

## City-wide wastewater genomic surveillance through the successive emergence of SARS-CoV-2 Alpha and Delta variants

Brunner, F. S.; Brown, M. R.; Bassano, I.; Denise, H.; Khalifa, M. S.; Wade, M. J.; van Aerle, R.; Kevill, J. L.; Jones, D. L.; Farkas, K.; Jeffries, A. R.; Cairns, E.; Wierzbicki, C.; Paterson, S.

### Water research

DOI:

[10.1016/j.watres.2022.119306](https://doi.org/10.1016/j.watres.2022.119306)

Published: 01/11/2022

Publisher's PDF, also known as Version of record

[Cyswllt i'r cyhoeddiad / Link to publication](#)

*Dyfyniad o'r fersiwn a gyhoeddwyd / Citation for published version (APA):*

Brunner, F. S., Brown, M. R., Bassano, I., Denise, H., Khalifa, M. S., Wade, M. J., van Aerle, R., Kevill, J. L., Jones, D. L., Farkas, K., Jeffries, A. R., Cairns, E., Wierzbicki, C., & Paterson, S. (2022). City-wide wastewater genomic surveillance through the successive emergence of SARS-CoV-2 Alpha and Delta variants. *Water research*, 226, Article 119306. <https://doi.org/10.1016/j.watres.2022.119306>

### Hawliau Cyffredinol / General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



# City-wide wastewater genomic surveillance through the successive emergence of SARS-CoV-2 Alpha and Delta variants

F.S. Brunner<sup>a, #</sup>, M.R. Brown<sup>b, c, #</sup>, I. Bassano<sup>b, d</sup>, H. Denise<sup>b</sup>, M.S. Khalifa<sup>b, e</sup>, M.J. Wade<sup>b, c</sup>, R. van Aerle<sup>f</sup>, J.L. Kevill<sup>f</sup>, D.L. Jones<sup>f, h</sup>, K. Farkas<sup>g</sup>, A.R. Jeffries<sup>i</sup>, The COVID-19 Genomics UK (COG-UK) Consortium<sup>j</sup>, E. Cairns<sup>a</sup>, C. Wierzbicki<sup>a</sup>, S. Paterson<sup>a, \*</sup>

<sup>a</sup> Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool, L69 7ZB, UK

<sup>b</sup> Environmental Monitoring for Health Protection, UK Health Security Agency, Nobel House, London SW1P 3HX, UK

<sup>c</sup> School of Engineering, Newcastle University, Newcastle-upon-Tyne NE1 7RU, UK

<sup>d</sup> Department of Infectious Disease, Imperial College London, London SW7 2AZ, UK

<sup>e</sup> Division of Biosciences, College of Health, Medicine and Life Sciences, Brunel University, London, UB8 3PH, UK.

<sup>f</sup> International Centre of Excellence for Aquatic Animal Health, Centre for Environment, Fisheries & Aquaculture Science, Dorset, DT4 8UB, UK

<sup>g</sup> Centre for Environmental Biotechnology, School of Natural Sciences, Bangor University, Bangor, Gwynedd LL57 2UW, UK

<sup>h</sup> Food Futures Institute, Murdoch University, Murdoch WA 6105, Australia

<sup>i</sup> Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4QD, UK

## ARTICLE INFO

### Keywords:

COVID-19

Wastewater-based epidemiology

Public health monitoring

Coronavirus variants

Wastewater sequencing

## ABSTRACT

Genomic surveillance of SARS-CoV-2 has provided a critical evidence base for public health decisions throughout the pandemic. Sequencing data from clinical cases has helped to understand disease transmission and the spread of novel variants. Genomic wastewater surveillance can offer important, complementary information by providing frequency estimates of all variants circulating in a population without sampling biases. Here we show that genomic SARS-CoV-2 wastewater surveillance can detect fine-scale differences within urban centres, specifically within the city of Liverpool, UK, during the emergence of Alpha and Delta variants between November 2020 and June 2021. Furthermore, wastewater and clinical sequencing match well in the estimated timing of new variant rises and the first detection of a new variant in a given area may occur in either clinical or wastewater samples. The study's main limitation was sample quality when infection prevalence was low in spring 2021, resulting in a lower resolution of the rise of the Delta variant compared to the rise of the Alpha variant in the previous winter. The correspondence between wastewater and clinical variant frequencies demonstrates the reliability of wastewater surveillance. However, discrepancies in the first detection of the Alpha variant between the two approaches highlight that wastewater monitoring can also capture missing information, possibly resulting from asymptomatic cases or communities less engaged with testing programmes, as found by a simultaneous surge testing effort across the city.

## 1. Introduction

Genomic surveillance has been a significant feature in the public health response to the SARS-CoV-2 pandemic (Wu et al., 2020; Zhu et al., 2020) because of its ability to detect the emergence of and track new variants of concern (VOC) (Robishaw et al., 2021). Important

examples include the B.1.1.7 (Rambaut et al., 2020), B.1.351 (Tegally et al., 2021), P.1 (Faria et al., 2021), B.1.617.2 (Cherian et al., 2021) and, most recently, the B.1.1.529 lineage (Qin et al., 2021), named VOC Alpha, Beta, Gamma, Delta, and Omicron, respectively. These VOC have demonstrated one or more attributes out of increased transmissibility, more severe disease, a reduction in antibody neutralisation, reduced

**Abbreviations:** VOC, Variant of concern; VUI, Variant under investigation; WWTP, Wastewater treatment plant; WBE, Wastewater based epidemiology; SNP, Single nucleotide polymorphism; qRT-PCR, Quantitative reverse transcriptase polymerase chain reaction.

\* Corresponding author.

E-mail address: [s.paterson@liverpool.ac.uk](mailto:s.paterson@liverpool.ac.uk) (S. Paterson).

# These authors contributed equally to the project.

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

<https://doi.org/10.1016/j.watres.2022.119306>

Received 14 July 2022; Received in revised form 13 October 2022; Accepted 26 October 2022

Available online 28 October 2022

0043-1354/© 2022 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

therapeutic response or reduced vaccine effectiveness (Davies et al., 2021; Robishaw et al., 2021). Thus, while vaccination currently provides substantial protection against all known VOC, continued genomic surveillance is essential to mitigate and contain their threat to public health. It informs the implementation and assessment of non-pharmaceutical interventions (e.g., social distancing, lockdowns, and regional, national, and international restrictions) and targeted surge testing. It also serves as an early warning system for the emergence and spread of novel variants (Mishra et al., 2021; Tegally et al., 2021).

Genomic surveillance of SARS-CoV-2 has primarily been driven by whole genome sequencing of clinical isolates, typically using residual RNA from diagnostic RT-qPCR tests. One million SARS-CoV-2 genomes were sequenced worldwide by April 2021, rising to over seven million by January 2022 on the GISAID database (Elbe and Buckland-Merrett, 2017). This has provided unprecedented insight into the joint evolution and epidemiology of the SARS-CoV-2 pandemic (Harvey et al., 2021; Ward et al., 2021). Nevertheless, the cost of clinical sequencing to generate these data has been and continues to be substantial (10 - 35 GBP per sample for consumables (Tyson et al., 2020), excluding similar costs for staff, logistics and data infrastructure). It may be unsustainable at the levels required to adequately inform public health authorities as SARS-CoV-2 becomes endemic and threatens public health for the foreseeable future, even in developed nations.

Wastewater-based surveillance is a complementary, cost-effective approach to clinical sequencing, which has gained significant attention throughout the COVID-19 pandemic (Hillary et al., 2021; Jahn et al., 2021; Mishra et al., 2021; Peccia et al., 2020; Rios et al., 2021; Smyth et al., 2022). Given that SARS-CoV-2 is shed in faeces by more than 50% of infected people (Foladori et al., 2020), it can be recovered from wastewater, its RNA extracted, and its presence and quantity in a wastewater catchment determined using RT-qPCR (Farkas et al., 2020), with trends generally tracking the rise and fall of corresponding clinical cases (Hillary et al., 2021; Peccia et al., 2020). This can be achieved for entire populations by sampling at the inlet of wastewater treatment plants, or at much finer spatial scales, such as across cities, by sampling within the sewer network.

More recently, the recovery of SARS-CoV-2 genomes from wastewater has opened up the possibility of detecting and tracking circulating SARS-CoV-2 variants (Hillary et al., 2021; Jahn et al., 2021; Peccia et al., 2020; Rios et al., 2021; Smyth et al., 2022). Such an approach is particularly attractive for population-level insights during periods of high prevalence, especially if capacity constraints reduce the proportion of sequenced positive RT-qPCR tests. Furthermore, it can detect asymptomatic cases and is proposed to capture communities under-represented by clinical testing, particularly in urban centres (Green et al., 2021; Polo et al., 2020).

