

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

5 F.S. Brunner<sup>1,\*</sup>, M.R. Brown<sup>2,3,\*</sup>, I. Bassano<sup>2,4</sup>, H. Denise<sup>2</sup>, M.S. Khalifa<sup>2,5</sup>, M. Wade<sup>2,3</sup>, J.L. Kevill<sup>6</sup>, D.L. Jones<sup>6,7</sup>, K. Farkas<sup>6</sup>, A.R. Jeffries<sup>8</sup>, The COVID-19 Genomics UK (COG-UK) Consortium<sup>9</sup>, E. Cairns<sup>1</sup>, C. Wierzbicki<sup>1</sup>, S. Paterson<sup>1</sup>

<sup>1</sup>Institute of Infection, Veterinary and Ecological Sciences, University of Liverpool CH64 7TE, UK

10 <sup>2</sup>Environmental Monitoring for Health Protection, UK Health Security Agency, Windsor House, London SW1H 0TL, UK

<sup>3</sup>School of Engineering, Newcastle University, Newcastle-upon-Tyne NE1 7RU, UK

<sup>4</sup>Department of Infectious Disease, Imperial College London, London SW7 2AZ, UK

<sup>5</sup>Division of Biosciences, College of Health, Medicine and Life Sciences, Brunel University, London, UB8 3PH, UK.

15 <sup>6</sup>Centre for Environmental Biotechnology, School of Natural Sciences, Bangor University, Bangor, Gwynedd LL57 2UW, UK

<sup>7</sup>Food Futures Institute, Murdoch University, Murdoch WA 6105, Australia

<sup>8</sup>Biosciences, College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4QD, UK

20 <sup>9</sup> <https://www.cogconsortium.uk>, Full list of consortium names and affiliations are in the appendix

\* These authors contributed equally to the project

### 25 **Abstract**

Genomic surveillance of SARS-CoV-2 has been essential to provide an evidence base for public health decisions throughout the SARS-CoV-2 pandemic. Sequencing data from clinical cases has provided data crucial to understanding disease transmission and the detection, surveillance, and containment of outbreaks of novel variants, which continue to pose fresh challenges. However, genomic wastewater surveillance can provide important complementary information by providing estimates of variant frequencies which do not suffer from sampling bias, and capturing all variants circulating in a population. 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. Overall, the correspondence

30  
35

between wastewater and clinical variant frequencies demonstrates the reliability of wastewater surveillance. Yet, discrepancies between the two approaches when the Alpha variant was first detected emphasises that wastewater monitoring can also  
40 capture missing information resulting from asymptomatic cases or communities less engaged with testing programmes, as found by a simultaneous surge testing effort across the city.

## **Introduction**

Genomic surveillance has been a significant feature in the public health response to  
45 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; Vöhringer et al. 2021). Important examples include the B.1.1.7 (Rambaut et al. 2020), B.1.351 (Tegally et al. 2020), P.1 (Faria et al. 2021), B.1617.2 (Cherian et al. 2021) and, most recently, the B.1.1.529 lineage (Qin et al. 2021), named VOC Alpha,  
50 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 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  
55 surveillance is essential to mitigate and contain the threat they pose 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, and serves as an early warning system for the emergence and spread of novel variants (Mishra et al. 2021; Tegally et al. 2020).  
60 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

65 evolution and epidemiology of the SARS-CoV-2 pandemic (Ward et al. 2021; Harvey  
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), plus similar costs for staff, logistics and data infrastructure). It  
may be unsustainable at the levels required to adequately inform public health  
70 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 (Jahn et al. 2021; Hillary et al. 2021; Peccia et al. 2020; Smyth et al. 2021;  
75 Mishra et al. 2021). 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 (Peccia et al. 2020; Hillary et al. 2021; Wade et al.  
80 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 (Jahn et  
85 al. 2021; Hillary et al. 2021; Peccia et al. 2020; Smyth et al. 2021; Brown et al. 2021).  
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,  
90 particularly in urban centres (Polo et al. 2020; Green et al. 2021).

Nevertheless, moving from detection and quantification of 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 subsequently 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 of VOC in particular (Jahn et al. 2021; Fontenele et al. 2021).

This 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 major factors limiting uptake (Green et al. 2021). Wastewater-based epidemiology 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 Alpha (B.1.1.7) and subsequently Delta (B.1.617.2) SARS-CoV-2 variants (Vöhringer et al. 2021). We show that wastewater genomic surveillance can reliably detect the emergence of these variants and their subsequent rise across a city.

