



# Rapid Increase of SARS-CoV-2 Variant B.1.1.7 Detected in Sewage Samples from England between October 2020 and January 2021

Thomas Wilton,<sup>a</sup> Erika Bujaki,<sup>a</sup> Dimitra Klapsa,<sup>a</sup> Manasi Majumdar,<sup>a</sup> Maria Zambon,<sup>b</sup> Martin Fritzsche,<sup>c</sup> Ryan Mate,<sup>c</sup>  Javier Martin<sup>a</sup>

<sup>a</sup>Division of Virology, National Institute for Biological Standards and Control, South Mimms, Potters Bar, Hertfordshire, UK

<sup>b</sup>Respiratory Virology and Polio Reference Service, Public Health England, London, UK

<sup>c</sup>Division of Analytical and Biological Sciences, National Institute for Biological Standards and Control, South Mimms, Potters Bar, Hertfordshire, UK

**ABSTRACT** SARS-CoV-2 variants with multiple amino acid mutations in the spike protein are emerging in different parts of the world, raising concerns regarding their possible impact on human immune response and vaccine efficacy against the virus. Recently, a variant named lineage B.1.1.7 was detected and shown to be rapidly spreading across the UK since November 2020. As surveillance for these SARS-CoV-2 variants of concern (VOCs) becomes critical, we have investigated the use of environmental surveillance (ES) for the rapid detection and quantification of B.1.1.7 viruses in sewage as a way of monitoring its expansion that is independent on the investigation of identified clinical cases. Next-generation sequencing analysis of amplicons synthesized from sewage concentrates revealed the presence of B.1.1.7 mutations in viral sequences, first identified in a sample collected in London on 10 November 2020 and shown to rapidly increase in frequency to >95% in January 2021, in agreement with clinical data over the same period. We show that ES can provide an early warning of VOCs becoming prevalent in the population and that, as well as B.1.1.7, our method can detect VOCs B.1.351 and P.1, first identified in South Africa and Brazil, respectively, and other viruses carrying critical spike mutation E484K, known to have an effect on virus antigenicity. Although we did not detect such mutation in viral RNAs from sewage, we did detect mutations at amino acids 478, 490, and 494, located close to amino acid 484 in the spike protein structure and known to also have an effect on antigenicity.

**IMPORTANCE** The recent appearance and growth of new SARS-CoV-2 variants represent a major challenge for the control of the COVID-19 pandemic. These variants of concern contain mutations affecting antigenicity, which raises concerns on their possible impact on human immune response to the virus and vaccine efficacy against them. Here, we show how environmental surveillance for SARS-CoV-2 can be used to help us understand virus transmission patterns and provide an early warning of variants becoming prevalent in the population. We describe the detection and quantification of variant B.1.1.7, first identified in southeast England in sewage samples from London (UK) before widespread transmission of this variant was obvious from clinical cases. Variant B.1.1.7 was first detected in a sample from early November 2020, with the frequency of B.1.1.7 mutations detected in sewage rapidly increasing to >95% in January 2021, in agreement with increasing SARS-CoV-2 infections associated with B.1.1.7 viruses.


**KEYWORDS** SARS-CoV-2, environmental surveillance, variant B.1.1.7, variant of concern, COVID-19, sewage, vaccine, wastewater, next-generation sequencing, direct detection, B.1.1.7, surveillance

**Citation** Wilton T, Bujaki E, Klapsa D, Majumdar M, Zambon M, Fritzsche M, Mate R, Martin J. 2021. Rapid increase of SARS-CoV-2 variant B.1.1.7 detected in sewage samples from England between October 2020 and January 2021. *mSystems* 6:e00353-21. <https://doi.org/10.1128/mSystems.00353-21>.

**Editor** Charles R. Langelier, University of California—San Francisco

© Crown copyright 2021. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Javier Martin, [javier.martin@nibsc.org](mailto:javier.martin@nibsc.org).

 Rapid increase of SARS-CoV-2 variant B.1.1.7 detected in sewage samples from England between October 2020 and January 2021 using next generation sequencing analysis of DNA amplicons directly synthesised from wastewater concentrates.