Nevertheless, moving from detecting and quantifying SARS-CoV-2 in wastewater by RT-qPCR to characterisation by genome sequencing is challenging. The low abundance of SARS-CoV-2 means enrichment through RNA concentration methods is necessary, simultaneously enriching PCR inhibitors and contaminating bacterial, viral, and human nucleic acids (Peccia et al., 2020). SARS-CoV-2 genomes in wastewater are also highly degraded and fragmented. In combination, this can result in poor and inconsistent amplification of target amplicons and, thus, patchy genome coverage. Even if amplification and sequencing are successful, data interpretation can be difficult. Wastewater harbours a mixed SARS-CoV-2 population. Therefore, sequences are derived from a pool of fragments, removing much of the phase information between polymorphic sites on the genome used to assign phylogeny and lineage. However, by reference against clinically-derived genomes of known SARS-CoV-2 lineages, wastewater data has the potential to detect and quantify polymorphisms characteristic of defined lineages and VOC in particular (Fontenele et al., 2021; Jahn et al., 2021).

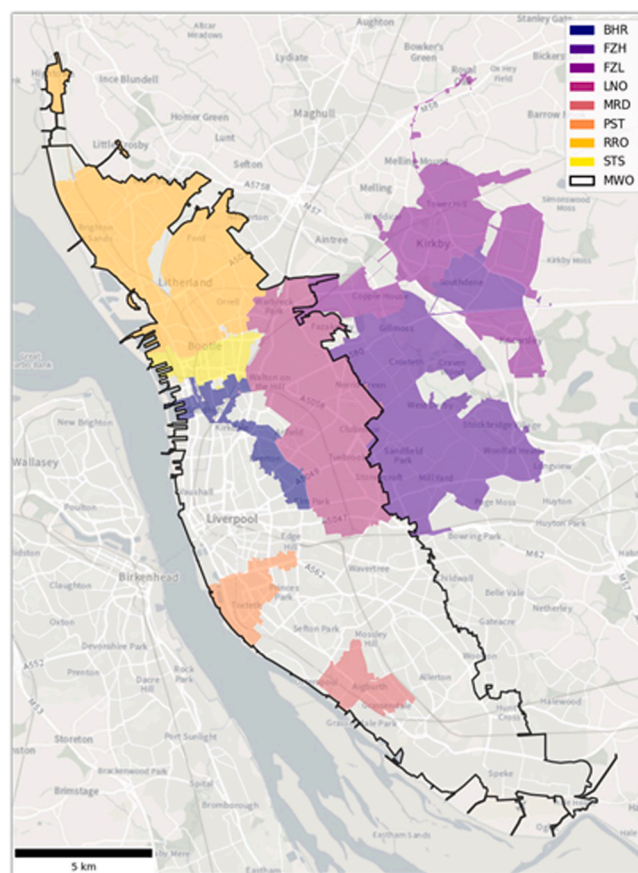
Our study demonstrates the utility of wastewater-based genomic surveillance of SARS-CoV-2 using longitudinal data collected from multiple locations in a single city – Liverpool, UK – between November

2020 and June 2021. During this time, Liverpool was the subject of a pilot study evaluating lateral flow tests for rapid asymptomatic testing (García-Fiñana et al., 2021). This pilot noted the link between social inequalities and testing uptake, with social deprivation and digital exclusion as significant factors limiting uptake (Green et al., 2021). Wastewater-based epidemiology (WBE) can provide valuable insight into some of the communities or areas of Liverpool that may be less accessible to conventional testing. This period in the UK also saw the emergence and establishment of the Alpha (B.1.1.7) and, subsequently, the Delta (B.1.617.2) SARS-CoV-2 variants. We show that wastewater genomic surveillance reliably detected the emergence of both and their subsequent rise across a city.

## 2. Materials and methods

### 2.1. Sample collection, concentration and RNA extraction

Wastewater grab samples (1 L per sample) were collected from eight locations across Liverpool's sewer network and from the main wastewater treatment plant (WWTP) at Sandon Docks between the 2nd of November 2020 and the 21st of June 2021, as part of the ongoing Environmental Monitoring for Health Protection programme (EMHP, part of NHS Test & Trace, now the UK Health Security Agency) in England (Fig. 1, Table S1). In addition, concurrent samples from four WWTPs in the southeast of England were collected as a control group between the 2nd of September 2020 and the 17th of January 2021. Samples were transported and subsequently stored at 4 - 6°C until



**Fig. 1.** Wastewater catchments of the 8 sewer network sampling locations and the WWTP across Liverpool. Network sample coverage is shown by coloured areas, WWTP catchment is shown by black outline. BHR= Bank Hall Relief, FZH= Fazakerley High, FZL= Fazakerley Low, LNO= Liverpool North, MRD= Mersey Road, PST= Park Street, RRO= Rimrose, STS= Strand SSO, MWO= Sandon Dock Main Works. See Table S1 for further catchment details.

analysis, minimising RNA degradation. Within 24 h of collection, all samples were centrifuged (10,000 x g, 4°C, 10 min) in sterile polypropylene tubes to remove suspended solids. The supernatant (50 ml) was transferred to 250 mL polycarbonate PPCO bottles containing 19–20 g of ammonium sulfate (Sigma-Aldrich, Cat. No. A4915). After the ammonium sulfate had dissolved, the samples were incubated at 4°C for 1 h before further centrifugation (10,000 x g, 4°C, 30 min) and supernatant removal. The pellet was resuspended in 200–500 µL of PBS. Concentrates were stored at 4°C until nucleic acid extraction. Nucleic acids were extracted from concentrates using NucliSens lysis buffer (BioMerieux, Marcy-l'Étoile, France, Cat. No. 280134 or 200292), NucliSens extraction reagent kit (BioMerieux, Cat. No. 200293) either manually (Farkas et al., 2021) or using the King-fisher 96 Flex system (Thermo Scientific, Waltham, MA, USA) according to the manufacturer instructions (Kevill et al., 2022), generating RNA extracts of 50–100 µL in volume. Extracts were stored at -80°C until further processing. Genome copies per litre (gc/L) of wastewater were calculated using One-step RT-qPCR for the SARS-CoV-2 N1, Phi6 and MNV targets using an RNA Ultrasense One-step RT-qPCR system (Life Technologies, Carlsbad, CA, USA, Cat. No. 11732927), on a Quant Studio Flex 6 (Applied Biosystems Inc., Waltham, MA, USA) as previously described (Kevill et al., 2022). Data were not subsequently normalised by flow rate, chemical composition, etc., since we were interested in the contribution of a variant to the proportion of viral RNA in a sample, not absolute case numbers.

## 2.2. SARS-CoV-2 RNA amplicon sequencing

Wastewater RNA extracts were purified and sequenced with a standardised EasySeq™ RC-PCR SARS CoV-2 (Nimagen) V1.0 protocol (Jeffries et al., 2021). In short, samples were cleaned with Mag-Bind® TotalPure NGS beads (Omega Bio-Tek) and then reverse transcribed using LunaScript® RT SuperMix Kit (New England Biolabs) and the EasySeq™ RC-PCR SARS CoV-2 (novel coronavirus) Whole Genome Sequencing kit v3.0 (NimaGen). Amplicons were pooled and libraries cleaned with Mag-Bind® Total Pure NGS beads (Omega Bio-Tek) before sequencing on an Illumina NovaSeq™ 6000 platform generating 2×150 bp paired-end reads.