## **Methods**

*Sample Collection, Concentration and RNA Extraction*

Wastewater grab samples (1 L per sample) were collected from eight locations across the Liverpool sewer network and from Liverpool's main wastewater treatment plant at Sandon Docks (WWTP) between the 2<sup>nd</sup> of November 2020 until the 21<sup>st</sup> of June 2021, as part of the ongoing Environmental Monitoring for Health Protection programme (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 between 2<sup>nd</sup> of September 2020 and 17<sup>th</sup> of January 2021 as a control group. Samples were transported and subsequently stored at 4 - 6°C until 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 (10,000xg, 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.

#### *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  
150 cleaned with Mag-Bind® Total Pure NGS beads (Omega Bio-Tek) before sequencing on an Illumina NovaSeq™ 6000 platform generating 2x150bp paired end reads.

### *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,  
155 reads were aligned to the SARS-CoV-2 reference genome (MN908947.3, (Wu et al. 2020)) using the Burrow-Wheeler Aligner (bwa) (Li and Durbin 2009). 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  
160 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 further used custom scripts to extract summary statistics and sequence quality indicators, such as genome  
165 coverage, mapped reads and read depth (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  
170 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

175 mutations of the B.1.1.7 (VOC-20DEC-01, alpha) and B.1.617.2 (VOC-21APR-02,  
delta) lineages. Since a proportion of signature mutations are shared amongst  
VOC/VUI, not all variants have a unique set of co-occurring mutations. B.1.1.7 has  
one unique pair of co-occurring mutations on amplicon 147, while B.1.617.2 only has  
one non-unique pair of mutations on amplicon 121, i.e., it is shared with other  
180 variants.

### *Statistical Analyses*

We used RStudio v. 1.4.1717 (RStudio Team 2021) and 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.  
185 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,  
190 (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,  $\text{logit}(p)$  with  $p$  the probability of observing the (variant)  
195 data and  $\text{logit}$  being the inverse of the standard logistic 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  
200 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” (Lenth 2021), Table S4).

205 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. 210 2021) on 26<sup>th</sup> of October 2021 and then filtered and grouped 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 215 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).

## Results

Across all catchments, we observed a significant increase in the mean frequency of 220 Alpha (B.1.1.7) signature SNPs/indels between the 2<sup>nd</sup> November 2020 and the 28<sup>th</sup> 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 Alpha began in mid 225 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 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).

230 We detected local differences 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  
235 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).

From 15<sup>th</sup> March to 26<sup>th</sup> June 2021, we observed a significant increase in the  
240 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). It is noteworthy that the observed transition from Alpha to Delta in  
245 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 (J. Kevill & D. Jones, unpublished data). 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  
250 sporadic during this period (Figs. 4, 5 and S4). Co-occurrence analysis was not informative for the Delta variant since the only pair of signature mutations co-occurring on the same amplicon of our panel is not unique to the Delta variant but also part of the B.1.629 and AY.3 variants, among others. This explains why we detected a co-occurrence signal as early as November for all sites (Fig. S4).

255 Finally, we also observed signature SNPs of other VOC and VUI in wastewater

samples across Liverpool (Fig. S6), however detections were sporadic and at low predicted frequencies, which corresponds to clinical cases. Although there were notable exceptions, including the B.1.351 (Beta), P.1 (Gamma), P.3 (Theta) and B.1.1.318 lineages (Fig. S6).

## 260 **Discussion**

Genomic surveillance of wastewater has already shown great promise throughout the unfolding SARS-CoV-2 pandemic (Polo et al. 2020; Farkas 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  
265 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 Alpha (B.1.1.7) in early winter 2020 and, despite lower quality data, the rise of Delta (B.1.617.2/AY.4) in spring 2021. This clear reflection of the rise of Alpha and Delta, respectively, in clinical and wastewater genomic data  
270 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  
275 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 is the case 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  
280 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  
285 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, etc  
(Polo et al. 2020)), and the increasing stochasticity of clinical detection with  
decreasing population size play a role. Certainly the relationships between population  
size, wastewater flow variation and SARS-CoV-2 variant detection warrant further  
290 investigation.

