Kraberger S, Hadfield J, *et al.* High-throughput sequencing of SARS-CoV-2 in
509 wastewater provides insights into circulating variants. *Water Res* 2021; **205**: 117710.
- 510 13 Smyth DS, Trujillo M, Gregory DA, *et al.* Tracking cryptic SARS-CoV-2 lineages detected in
511 NYC wastewater. *Nat Commun* 2022; **13**: 635.

- 512 14 Markt R, Endler L, Amman F, *et al.* Detection and abundance of SARS-CoV-2 in wastewater in
513 Liechtenstein, and the estimation of prevalence and impact of the B.1.1.7 variant. *J Water*
514 *Health* 2021; : jwh2021180.
- 515 15 Xiao A, Wu F, Bushman M, *et al.* Metrics to relate COVID-19 wastewater data to clinical testing
516 dynamics. *Water Res* 2022; **212**: 118070.
- 517 16 Brunner FS, Brown MR, Bassano I, *et al.* City-wide wastewater genomic surveillance through
518 the successive emergence of SARS-CoV-2 Alpha and Delta variants. *Infectious Diseases*
519 (except HIV/AIDS), 2022 DOI:10.1101/2022.02.16.22269810.
- 520 17 Rios G, Lacoux C, Leclercq V, *et al.* Monitoring SARS-CoV-2 variants alterations in Nice
521 neighborhoods by wastewater nanopore sequencing. *Lancet Reg Health - Eur* 2021; **10**:
522 100202.
- 523 18 Amman F, Markt R, Endler L, *et al.* Viral variant-resolved wastewater surveillance of SARS-
524 CoV-2 at national scale. *Nat Biotechnol* 2022; published online July 18. DOI:10.1038/s41587-
525 022-01387-y.
- 526 19 Wade MJ, Lo Jacomo A, Armenise E, *et al.* Understanding and managing uncertainty and
527 variability for wastewater monitoring beyond the pandemic: Lessons learned from the
528 United Kingdom national COVID-19 surveillance programmes. *J Hazard Mater* 2022; **424**:
529 127456.
- 530 20 Karim SSA, Karim QA. Omicron SARS-CoV-2 variant: a new chapter in the COVID-19
531 pandemic. *The Lancet* 2021; **398**: 2126–8.
- 532 21 Viana R, Moyo S, Amoako DG, *et al.* Rapid epidemic expansion of the SARS-CoV-2 Omicron
533 variant in southern Africa. *Nature* 2022; **603**: 679–86.
- 534 22 Qin S, Cui M, Sun S, *et al.* Genome Characterization and Potential Risk Assessment of the
535 Novel SARS-CoV-2 Variant Omicron (B.1.1.529). *Zoonoses* 2021; **1**.
536 DOI:10.15212/ZOONOSES-2021-0024.
- 537 23 Sharma V, Rai H, Gautam DNS, Prajapati PK, Sharma R. Emerging evidence on Omicron
538 (B.1.1.529) SARS-CoV-2 variant. *J Med Virol* 2022; **94**: 1876–85.
- 539 24 Kevill JL, Pellett C, Farkas K, *et al.* A comparison of precipitation and filtration-based SARS-
540 CoV-2 recovery methods and the influence of temperature, turbidity, and surfactant load in
541 urban wastewater. *Sci Total Environ* 2022; **808**: 151916.
- 542 25 Jeffries A, Paterson S, Loose M, van Aerle R. Wastewater Sequencing using the EasySeq™ RC-
543 PCR SARS CoV-2 (Nimagen) V1.0. *protocols.io* 2021; : 8.
- 544 26 Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables
545 reproducible computational workflows. *Nat Biotechnol* 2017; **35**: 316–9.
- 546 27 Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor.
547 *Bioinformatics* 2018; **34**: i884–90.
- 548 28 Grubaugh ND, Gangavarapu K, Quick J, *et al.* An amplicon-based sequencing framework for
549 accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol* 2019;
550 **20**: 8.

- 551 29 Koboldt DC, Chen K, Wylie T, *et al.* VarScan: variant detection in massively parallel
552 sequencing of individual and pooled samples. *Bioinformatics* 2009; **25**: 2283–5.
- 553 30 R Core Team. R: A Language and Environment for Statistical Computing. 2021.
554 <https://www.R-project.org/>.
- 555 31 RStudio Team. RStudio: Integrated Development Environment for R. 2021.
556 <http://www.rstudio.com/>.
- 557 32 Becker RA, Wilks AR, Brownrigg R, Minka TP, Deckmyn A. maps: Draw Geographical Maps.
558 2021. <https://CRAN.R-project.org/package=maps>.
- 559 33 Ritz C, Baty F, Streibig JC, Gerhard D. Dose-Response Analysis Using R. *PLOS ONE* 2015; **10**:
560 e0146021.
- 561 34 Chadeau-Hyam M, Tang D, Eales O, *et al.* The Omicron SARS-CoV-2 epidemic in England
562 during February 2022. *Epidemiology*, 2022 DOI:10.1101/2022.03.10.22272177.
- 563 35 Wastewater monitoring comes of age. *Nat Microbiol* 2022; **7**: 1101–2.
- 564 36 Wolfe MK, Duong D, Hughes B, Chan-Herur V, White BJ, Boehm AB. Detection of monkeypox
565 viral DNA in a routine wastewater monitoring program. *Infectious Diseases (except*
566 *HIV/AIDS)*, 2022 DOI:10.1101/2022.07.25.22278043.
- 567