## 2.3. Bioinformatics analysis

We processed raw reads using the *ncov2019-artic-nf* v3 pipeline (<https://github.com/connor-lab/ncov2019-artic-nf>) using default parameters. Briefly, reads were trimmed using Trim Galore v0.6.5 ([www.bioinformatics.babraham.ac.uk/projects/trim\\_galore](http://www.bioinformatics.babraham.ac.uk/projects/trim_galore)) and aligned to the SARS-CoV-2 reference genome (MN908947.3, (Wu et al., 2020)) using the Burrow-Wheeler Aligner (bwa) v0.7.17 (Li and Durbin, 2009). Primer sequences were trimmed using iVar v1.3 and a bed file containing the genome positions of the v3.0 primers that were used to

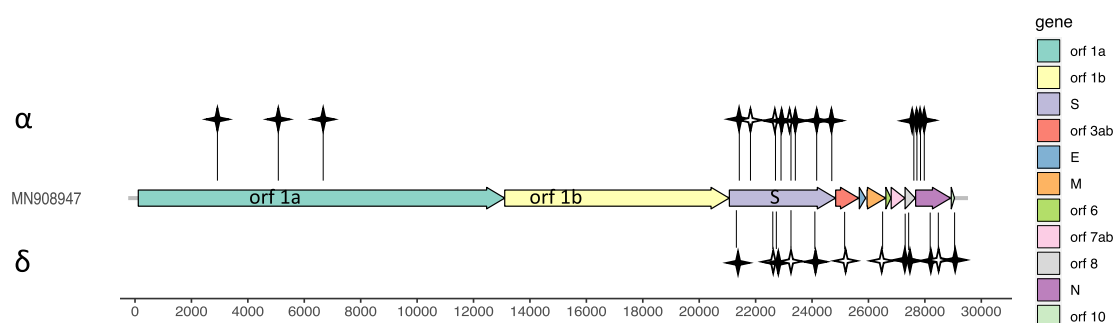
generate the amplicons. We then identified Single Nucleotide Polymorphisms (SNPs) and insertions/deletions (Indels) from BAM files using samtools (v1.13, (Danecek et al., 2021)) and VarScan (v2.4.4,  $P < 0.05$ , all other settings default, (Koboldt et al., 2012)) on 100,000 sequencing reads with an alignment score  $> 10$ . Next, we filtered the identified SNPs and Indels against signature mutations of known VOC and variants under investigation (VUI), as defined by Public Health England (PHE) at the time of writing ([https://github.com/phe-genomics/variant\\_definitions](https://github.com/phe-genomics/variant_definitions), Table S2, Fig. 2). We used custom scripts to extract summary statistics and sequence quality indicators, such as genome coverage, mapped reads and read depth (Fig. S1, Table S3).

To aid VOC and VUI identification at low frequencies from wastewater samples, we adopted a recently described amplicon-level co-occurrence approach (Jahn et al., 2021). Briefly, co-occurring mutations were called from BAM files using CoOccurrence adJusted Analysis and Calling (COJAC) (Jahn et al., 2021), facilitating the identification of signature mutations co-occurring on the same sequencing read, that is, a read or paired read coming from the same amplicon, thus one SARS-CoV-2 virion. This greatly improves confidence in variant detection, especially at low frequencies, since co-occurring mutations are less likely to arise through sequencing error than individual SNPs (Jahn et al., 2021). We extracted co-occurrence signature mutations of the B.1.1.7 (VOC-20DEC-01, Alpha) and B.1.617.2 (VOC-21APR-02, Delta) lineages. Since several signature mutations are shared amongst VOC-/VUI, not all variants have a unique set of co-occurring mutations. B.1.1.7 has unique pairs of co-occurring mutations on amplicon 146 (genome positions 27972 (Q27\*) and 28048 (R52I)) and amplicon 147 (genome positions 28111 (Y73C) and 28280 (D3L)), while B.1.617.2 only has one non-unique pair of mutations on amplicon 121 (genome positions 22917 (L452R) and 22995 (T478K)), i.e., it is shared with other variants.

## 2.4. Statistical analyses

We used R version 4.1.1 (R Core Team, 2021) for all statistical analyses and ggplot2 (Wickham, 2016) for visualisations.

Prior to analysis, all unique signature mutations (SNPs/Indels) of a given variant (Fig. 2) were identified in each sample, mutations with a read depth  $< 10$  removed and frequencies of 1.0 and 0 rescaled to 0.99 and 0.01 for beta regression compatibility, respectively. We modelled the relationship of the mean frequency of each variant's signature SNPs/indels with location (i.e. differing network sites) and time during respective variant emergences with beta linear regression (betareg, “betareg” v.3.1.4, (Cribari-Neto and Zeileis, 2010)), given allele frequencies are in the standard unit interval [0, 1]. To do so, we set the mean frequency of the unique signature SNPs/indels of a given variant as the dependent variable and wastewater site, date, and their interaction as predictor variables. All models were fit by maximum likelihood using the logit link function, *logit(p)* with *p* the probability of observing the (variant) data and *logit* being the inverse of the standard logistic



**Fig. 2.** SARS-CoV-2 genome structure and signature mutations of VOC Alpha and Delta. Black stars show unique mutations for the Alpha or Delta variant. White stars show mutations shared with at least one other VOC or VUI and therefore not used for mean frequency estimates. Details of all mutations can be found in Table S2. Genome structure adopted from Wu et al. (2020).



function, and included site as an additional regressor for the precision parameter when it improved the model fit (see Table S4 for final model structures), as indicated by Akaike Information Criterion (AIC) and likelihood ratio tests (Cribari-Neto and Zeileis, 2010). To account for missing data in SNP/indel frequencies, a weighting factor was applied using the number of used signature SNPs/indels (weights =  $n$ ). We assessed model validity by visual checks of homoscedasticity of the standardised weighted residuals and linearity of the model fit (Fig. S2). We then extracted likelihood ratio tests of estimated marginal means for each predictor variable (joint tests, “emmeans”, Table S4).

We also compared the frequency of detected VOC/VUI signature SNPs/indels in wastewater samples to the frequency of VOC/VUI identified in clinical cases by the COVID-19 Genomics (COG-UK) Consortium between the 2<sup>nd</sup> of November 2020 and the 21<sup>st</sup> of June 2021 across Liverpool. We extracted counts of genomically confirmed cases for all circulating lineages from the CLIMB platform (Nicholls et al., 2021) on the 26<sup>th</sup> of October 2021 and then filtered and grouped them by the outer postcodes covered by the catchment areas of the WWTP and the eight sewer network sites. For the Delta variant, confirmed clinical cases of the B.617.2 lineage and its subvariant AY.4 were combined. Where outer postcodes spanned multiple wastewater catchments, we included clinical cases in counts for all those sites, divided by the number of overlapping wastewater catchments. Additionally, we obtained total daily infection numbers for the upper-tier local authority of Liverpool from UK Government statistics (<https://coronavirus.data.gov.uk>, Fig. S5).

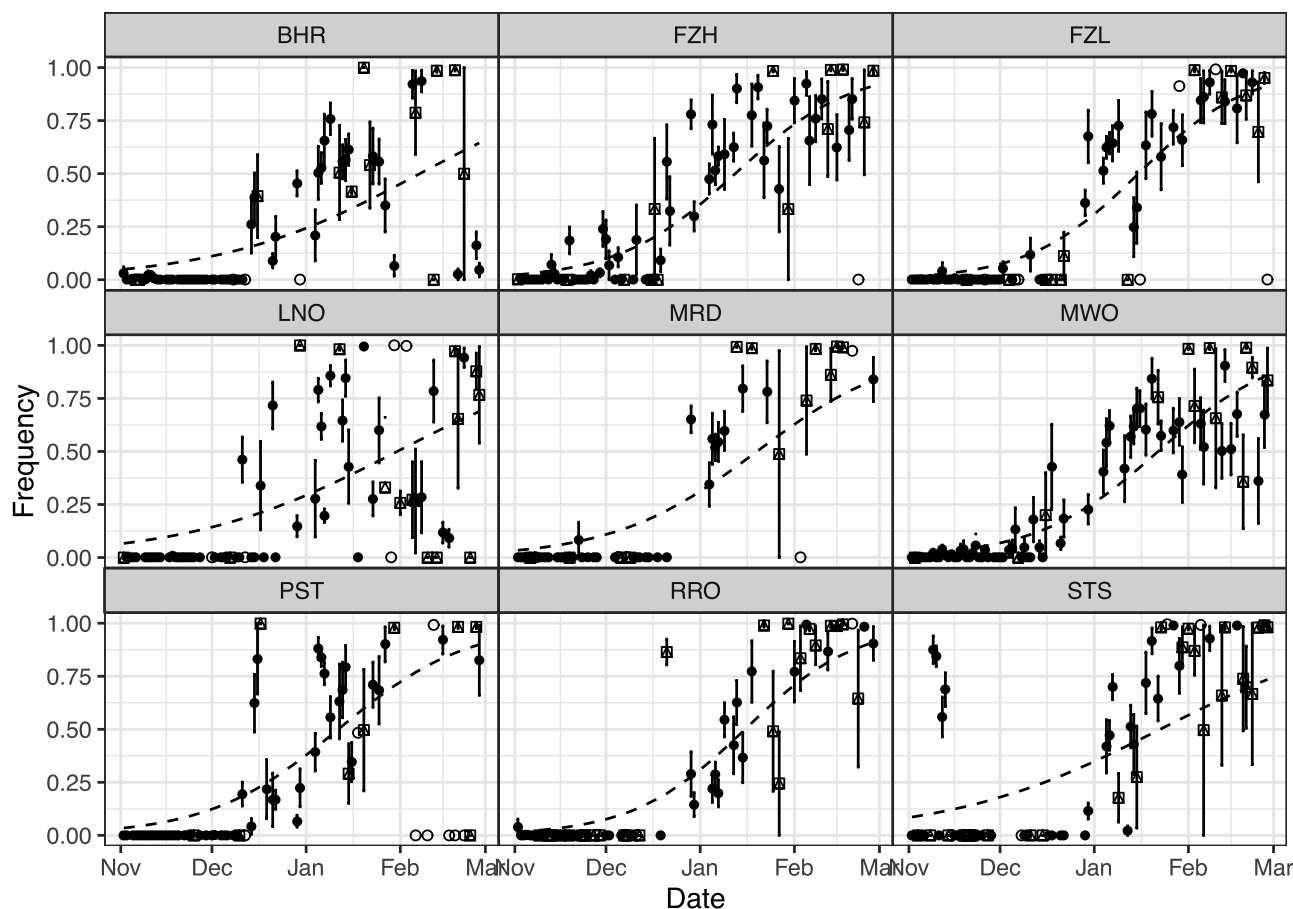
We modelled the frequency change of variants in clinical data over time with beta linear regression in the same way as for wastewater