This mixed pattern of wastewater Alpha 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  
295 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 Alpha 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  
300 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 frequencies in corresponding  
305 clinical and wastewater data, we find that each variant, in turn, reaches complete  
dominance, with estimated frequencies of 100% and ~75% in clinical and wastewater  
samples, respectively. This suggests improved statistical methods may be required  
to estimate lineage frequencies from pooled sequencing data, as produced from  
wastewater samples. Here we have relied on a relatively crude estimation by taking  
310 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 (Quince et al. 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. The use of a relatively short amplicon scheme, as used here, may help to boost sensitivity given the likely degradation of SARS-CoV-2 RNA in wastewater; and from initial work gives better genome coverage than the longer ARTIC v3 amplicons more widely used for sequencing SARS-CoV-2 (data not shown). As previously noted, however, the use of short amplicons limits the use of co-occurrence analysis for variant detection. It is also 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 (Polo et al. 2020; Hillary et al. 2021; Kevill et al. 2022).

In conclusion, we show that wastewater genomic sequencing can detect the emergence and rise of new SARS-CoV-2 variants. Results correspond well with those obtained through genomic sequencing of clinical samples. In some cases, variants are observed prior to clinical detections, which may be particularly useful in areas or communities with low testing uptake.

335

### **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

340 (NE/V003860/1) and DHSC UK (2020\_097). 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.

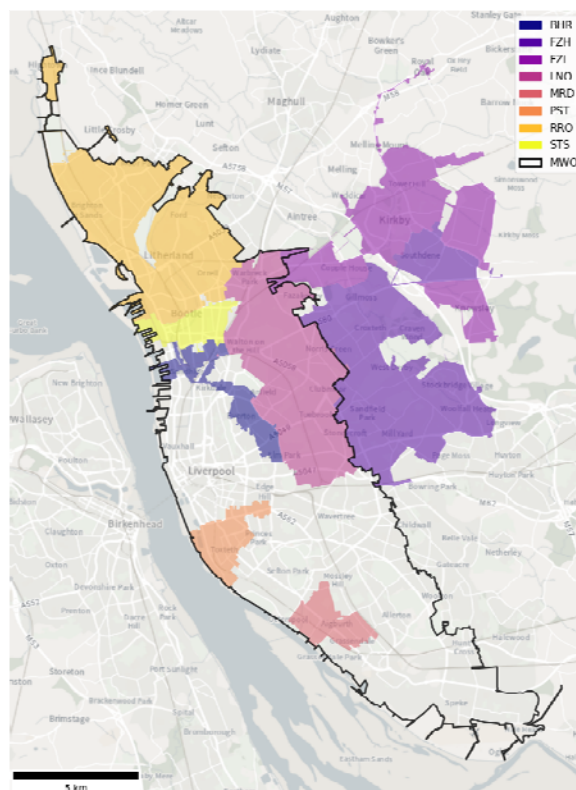
#### 345 **Data accessibility statement**

Wastewater sequencing data will be deposited and publicly available on the European Nucleotide Archive by 1<sup>st</sup> July 2022. The clinical case data used in this study is visualised at <https://www.cogconsortium.uk/tools-analysis/public-data-analysis-2/>. A filtered, privacy conserving version of the lineage-LTLA-week dataset  
350 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.

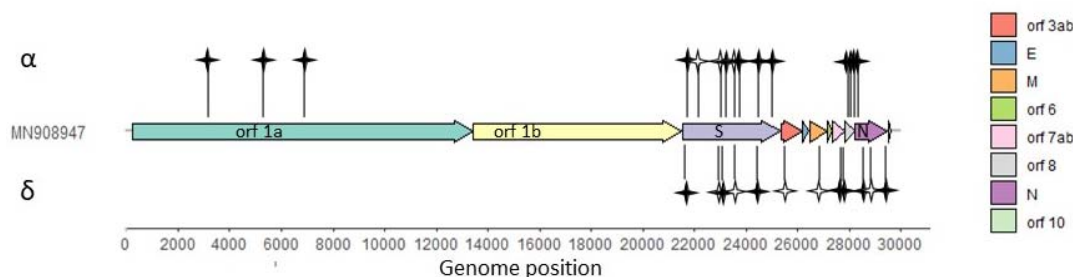
#### **Ethics Statement**

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

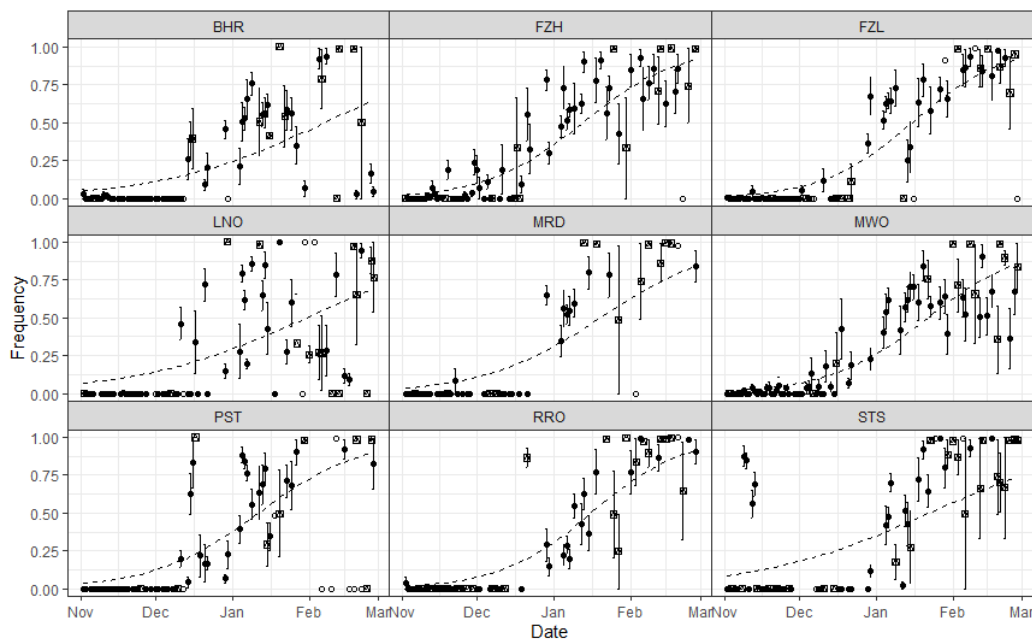
## Figures



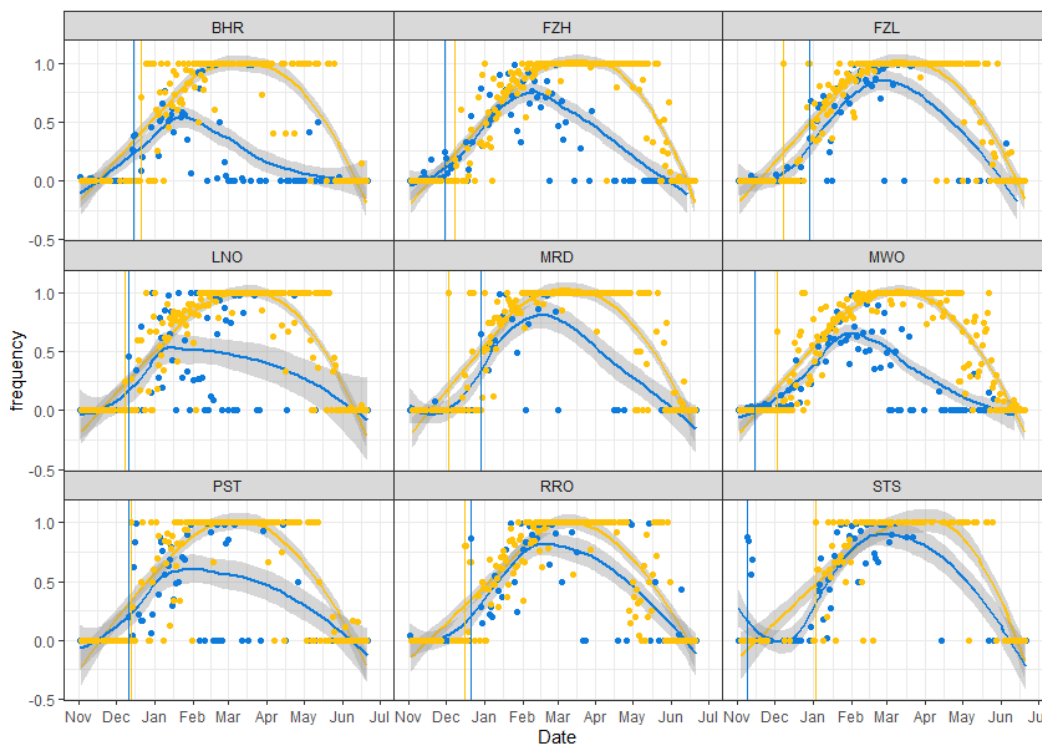
360 **Figure 1.** Wastewater catchments of the 8 sewer network sampling locations (coloured) and the WWTW (MWO, black outline) across Liverpool. BHR= Bank Hall Relief, FZH= Fazakerley High, FZL= Fazakerley Low, LNO= Liverpool North, MRD= Mersey Road, PST= Park Street, RRO= Rimrose, STS= Strand SSO, MWO= Sandon Dock Main Works. See Table S1 for further catchment details.