variant frequencies. To test the time match between a respective variant's frequency in wastewater and clinical samples, we used Spearman's rank correlation with a series of possible time lag settings to find the time frame shift with the best match for each sampling area.

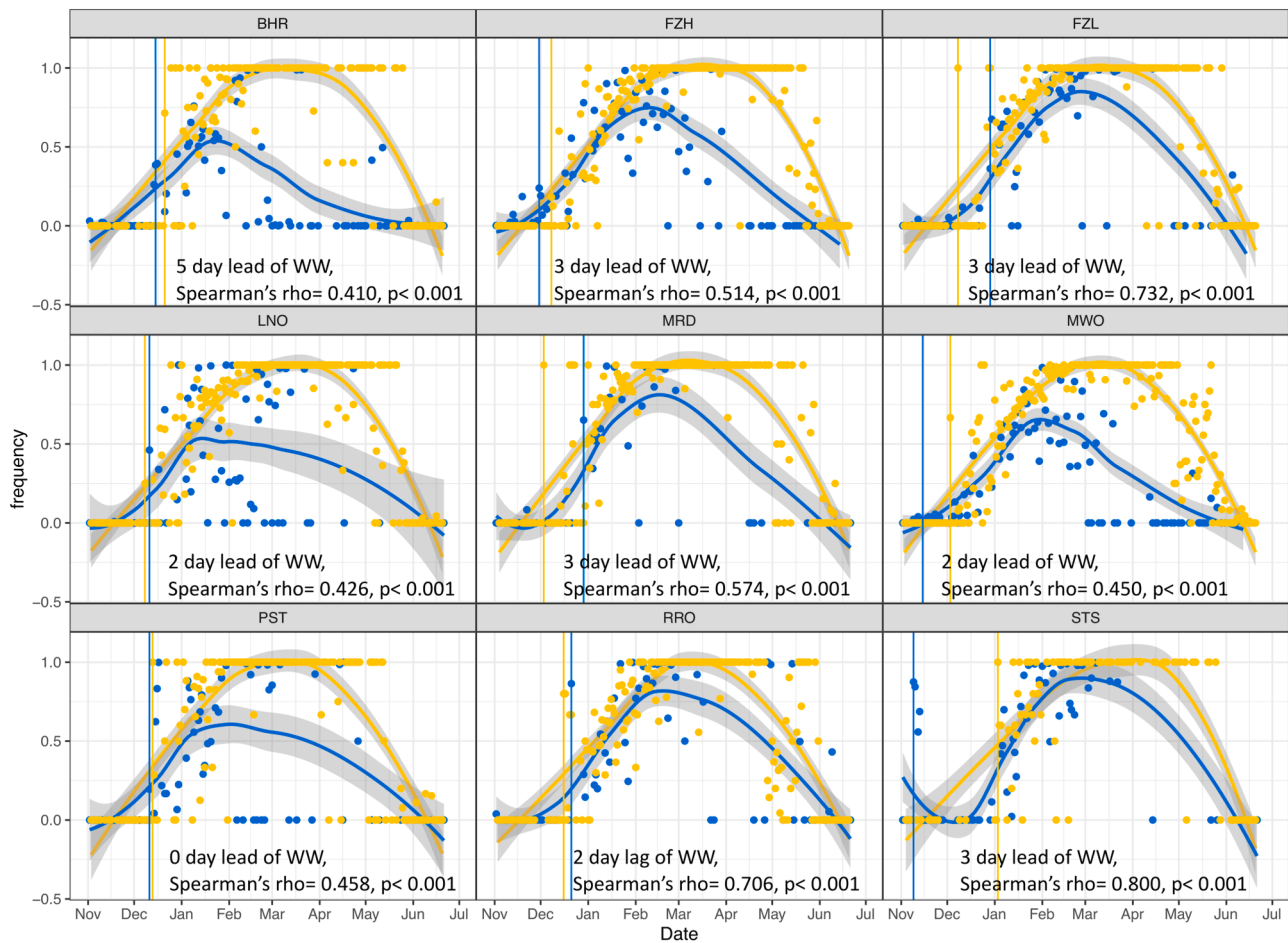
### 3. Results

#### 3.1. The rise of Alpha variant

Across all catchments, we observed a significant increase in the mean frequency of Alpha (B.1.1.7) signature SNPs/indels between the 2<sup>nd</sup> of November 2020 and the 28<sup>th</sup> of February 2021 ( $F_1 = 13667$ ,  $P < 0.001$ , Fig. 3, Table S4). This closely corresponds with the observed rise in Alpha clinical cases across Liverpool for the same period (Fig. 4) and wastewater data from four WWTPs in the southeast of England ( $F_1 = 13829$ ,  $P < 0.001$ , Table S4, Fig. S3). For most sites, the rise of the Alpha variant began in mid to late December, with peak frequencies observed in late January and early February (Figs. 3 and 4). As defined by co-occurrence analysis, the earliest wastewater Alpha variant detection preceded clinical detections in five of the nine sites by up to 55 days (Fig. 4). The contrary was observed in the remaining four sites, with clinical samples picking up Alpha up to 26 days earlier (Fig. 4). The best time match between Alpha frequencies in wastewater and clinical samples also depended on catchment. The closest match varied from a 5-day lead to a 2-day lag in the wastewater when testing a range of 5-day lag to 5-day lead of wastewater frequencies (Fig. 4, Table S5). However, this does not match the relative pattern of the earliest detection in the two sample types.



**Fig. 3.** Mean frequency of B.1.1.7 (Alpha) signature SNPs/Indels during its rise in each wastewater catchment, 2<sup>nd</sup> of November 2020 to 28<sup>th</sup> of February 2021. Points and error bars show means and standard errors across all unique Alpha-specific mutations with sufficient sequencing coverage in each sample. Dashed lines show the best fit beta regression line for this period (Table S4). Point shape indicates the number of unique Alpha-specific mutations used in the mean calculation for a given sample: empty circles: 1 mutation, crossed square: 2 to 5 mutations, filled circles: >5 mutations.



**Fig. 4.** Mean frequency of B.1.1.7 (Alpha) signature SNPs/Indels detected in wastewater and clinical samples in each catchment, 2<sup>nd</sup> of November 2020 to 26<sup>th</sup> of June 2021. Points show the mean frequency of unique Alpha-specific mutations for a given wastewater sample (blue) and the frequency of Alpha variant clinical cases from a given date (yellow). Coloured lines show the respective local polynomial regression fit including shaded 95% confidence intervals. Vertical lines indicate the first confirmed clinical case of Alpha variant (yellow) and the first wastewater detection of co-occurring Alpha-specific mutations on amplicon 147 (blue). The strongest correlation between a 5-day lead and 5-day lag of wastewater data respective to clinical data is reported at the bottom of each graph.