365 **Figure 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  
 370 (Wu et al. 2020).

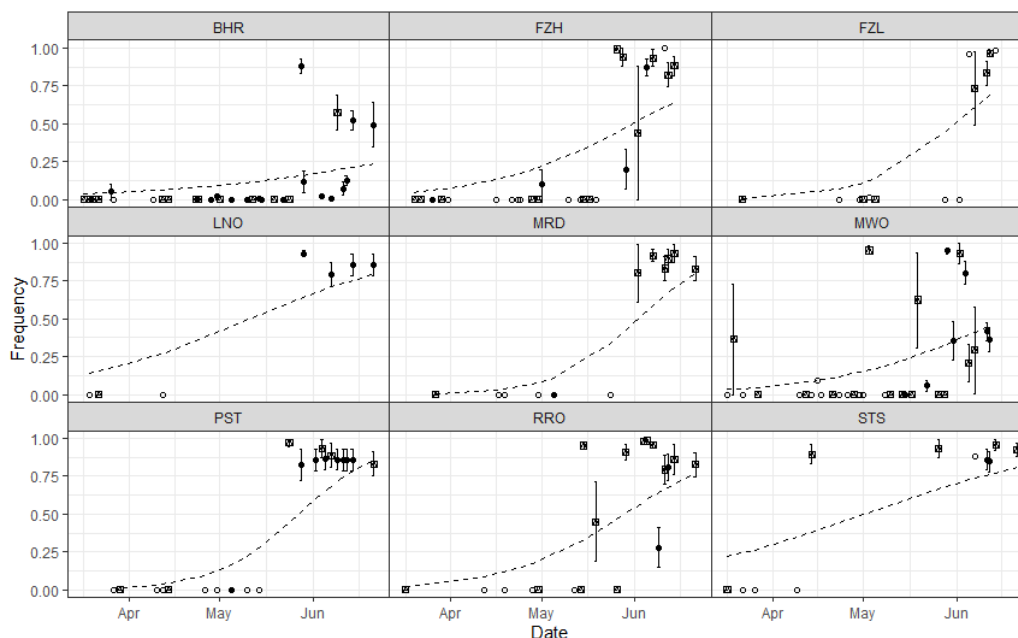


375 **Figure 3.** Mean frequency of B.1.1.7 (Alpha) signature SNPs/Indels in each wastewater catchment, 2<sup>nd</sup> November 2020 to 28<sup>th</sup> February 2021. Points and error bars show means and standard errors across all unique Alpha mutations with sufficient sequencing coverage in each sample.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



380

385 **Figure 4.** Mean frequency of B.1.1.7 (Alpha) signature SNPs/Indels detected in wastewater versus clinical samples in each catchment, 2<sup>nd</sup> November 2020 to 26<sup>th</sup> June 2021. Points show the mean frequency of unique Alpha mutations for a given wastewater sample (blue) and the frequency of Alpha 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 (yellow) and the first wastewater detection of co-occurring Alpha mutations on amplicon 147 (blue).



390 **Figure 5.** Mean frequency of B.1.617.2 (Delta) signature SNPs/Indels in each wastewater catchment, 15<sup>th</sup> March 2021 to 26<sup>th</sup> June 2021. Points and error bars show means and standard errors across all unique Alpha 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 mutations used in the mean calculation for a given sample: empty circles: 1 mutation, crossed square: 2 to 5 mutations, filled circles: >5 mutations.

## References

- 400 Brown, Mathew R, Matthew J Wade, Shannon McIntyre-Nolan, Irene Bassano, Hubert Denise, David Bass, John Bentley, et al. 2021. 'Wastewater Monitoring of SARS-CoV-2 Variants in England: Demonstration Case Study for Bristol (Dec 2020 - March 2021) Summary for SAGE 08/04/21', 14.
- 405 Cherian, Sarah, Varsha Potdar, Santosh Jadhav, Pragya Yadav, Nivedita Gupta, Mousmi Das, Partha Rakshit, et al. 2021. 'Convergent Evolution of SARS-CoV-2 Spike Mutations, L452R, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India'. Preprint. Molecular Biology. <https://doi.org/10.1101/2021.04.22.440932>.
- Cribari-Neto, Francisco, and Achim Zeileis. 2010. 'Beta Regression in R'. *Journal of Statistical Software* 34 (2). <https://doi.org/10.18637/jss.v034.i02>.
- 410 Danecek, Petr, James K. Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O. Pollard, Andrew Whitwham, et al. 2021. 'Twelve Years of SAMtools



- and BCFtools.' *GigaScience* 10 (2).  
<https://doi.org/10.1093/gigascience/giab008>.
- 415 Davies, Nicholas G., Sam Abbott, Rosanna C. Barnard, Christopher I. Jarvis, Adam J. Kucharski, James D. Munday, Carl A. B. Pearson, et al. 2021. 'Estimated Transmissibility and Impact of SARS-CoV-2 Lineage B.1.1.7 in England'. *Science* 372 (6538): eabg3055–eabg3055.  
<https://doi.org/10.1126/science.abg3055>.
- 420 Elbe, Stefan, and Gemma Buckland-Merrett. 2017. 'Data, Disease and Diplomacy: GISAID's Innovative Contribution to Global Health'. *Global Challenges* 1 (1): 33–46. <https://doi.org/10.1002/gch2.1018>.
- Faria, Nuno R., Thomas A. Mellan, Charles Whittaker, Ingra M. Claro, Darlan da S. Candido, Swapnil Mishra, Myuki A. E. Crispim, et al. 2021. 'Genomics and Epidemiology of the P.1 SARS-CoV-2 Lineage in Manaus, Brazil'. *Science* 372 (6544): 815–21. <https://doi.org/10.1126/science.abh2644>.
- 425 Farkas, Kata, Luke S. Hillary, Shelagh K. Malham, James E. McDonald, and David L. Jones. 2020. 'Wastewater and Public Health: The Potential of Wastewater Surveillance for Monitoring COVID-19'. *Current Opinion in Environmental Science & Health* 17 (October): 14–20.  
<https://doi.org/10.1016/j.coesh.2020.06.001>.
- 430 Farkas, Kata, Luke S. Hillary, Jamie Thorpe, David I. Walker, James A. Lowther, James E. McDonald, Shelagh K. Malham, and Davey L. Jones. 2021. 'Concentration and Quantification of SARS-CoV-2 RNA in Wastewater Using Polyethylene Glycol-Based Concentration and QRT-PCR'. *Methods and Protocols* 4 (1): 17. <https://doi.org/10.3390/mps4010017>.
- 435 Foladori, Paola, Francesca Cutrupi, Nicola Segata, Serena Manara, Federica Pinto, Francesca Malpei, Laura Bruni, and Giuseppina La Rosa. 2020. 'SARS-CoV-2 from Faeces to Wastewater Treatment: What Do We Know? A Review'. *Science of The Total Environment* 743 (November): 140444.  
<https://doi.org/10.1016/j.scitotenv.2020.140444>.
- 440 Fontenele, Rafaela S., Simona Kraberger, James Hadfield, Erin M. Driver, Devin Bowes, LaRinda A. Holland, Temitope O.C. Faleye, et al. 2021. 'High-Throughput Sequencing of SARS-CoV-2 in Wastewater Provides Insights into Circulating Variants'. *Water Research* 205 (October): 117710.  
<https://doi.org/10.1016/j.watres.2021.117710>.
- 445 García-Fiñana, Marta, David M Hughes, Christopher P Cheyne, Girvan Burnside, Mark Stockbridge, Tom A Fowler, Veronica L Fowler, Mark H Wilcox, Malcolm G Semple, and Iain Buchan. 2021. 'Performance of the Innova SARS-CoV-2 Antigen Rapid Lateral Flow Test in the Liverpool Asymptomatic Testing Pilot: Population Based Cohort Study'. *BMJ*, July, n1637.  
<https://doi.org/10.1136/bmj.n1637>.
- 450 Green, Mark A., Marta García-Fiñana, Ben Barr, Girvan Burnside, Christopher P. Cheyne, David Hughes, Matthew Ashton, Sally Sheard, and Iain E. Buchan. 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)'. *The Lancet Regional Health - Europe* 6 (July): 100107.  
<https://doi.org/10.1016/j.lanepe.2021.100107>.
- 455 Harvey, William T., Alessandro M. Carabelli, Ben Jackson, Ravindra K. Gupta, Emma C. Thomson, Ewan M. Harrison, Catherine Ludden, et al. 2021. 'SARS-CoV-2 Variants, Spike Mutations and Immune Escape'. *Nature Reviews Microbiology* 19 (7): 409–24. <https://doi.org/10.1038/s41579-021-00573-0>.
- 460 Hillary, Luke S., Kata Farkas, Kathryn H. Maher, Anita Lucaci, Jamie Thorpe, Marco A. Distaso, William H. Gaze, et al. 2021. 'Monitoring SARS-CoV-2 in Municipal Wastewater to Evaluate the Success of Lockdown Measures for
- 465