We detected local differences in the rise of the Alpha variant between wastewater catchments (date: site interaction,  $F_8 = 32.8$ ,  $P < 0.001$ , Table S4). This was most notable in Strand SSO (STS), where we observed a high frequency of Alpha signature SNPs from four samples in early November (Fig. 3), though this signal diminished before further detections in early January. Similarly, we observed Alpha signature SNPs at a low to moderate frequency at Fazakerley High (FZH) as early as mid-November, while they were barely detected until late December in Mersey Road (MRD, Fig. 3). This suggests Alpha spread through parts of the north of the city earlier than through the south (Fig. 1), a finding corroborated by co-occurrence analysis but not clinical data (Fig. 4).

### 3.2. The decline of Alpha and rise of Delta variant

From the 15<sup>th</sup> of March to the 26<sup>th</sup> of June 2021, we observed a significant increase in the frequency of Delta signature SNPs in all wastewater catchments ( $F_1 = 964.1$ ,  $P < 0.001$ , Table S3, Fig. 5), with variation in temporal trends across the city (date: site interaction,  $F_8 = 32.4$ ,  $P < 0.001$ , Table S4, Fig. 5). This coincides with a rise in clinical cases of the same variant (B.1.617.2 and AY.4, Fig. S4) and Alpha's decline (Figs. 3 and 4). The best time match between Delta frequencies in wastewater and clinical samples again varied by catchment from a 4-day lead to a 5-day lag in the wastewater. In all catchments, a significant correlation of frequencies in wastewater and clinical samples was confirmed (Fig. S4, Table S6).

It is noteworthy that the observed transition from Alpha to Delta in wastewater was abrupt (Figs. 4, 5 and S4). From April to early June, infection numbers were low across Liverpool (Fig. S5), and wastewater SARS-CoV-2 concentrations were consequently low (Fig. S6). This is reflected in the observed reduction in mapped reads and genome coverage for this period (Fig. S1). Indeed, the detection of Alpha and Delta signature SNPs was more sporadic during this period (Figs. 4, 5 and S4). Lower SARS-CoV-2 concentrations – and associated lower data quality – probably also contributes to the lower estimates of Alpha frequency in wastewater relative to clinical data from around March 2021.

Co-occurrence analysis was less discriminatory for Delta than Alpha and, thus, a less reliable indicator of its presence in wastewater catchments. We note an apparent co-occurrence signal for Delta, i.e., the detection of L452R and T478K co-occurring on amplicon 121 of our primer panel, as early as November for all wastewater catchments (Fig. S4). This was too early to be Delta, its sub-lineages (e.g., AY.2 and AY.3) or known VUI B.1.629, B.1.630 and B.1.633 (first detected globally in February and March 2021), which also carry this pair of co-occurring mutations. The detected signal, thus, must reflect other circulating variants carrying these SNPs or false positives due to reliance on co-occurrence in a single amplicon.

## 4. Discussion

Genomic surveillance of wastewater has already shown great

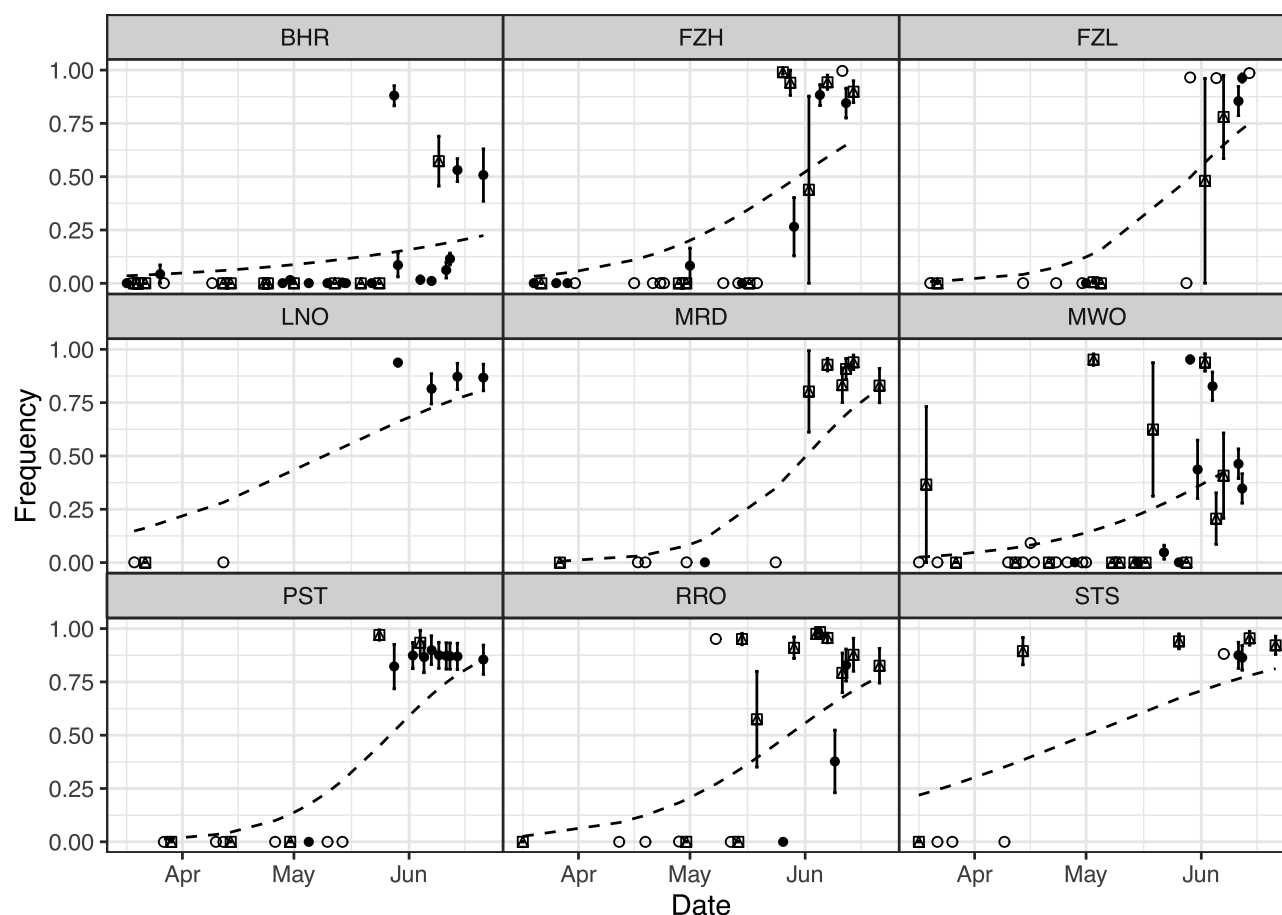


Fig. 5. Mean frequency of B.1.617.2 (Delta) signature SNPs/Indels during its rise in each wastewater catchment, 15<sup>th</sup> of March 2021 to 26<sup>th</sup> of June 2021. Points and error bars show means and standard errors across all unique Delta-specific mutations with sufficient sequencing coverage in each sample. Dashed lines show the best fit beta regression line for this period (Table S4). Point shape indicates the number of unique Delta-specific mutations used in the mean calculation for a given sample: empty circles: 1 mutation, crossed square: 2 to 5 mutations, filled circles: >5 mutations.

promise throughout the unfolding SARS-CoV-2 pandemic (Farkas et al., 2020; Polo et al., 2020), including the detection of VOC (Fontenele et al., 2021), in some cases prior to clinical detection (Jahn et al., 2021). Here, we have demonstrated that wastewater monitoring can also reveal fine-scale, local differences in the spread of VOC across urban centres. Spatiotemporal differences in variant frequencies across Liverpool were recoverable throughout the rise of the Alpha variant (B.1.1.7) in early winter 2020 and, despite lower quality data, the rise of the Delta variant (B.1.617.2/AY.4) in spring 2021. This clear reflection of the rise of Alpha and Delta variants, respectively, in clinical and wastewater genomic data, demonstrates the reliability of this approach.

As seen for the Alpha variant here and by Jahn et al. (2021), genomic surveillance of wastewater can detect VOC earlier than clinical testing. In both instances, co-occurrence analysis improved confidence in (early) low-frequency variant detection by identifying multiple linked mutations from the same virion instead of solely relying on single signature mutations. This requires the co-occurrence of mutations unique to a given variant on amplicons of the used sequencing scheme. When no unique mutation set is available, as for the Delta variant and the NimaGen SARS-CoV-2 whole-genome sequencing kit used here, reliable variant detection via co-occurrence analysis is not possible. The software developers have acknowledged this limitation, and the design of primers to create appropriate co-occurrence amplicons for relevant sequencing schemes is suggested as a workaround (Jahn et al., 2021). It is important to note that even in cases where co-occurrence analysis is applicable, our fine-scale local data highlighted that wastewater monitoring sometimes detects new variants earlier than clinical testing, but not always. The

reasons for this are yet unclear. It is likely that the inherent variability of wastewater detections, due to variations in viral shedding rates and dilution from rainfall (Polo et al., 2020), and the increasing stochasticity of clinical detection with decreasing population size play a role. It is also worth noting that these samples were almost entirely grab samples, which likely sample fewer clinical infections than composite samples. Certainly, the relationships between population size, wastewater flow variation and SARS-CoV-2 variant detection warrant further investigation.