- Controlling COVID-19 in the UK'. *Water Research* 200: 117214–117214. <https://doi.org/10.1016/j.watres.2021.117214>.
- 470 Jahn, Katharina, David Dreifuss, Ivan Topolsky, Anina Kull, Pravin Ganesanandamoorthy, Xavier Fernandez-Cassi, Carola Bänziger, et al. 2021. 'Detection and Surveillance of SARS-CoV-2 Genomic Variants in Wastewater'. *MedRxiv*. <https://doi.org/10.1101/2021.01.08.21249379>.
- Jeffries, Aaron, Steve Paterson, Matthew Loose, and Ronny van Aerle. 2021. 'Wastewater Sequencing Using the EasySeq™ RC-PCR SARS CoV-2 (Nimagen) V1.0'. *Protocols.io*, 8.
- 475 Kevill, Jessica L., Cameron Pellett, Kata Farkas, Mathew R. Brown, Irene Bassano, Hubert Denise, James E. McDonald, et al. 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'. *Science of The Total Environment* 808 (February): 151916. <https://doi.org/10.1016/j.scitotenv.2021.151916>.
- 480 Koboldt, Daniel C., Qunyuan Zhang, David E. Larson, Dong Shen, Michael D. McLellan, Ling Lin, Christopher A. Miller, Elaine R. Mardis, Li Ding, and Richard K. Wilson. 2012. 'VarScan 2: Somatic Mutation and Copy Number Alteration Discovery in Cancer by Exome Sequencing'. *Genome Research* 22 (3): 568–76. <https://doi.org/10.1101/gr.129684.111>.
- 485 Lenth, Russell V. 2021. *Emmeans: Estimated Marginal Means, Aka Least-Squares Means* (version R package version 1.7.1-1). <https://CRAN.R-project.org/package=emmeans>.
- 490 Li, Heng, and Richard Durbin. 2009. 'Fast and Accurate Short Read Alignment with Burrows-Wheeler Transform'. *Bioinformatics (Oxford, England)* 25 (14): 1754–60. <https://doi.org/10.1093/bioinformatics/btp324>.
- Mishra, Swapnil, Sören Mindermann, Mrinank Sharma, Charles Whittaker, Thomas A Mellan, Thomas Wilton, Dimitra Klapsa, et al. 2021. 'Changing Composition of SARS-CoV-2 Lineages and Rise of Delta Variant in England'. *EClinicalMedicine* 39 (September): 101064. <https://doi.org/10.1016/j.eclinm.2021.101064>.
- 500 Nicholls, Samuel M., Radoslaw Poplawski, Matthew J. Bull, Anthony Underwood, Michael Chapman, Khalil Abu-Dahab, Ben Taylor, et al. 2021. 'CLIMB-COVID: Continuous Integration Supporting Decentralised Sequencing for SARS-CoV-2 Genomic Surveillance'. *Genome Biology* 22 (1): 196, s13059-021-02395-y. <https://doi.org/10.1186/s13059-021-02395-y>.
- 505 Peccia, Jordan, Alessandro Zulli, Doug E. Brackney, Nathan D. Grubaugh, Edward H. Kaplan, Arnau Casanovas-Massana, Albert I. Ko, et al. 2020. 'Measurement of SARS-CoV-2 RNA in Wastewater Tracks Community Infection Dynamics'. *Nature Biotechnology* 38 (10): 1164–67. <https://doi.org/10.1038/s41587-020-0684-z>.
- Polo, David, Marcos Quintela-Baluja, Alexander Corbishley, Davey L. Jones, Andrew C. Singer, David W. Graham, and Jesús L. Romalde. 2020. 'Making Waves: Wastewater-Based Epidemiology for COVID-19 – Approaches and Challenges for Surveillance and Prediction'. *Water Research* 186 (November): 116404. <https://doi.org/10.1016/j.watres.2020.116404>.
- 510 Qin, Si, Mengnan Cui, Siqi Sun, Jiyang Zhou, Zongmin Du, Yujun Cui, and Hang Fan. 2021. 'Genome Characterization and Potential Risk Assessment of the Novel SARS-CoV-2 Variant Omicron (B.1.1.529)'. *Zoonoses* 1 (1). <https://doi.org/10.15212/ZOONOSES-2021-0024>.
- 515 Quince, Christopher, Anders Lanzén, Russell J Davenport, and Peter J Turnbaugh. 2011. 'Removing Noise From Pyrosequenced Amplicons'. *BMC Bioinformatics* 12 (1): 38. <https://doi.org/10.1186/1471-2105-12-38>.