The mixed pattern of wastewater Alpha variant detections preceding confirmed clinical cases in some parts of Liverpool, yet vice versa in others, highlights the complementarity of the approaches. Wastewater monitoring has the notable advantages of being more cost-effective per unit of population and is less biased by testing frequencies in different communities (Polo et al., 2020), while sequencing of clinical samples provides greater specificity and the opportunity for contact tracing. Indeed, our finding that the Alpha variant was detected in wastewater in North Liverpool much earlier than clinical cases had indicated, corresponds well with findings from a large-scale asymptomatic testing campaign, which found that testing uptake was lower in North Liverpool, yet the rate of positive tests higher (Green et al., 2021). Clearly, a combination of genomic surveillance of clinical cases and wastewater is most likely to detect new variants as early as possible and provides the most precise picture of unfolding variant dynamics to inform public health measures (Mishra et al., 2021).

Intriguingly, when comparing peak Alpha and Delta variant frequencies in corresponding clinical and wastewater data, we find that

each variant, in turn, reaches complete dominance in clinical but not wastewater samples, with estimated maximum frequencies of 100% and ~75% in clinical and wastewater samples, respectively. Correspondingly, correlations between clinical and wastewater data appear stronger in sampling areas with higher maximum frequency estimates in wastewater. It is notable that viral concentrations were low from March – May 2021, which corresponded to public health restrictions in the UK to contain infection numbers, and which led to lower sequence data quality for these samples. Improved viral concentration methods may mitigate this limitation (Kevill et al., 2022). Equally, better statistical methods may be required to estimate lineage frequencies from pooled sequencing data, as produced from wastewater samples (Amman et al., 2022; Karthikeyan et al., 2022). Here we have relied on a relatively crude estimation by taking the mean of signature SNP/Indel frequencies. However, genetic variation may mean that a SNP/Indel may not be present on all branches within a lineage, whereas a single fully phased viral genome from a clinical sample would be reliably assigned to a lineage. Following methods for analysing metagenomic amplicon data (Grubaugh et al., 2019; Quince et al., 2021; 2011), the development of statistical methods to infer lineage proportions from multiple amplicons while controlling for sequencing error may be productive.

The limited quality of sequences obtained from wastewater during April and May 2020 also highlights the current limits of variant detection via wastewater sequencing when case numbers, and hence SARS-CoV-2 concentrations, are low. While the rise of Delta was evident in our results (Fig. 5), the transition from Alpha to Delta, compared to the gradual emergence of Alpha, was less visible and more abrupt. If new, more transmissible variants are associated with more rapid emergence and a quicker rise in frequency, this would highlight the need to develop more sensitive and accurate methods of variant detection and quantification within wastewater. It is, however, anticipated that the increased adoption of wastewater-based epidemiology will drive innovation in wastewater sampling, concentration and RNA extraction, improving viral qPCR and sequencing sensitivity (Hillary et al., 2021; Kevill et al., 2022; Polo et al., 2020).

## 5. Conclusions

- We show that wastewater genomic sequencing can reliably detect the emergence and rise of new SARS-CoV-2 variants.
- Variant frequency estimates from wastewater sequencing correspond well with those obtained through genomic sequencing of clinical samples.
- In some cases, variants are observed in wastewater before clinical detections, which may be particularly useful in areas or communities with low testing uptake.

## Author contributions

MRB and SP conceived and designed the study. ARJ developed laboratory methods for data collection. JLK, KF, EC and CW collected the wastewater data. COG-UK provided clinical and community testing data. FSB and MRB processed and analysed the data. IB, HB, MSK, RvA and SP provided bioinformatic support for processing the data. MJW, DLJ and SP provided supervision of the work. FSB, MRB, and SP drafted the manuscript. All authors reviewed and edited the manuscript. All authors approved the final version of the report.

## Data accessibility statement

Wastewater sequencing data is publicly available on the European Nucleotide Archive under Study ID PRJEB53325 (ERP138109). The clinical case data used in this study are visualised at <https://www.cogconsortium.uk/tools-analysis/public-data-analysis-2/>. A filtered, privacy conserving version of the lineage-LTLA-week dataset is publicly available online (<https://covid19.sanger.ac.uk/downloads>) and gives

access to almost all used data, despite a small number of cells having been suppressed to conserve patient privacy.

## Ethics statement

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

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data Availability

Data is available according to the Data Availability statement in the manuscript

## Acknowledgements

We thank R. Crompton, B. Jones, G. Airey, T. Foster, N. Kadu, C. Nelson and A. Lucaci for help with processing samples and sequencing and United Utilities for providing wastewater catchment mapping data.

Funding was provided by NERC (NE/V003860/1) and DHSC UK (2020\_097). This report is independent research funded by the Department of Health and Social Care. The views expressed in this publication are those of the authors and not necessarily those of the Department of Health and Social Care. COG-UK is supported by funding from the Medical Research Council (MRC) part of UK Research & Innovation (UKRI), the National Institute of Health Research (NIHR) [grant code: MC\_PC\_19027], and Genome Research Limited, operating as the Wellcome Sanger Institute.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.watres.2022.119306](https://doi.org/10.1016/j.watres.2022.119306).

## References

- Amman, F., Markt, R., Endler, L., Hupfau, S., Agerer, B., Schedl, A., Richter, L., Zechmeister, M., Bicher, M., Heiler, G., Triska, P., Thornton, M., Penz, T., Senekowitsch, M., Laine, J., Keszei, Z., Klimek, P., Nägele, F., Mayr, M., Daleiden, B., Steinlechner, M., Niederstätter, H., Heidinger, P., Rauch, W., Scheffknecht, C., Vogl, G., Weichlinger, G., Wagner, A.O., Slipko, K., Masseron, A., Radu, E., Allerberger, F., Popper, N., Bock, C., Schmid, D., Oberacher, H., Kreuzinger, N., Insam, H., Berghaler, A., 2022. Viral variant-resolved wastewater surveillance of SARS-CoV-2 at national scale. *Nat. Biotechnol.* 1–9. <https://doi.org/10.1038/s41587-022-01387-y>.
- Cherian, S., Potdar, V., Jadhav, S., Yadav, P., Gupta, N., Das, M., Rakshit, P., Singh, S., Abraham, P., Panda, S., Team, N., 2021. SARS-CoV-2 Spike Mutations, L452R, T478K, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra. *India. Microorganisms* 9, 1542. <https://doi.org/10.3390/microorganisms9071542>.
- Cribari-Neto, F., Zeileis, A., 2010. Beta Regression in R. *J. Stat. Softw.* 34 <https://doi.org/10.18637/jss.v034.i02>.
- Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., Li, H., 2021. Twelve years of SAMtools and BCFtools. *GigaScience* 10. <https://doi.org/10.1093/gigascience/giab008>.
- Davies, N.G., Abbott, S., Barnard, R.C., Jarvis, C.I., Kucharski, A.J., Munday, J.D., Pearson, C.A.B., Russell, T.W., Tully, D.C., Washburne, A.D., Wenseleers, T., Gimma, A., Waites, W., Wong, K.L.M., van Zandvoort, K., Silverman, J.D., Diaz-Ordaz, K., Keogh, R., Eggo, R.M., Funk, S., Jit, M., Atkins, K.E., Edmunds, W.J., 2021. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* 372. <https://doi.org/10.1126/science.abg3055> eabg3055–eabg3055.



- Elbe, S., Buckland-Merrett, G., 2017. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.* 1, 33–46. <https://doi.org/10.1002/gch2.1018>.
- Faria, N.R., Mellan, T.A., Whittaker, C., Claro, I.M., Candido, D.da S., Mishra, S., Crispin, M.A.E., Sales, F.C.S., Hawryluk, I., McCrone, J.T., Hulsmit, R.J.G., Franco, L.A.M., Ramundo, M.S., de Jesus, J.G., Andrade, P.S., Coletti, T.M., Ferreira, G.M., Silva, C.A.M., Manuli, E.R., Pereira, R.H.M., Peixoto, P.S., Kraemer, M.U.G., Gaburo, N., Camilo, C.da C., Hoeltgebaum, H., Souza, W.M., Rocha, E.C., de Souza, L.M., de Pinho, M.C., Araujo, L.J.T., Malta, F.S.V., de Lima, A.B., Silva, J.do P., Zauli, D.A.G., Ferreira, A.C.de S., Schnekenberg, R.P., Laydon, D.J., Walker, P.G.T., Schlüter, H.M., dos Santos, A.L.P., Vidal, M.S., Del Caro, V.S., Filho, R.M.F., dos Santos, H.M., Aguiar, R.S., Proença-Modena, J.L., Nelson, B., Hay, J.A., Monod, M., Miscoiridou, X., Coupland, H., Sonabend, R., Vollmer, M., Gandy, A., Prete, C.A., Nascimento, V.H., Suchard, M.A., Bowden, T.A., Pond, S.L.K., Wu, C.-H., Ratmann, O., Ferguson, N.M., Dye, C., Loman, N.J., Lemey, P., Rambaut, A., Fraiji, N.A., Carvalho, M.do P.S.S., Pybus, O.G., Flaxman, S., Bhatt, S., Sabino, E.C., 2021. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Sci.* 372, 815–821. <https://doi.org/10.1126/science.abb2644>.
- Farkas, K., Hillary, L.S., Malham, S.K., McDonald, J.E., Jones, D.L., 2020. Wastewater and public health: the potential of wastewater surveillance for monitoring COVID-19. *Curr. Opin. Environ. Sci. Health* 17, 14–20. <https://doi.org/10.1016/j.coesh.2020.06.001>.
- Farkas, K., Hillary, L.S., Thorpe, J., Walker, D.L., Lowther, J.A., McDonald, J.E., Malham, S.K., Jones, D.L., 2021. Concentration and Quantification of SARS-CoV-2 RNA in Wastewater Using Polyethylene Glycol-Based Concentration and qRT-PCR. *Methods Protoc* 4, 17. <https://doi.org/10.3390/mps4010017>.
- Foladori, P., Cutrupi, F., Segata, N., Manara, S., Pinto, F., Malpei, F., Bruni, L., La Rosa, G., 2020. SARS-CoV-2 from faeces to wastewater treatment: What do we know? A review. *Sci. Total Environ.* 743, 140444. <https://doi.org/10.1016/j.scitotenv.2020.140444>.
- Fontenele, R.S., Kraberger, S., Hadfield, J., Driver, E.M., Bowes, D., Holland, L.A., Faleye, T.O.C., Adhikari, S., Kumar, R., Inchausti, R., Holmes, W.K., Deitrick, S., Brown, P., Duty, D., Smith, T., Bhatnagar, A., Yeager, R.A., Holm, R.H., von Reitzenstein, N.H., Wheeler, E., Dixon, K., Constantine, T., Wilson, M.A., Lim, E.S., Jiang, X., Halden, R.U., Scotch, M., Varsani, A., 2021. High-throughput sequencing of SARS-CoV-2 in wastewater provides insights into circulating variants. *Water Res* 205, 117710. <https://doi.org/10.1016/j.watres.2021.117710>.
- García-Fiñana, M., Hughes, D.M., Cheyne, C.P., Burnside, G., Stockbridge, M., Fowler, T.A., Fowler, V.L., Wilcox, M.H., Semple, M.G., Buchan, I., 2021. Performance of the Innova SARS-CoV-2 antigen rapid lateral flow test in the Liverpool asymptomatic testing pilot: population based cohort study. *BMJ* n1637. <https://doi.org/10.1136/bmj.n1637>.
- Green, M.A., García-Fiñana, M., Barr, B., Burnside, G., Cheyne, C.P., Hughes, D., Ashton, M., Sheard, S., Buchan, I.E., 2021. Evaluating social and spatial inequalities of large scale rapid lateral flow SARS-CoV-2 antigen testing in COVID-19 management: An observational study of Liverpool, UK (November 2020 to January 2021). *Lancet Reg. Health - Eur.* 6, 100107. <https://doi.org/10.1016/j.lanepe.2021.100107>.
- Grubaugh, N.D., Gangavarapu, K., Quick, J., Matteson, N.L., De Jesus, J.G., Main, B.J., Tan, A.L., Paul, L.M., Brackney, D.E., Grewal, S., Gurfield, N., Van Rompay, K.K.A., Isern, S., Michael, S.F., Coffey, L.L., Loman, N.J., Andersen, K.G., 2019. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol* 20, 8. <https://doi.org/10.1186/s13059-018-1618-7>.
- Harvey, W.T., Carabelli, A.M., Jackson, B., Gupta, R.K., Thomson, E.C., Harrison, E.M., Ludden, C., Reeve, R., Rambaut, A., Peacock, S.J., Robertson, D.L., 2021. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* 19, 409–424. <https://doi.org/10.1038/s41579-021-00573-0>.
- Hillary, L.S., Farkas, K., Maher, K.H., Lucaci, A., Thorpe, J., Distaso, M.A., Gaze, W.H., Paterson, S., Burke, T., Connor, T.R., McDonald, J.E., Malham, S.K., Jones, D.L., 2021. Monitoring SARS-CoV-2 in municipal wastewater to evaluate the success of lockdown measures for controlling COVID-19 in the UK. *Water Res* 200, 117214. <https://doi.org/10.1016/j.watres.2021.117214>.
- Jahn, K., Dreifuss, D., Topolsky, I., Kull, A., Ganesanandamoorthy, P., Fernandez-Cassi, X., Bänziger, C., Devaux, A.J., Stachler, E., Caduff, L., Cariti, F., Corzón, A.T., Fuhrmann, L., Chen, C., Jablonski, K.P., Nadeau, S., Feldkamp, M., Beisel, C., Aquino, C., Stadler, T., Ort, C., Kohn, T., Julian, T.R., Beerenwinkel, N., 2021. Detection and surveillance of SARS-CoV-2 genomic variants in wastewater. *medRxiv*. <https://doi.org/10.1101/2021.01.08.21249379>.
- Jeffries, A., Paterson, S., Loose, M., van Aerle, R., 2021. Wastewater Sequencing using the EasySeqTM RC-PCR SARS-CoV-2 (Nimigen) V1.0. protocols.io 8.
- Karthikeyan, S., Levy, J.L., De Hoff, P., Humphrey, G., Birmingham, A., Jepsen, K., Farmer, S., Tubb, H.M., Valles, T., Tribelhorn, C.E., Tsai, R., Aigner, S., Sathe, S., Moshiri, N., Henson, B., Mark, A.M., Hakim, A., Baer, N.A., Barber, T., Belda-Ferre, P., Chacón, M., Cheung, W., Cresini, E.S., Eisner, E.R., Lastrella, A.L., Lawrence, E.S., Marotz, C.A., Ngo, T.T., Ostrander, T., Plascencia, A., Salido, R.A., Seaver, P., Smoot, E.W., McDonald, D., Neuhauser, R.M., Scioscia, A.L., Satterlund, A.M., Simmons, E.H., Abelman, D.B., Brenner, D., Bruner, J.C., Buckley, A., Ellison, M., Gattas, J., Gonias, S.L., Hale, M., Hawkins, F., Ikeda, L., Jhaveri, H., Johnson, T., Kellen, V., Kremer, B., Matthews, G., McLawhon, R.W., Ouellet, P., Park, D., Pradenas, A., Reed, S., Riggs, L., Sanders, A., Sollenberger, B., Song, A., White, B., Winbush, T., Aceves, C.M., Anderson, Catelyn, Gangavarapu, K., Hufbauer, E., Kurzban, E., Lee, J., Matteson, N.L., Parker, E., Perkins, S.A., Ramesh, K.S., Robles-Sikisaka, R., Schwab, M.A., Spencer, E., Wohl, S., Nicholson, L., McHardy, I.H., Dimmock, D.P., Hobbs, C.A., Bakhtar, O., Harding, A., Mendoza, A., Bolze, A., Becker, D., Cirulli, E.T., Isaksson, M., Schiabor Barrett, K.M., Washington, N.L., Malone, J.D., Schafer, A.M., Gurfield, N., Stous, S., Fielding-Miller, R., Garfein, R.S., Gaines, T., Anderson Cheryl, Martin, N.K., Schooley, R., Austin, B., MacCannell, D. R., Kingsmore, S.F., Lee, W., Shah, S., McDonald, E., Yu, A.T., Zeller, M., Fisch, K.M., Longhurst, C., Maysent, P., Pride, D., Khosla, P.K., Laurent, L.C., Yeo, G.W., Andersen, K.G., Knight, R., 2022. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. *Nature* 609, 101–108. <https://doi.org/10.1038/s41586-022-05049-6>.