- 520 R Core Team. 2021. *R: A Language and Environment for Statistical Computing* (version 4.1.1). Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Rambaut, Andrew, Nick Loman, Oliver Pybus, Wendy Barclay, Jeff Barrett, Alessandro Carabelli, Tom Connor, et al. 2020. 'Preliminary Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in the UK Defined by a Novel Set of Spike Mutations'. <https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563>.
- 525
- Robishaw, Janet D, Scott M Alter, Joshua J Solano, Richard D Shih, David L DeMets, Dennis G Maki, and Charles H Hennekens. 2021. 'Genomic Surveillance to Combat COVID-19: Challenges and Opportunities'. *The Lancet Microbe* 2 (9): e481–84. [https://doi.org/10.1016/S2666-5247\(21\)00121-X](https://doi.org/10.1016/S2666-5247(21)00121-X).
- 530
- RStudio Team. 2021. *RStudio: Integrated Development Environment for R* (version 1.4.1717). Boston, MA: RStudio, PBC. <http://www.rstudio.com/>.
- 535
- Smyth, Davida S, Monica Trujillo, Devon A Gregory, Kristen Cheung, Anna Gao, Maddie Graham, Yue Guan, et al. 2021. 'Tracking Cryptic SARS-CoV-2 Lineages Detected in NYC Wastewater'. *MedRxiv*, 2021.07.26.21261142-2021.07.26.21261142.
- 540
- Tegally, H., E. Wilkinson, M. Giovanetti, A. Iranzadeh, V. Fonseca, J. Giandhari, D. Doolabh, et al. 2020. 'Emergence and Rapid Spread of a New Severe Acute Respiratory Syndrome-Related Coronavirus 2 (SARS-CoV-2) Lineage with Multiple Spike Mutations in South Africa'. *MedRxiv* 2. <https://doi.org/10.1101/2020.12.21.20248640>.
- 545
- Tyson, John R, Phillip James, David Stoddart, Natalie Sparks, Arthur Wickenhagen, Grant Hall, Ji Hyun Choi, et al. 2020. 'Improvements to the ARTIC Multiplex PCR Method for SARS-CoV-2 Genome Sequencing Using Nanopore.' *BioRxiv*: *The Preprint Server for Biology*. <https://doi.org/10.1101/2020.09.04.283077>.
- 550
- Vöhringer, Harald S., Theo Sanderson, Matthew Sinnott, Nicola De Maio, Thuy Nguyen, Richard Goater, Frank Schwach, et al. 2021. 'Genomic Reconstruction of the SARS-CoV-2 Epidemic in England'. *Nature*, October. <https://doi.org/10.1038/s41586-021-04069-y>.
- 555
- Wade, Matthew, Davey Jones, Andrew Singer, Alwyn Hart, Alex Corbishley, Mario Morvan, Chaoyi Zhang, et al. 2020. 'Wastewater COVID-19 Monitoring in the UK: Summary for SAGE – 19/11/20'.
- Ward, Thomas, Alex Glaser, Alexander Johnsen, Feng Xu, Ian Hall, and Lorenzo Pellis. 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 (11): e056636. <https://doi.org/10.1136/bmjopen-2021-056636>.
- 560
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. New York: Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- 565
- Wu, Fan, Su Zhao, Bin Yu, Yan Mei Chen, Wen Wang, Zhi Gang Song, Yi Hu, et al. 2020. 'A New Coronavirus Associated with Human Respiratory Disease in China'. *Nature* 579 (7798): 265–69. <https://doi.org/10.1038/s41586-020-2008-3>.
- 570
- Zhu, Na, Dingyu Zhang, Wenling Wang, Xingwang Li, Bo Yang, Jingdong Song, Xiang Zhao, et al. 2020. 'A Novel Coronavirus from Patients with Pneumonia in China, 2019'. *New England Journal of Medicine* 382 (8): 727–33. <https://doi.org/10.1056/NEJMoa2001017>.