- Kevill, J.L., Pellett, C., Farkas, K., Brown, M.R., Bassano, I., Denise, H., McDonald, J.E., Malham, S.K., Porter, J., Warren, J., Evens, N.P., Paterson, S., Singer, A.C., Jones, D. L., 2022. A comparison of precipitation and filtration-based SARS-CoV-2 recovery methods and the influence of temperature, turbidity, and surfactant load in urban wastewater. *Sci. Total Environ.* 808, 151916. <https://doi.org/10.1016/j.scitotenv.2021.151916>.
- Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L., Wilson, R.K., 2012. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 22, 568–576. <https://doi.org/10.1101/gr.129684.111>.
- Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinforma. Oxf. Engl.* 25, 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
- Mishra, S., Mindermann, S., Sharma, M., Whittaker, C., Mellan, T.A., Wilton, T., Klapsa, D., Mate, R., Fritzsche, M., Zambon, M., Ahuja, J., Howes, A., Miscoiridou, X., Nason, G.P., Ratmann, O., Semenova, E., Leech, G., Sandkühler, J.F., Rogers-Smith, C., Vollmer, M., Unwin, H.J.T., Gal, Y., Chand, M., Gandy, A., Martin, J., Volz, E., Ferguson, N.M., Bhatt, S., Brauer, J.M., Flaxman, S., 2021. Changing composition of SARS-CoV-2 lineages and rise of Delta variant in England. *EClinicalMedicine* 39, 101064. <https://doi.org/10.1016/j.eclinm.2021.101064>.
- Nicholls, S.M., Poplawski, R., Bull, M.J., Underwood, A., Chapman, M., Abu-Dahab, K., Taylor, B., Colquhoun, R.M., Rowe, W.P.M., Jackson, B., Hill, V., O'Toole, A., Rey, S., Southgate, J., Amato, R., Livett, R., Gonçalves, S., Harrison, E.M., Peacock, S. J., Aanensen, D.M., Rambaut, A., Connor, T.R., Loman, N.J., 2021. CLIMB-COVID: continuous integration supporting decentralised sequencing for SARS-CoV-2 genomic surveillance. *Genome Biol* 22, 196. <https://doi.org/10.1186/s13059-021-02395-y>.
- Peccia, J., Zulli, A., Brackney, D.E., Grubaugh, N.D., Kaplan, E.H., Casanovas-Massana, A., Ko, A.I., Malik, A.A., Wang, D., Wang, M., Warren, J.L., Weinberger, D. M., Arnold, W., Omer, S.B., 2020. Measurement of SARS-CoV-2 RNA in wastewater tracks community infection dynamics. *Nat. Biotechnol.* 38, 1164–1167. <https://doi.org/10.1038/s41587-020-0684-z>.
- Polo, D., Quintela-Balaja, M., Corbishley, A., Jones, D.L., Singer, A.C., Graham, D.W., Romalde, J.L., 2020. Making waves: Wastewater-based epidemiology for COVID-19 – approaches and challenges for surveillance and prediction. *Water Res* 186, 116404. <https://doi.org/10.1016/j.watres.2020.116404>.
- Qin, S., Cui, M., Sun, S., Zhou, J., Du, Z., Cui, Y., Fan, H., 2021. Genome Characterization and Potential Risk Assessment of the Novel SARS-CoV-2 Variant Omicron (B.1.1.529). *Zoonoses* 1. <https://doi.org/10.15212/ZOONOSSES-2021-0024>.
- Quince, C., Lanzen, A., Davenport, R.J., Turnbaugh, P.J., 2011. Removing Noise From Pyrosequenced Amplicons. *BMC Bioinformatics* 12, 38. <https://doi.org/10.1186/1471-2105-12-38>.
- Quince, C., Nurk, S., Raguideau, S., James, R., Soyer, O.S., Summers, J.K., Limasset, A., Eren, A.M., Chikhi, R., Darling, A.E., 2021. STRONG: metagenomics strain resolution on assembly graphs. *Genome Biol* 22, 214. <https://doi.org/10.1186/s13059-021-02419-7>.
- R Core Team, 2021. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Rambaut, A., Loman, N., Pybus, O., Barclay, W., Barrett, J., Carabelli, A., Connor, T., Peacock, T., Robertson, D.L., Volz, E., COG UK, 2020. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations.
- Rios, G., Lacoux, C., Leclercq, V., Diamant, A., Lebrigand, K., Lazuka, A., Soyeux, E., Lacroix, S., Fassy, J., Couesnon, A., Thierry, R., Mari, B., Pradier, C., Waldmann, R., Barbry, P., 2021. Monitoring SARS-CoV-2 variants alterations in Nice neighborhoods by wastewater nanopore sequencing. *Lancet Reg. Health - Eur.* 10, 100202. <https://doi.org/10.1016/j.lanepe.2021.100202>.
- Robishaw, J.D., Alter, S.M., Solano, J.J., Shih, R.D., DeMets, D.L., Maki, D.G., Hennekens, C.H., 2021. Genomic surveillance to combat COVID-19: challenges and opportunities. *Lancet Microbe* 2, e481–e484. [https://doi.org/10.1016/S2666-5247\(21\)00121-X](https://doi.org/10.1016/S2666-5247(21)00121-X).
- Smyth, D.S., Trujillo, M., Gregory, D.A., Cheung, K., Gao, A., Graham, M., Guan, Y., Guldenpfennig, C., Hoxie, I., Kannoly, S., Kubota, N., Lyddon, T.D., Markman, M., Rushford, C., San, K.M., Sompanya, G., Spagnolo, F., Suarez, R., Teixeira, E., Daniels, M., Johnson, M.C., Dennehy, J.J., 2022. Tracking cryptic SARS-CoV-2 lineages detected in NYC wastewater. *Nat. Commun.* 13, 635. <https://doi.org/10.1038/s41467-022-28246-3>.
- Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E.J., Msomi, N., Mlisana, K., von Gottberg, A., Walaza, S., Allam, M., Ismail, A., Mohale, T., Glass, A.J., Engelbrecht, S., Van Zyl, G., Preiser, W., Petrucci, F., Sigal, A., Hardie, D., Marais, G., Hsiao, N., Korsman, S., Davies, M.-A., Tyers, L., Mudau, I., York, D., Maslo, C., Goedhals, D., Abrahams, S., Laguda-Akingba, O., Alisoltani-Dehkordi, A., Godzik, A., Wibmer, C.K., Sewell, B.T., Lourenço, J., Alcantara, L.C.J., Kosakovsky Pond, S.L., Weaver, S., Martin, D., Lessells, R.J., Bhiman, J.N., Williamson, C., de Oliveira, T., 2021. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* 592, 438–443. <https://doi.org/10.1038/s41586-021-03402-9>.
- Tyson, J.R., James, P., Stoddart, D., Sparks, N., Wickenhagen, A., Hall, G., Choi, J.H., Lapointe, H., Kamelian, K., Smith, A.D., Prystajek, N., Goodfellow, I., Wilson, S.J.,

- Harrigan, R., Snutch, T.P., Loman, N.J., Quick, J., 2020. Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore. *BioRxiv Prepr. Serv. Biol.* <https://doi.org/10.1101/2020.09.04.283077>.
- Ward, T., Glaser, A., Johnsen, A., Xu, F., Hall, I., Pellis, L., 2021. Growth, reproduction numbers and factors affecting the spread of SARS-CoV-2 novel variants of concern in the UK from October 2020 to July 2021: a modelling analysis. *BMJ Open* 11, e056636. <https://doi.org/10.1136/bmjopen-2021-056636>.
- Wickham, H., 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag, New York, New York.
- Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y., Yuan, M.L., Zhang, Y.L., Dai, F.H., Liu, Y., Wang, Q.M., Zheng, J.J., Xu, L., Holmes, E.C., Zhang, Y.Z., 2020. A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. <https://doi.org/10.1038/s41586-020-2008-3>.
- Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., Zhan, F., Ma, X., Wang, D., Xu, W., Wu, G., Gao, G.F., Tan, W., 2020. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* 382, 727–733. <https://doi.org/10.1056/NEJMoa2001017>.