**Received** 24 March 2021

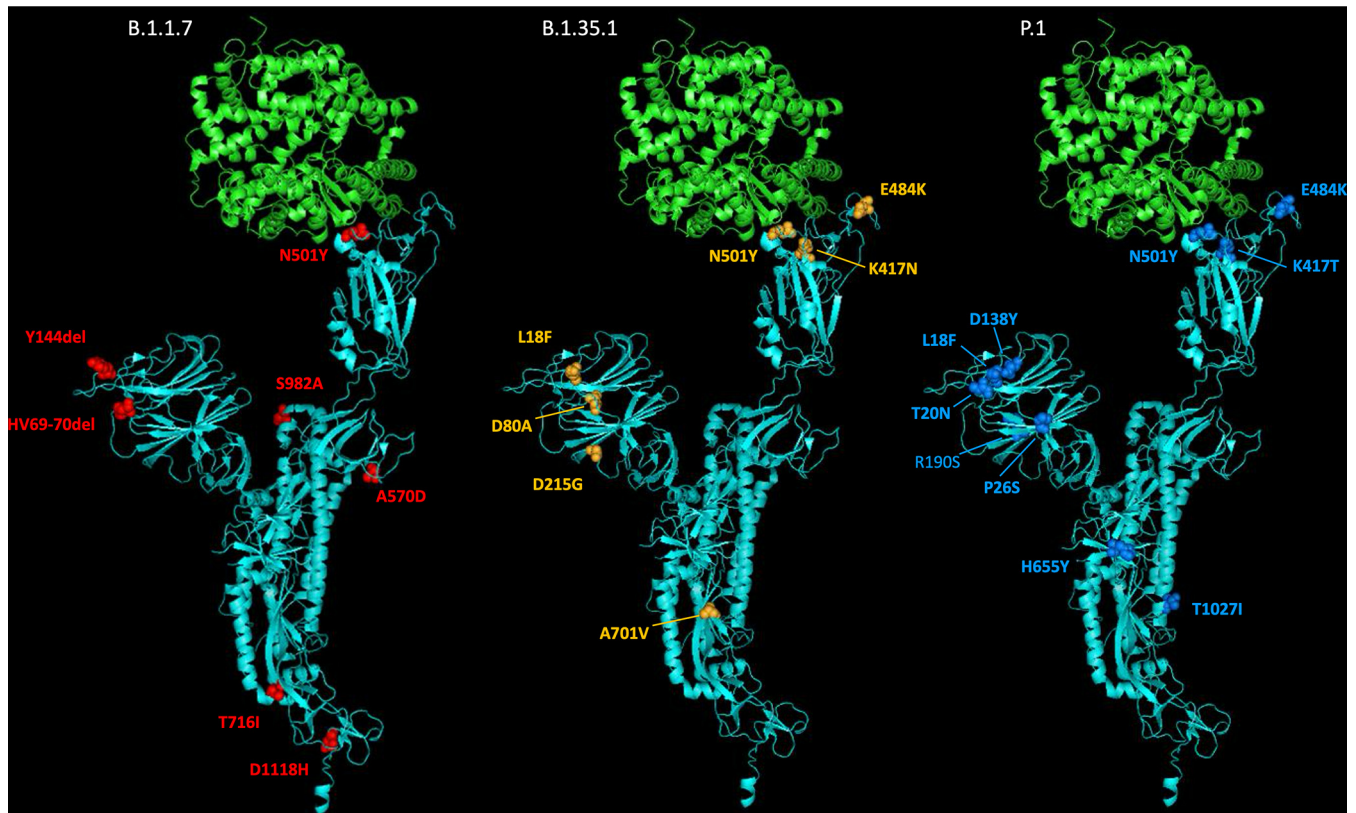
**Accepted** 19 May 2021

**Published** 15 June 2021

A novel SARS-CoV-2 lineage B.1.1.7 was recently identified and shown to be rapidly expanding across the UK, particularly in London and in the southeast and the east of England (1, 2). This lineage was detected in November 2020 and likely originated in September 2020 in Kent, southeast England. B.1.1.7 viruses possess a striking constellation of nucleotide changes with an unusual number of missense mutations in the spike protein of potential biological relevance (3) (Fig. 1). For this reason, lineage B.1.1.7, was designated SARS-CoV-2 variant of concern (VOC) 202012/01 by Public Health England (PHE) since the mutations found might impact transmission, immune response, and disease severity. Mutation N501Y, located within the receptor-binding domain (RBD), can increase binding affinity to human and murine angiotensin-converting enzyme 2 (ACE-2) receptor and cell infectivity in mice (4–6). Substitution P681H is immediately adjacent to the furin cleavage site, important for viral pathogenesis (7, 8), whereas deletion of amino acids 69 and 70 has been identified in multiple lineages associated with different RBD mutations and has been related to the evasion to human immune response (9). The deletion of nucleotides 21765 to 21770 coding for amino acids 69 and 70 has been associated with S-gene target failure (SGTF) in diagnostic PCR testing of samples showing positive results for other gene targets (1) and is used to estimate the proportion of COVID-19 confirmed cases likely associated with infection by variant B.1.1.7. Epidemiological evidence and phylogenetic analysis of viral sequences suggest that viruses of the B.1.1.7 lineage may have increased transmissibility compared to ancestral isolates (2). Whole-genome sequencing and SGTF data have shown a rapid increase in the prevalence of B.1.1.7 variant through time with the proportion of sequences from clinical samples associated with B.1.1.7 in England, reaching around 94% during the last week of January 2021 (1, 2, 10). Additional analyses indicate exponential growth of lineage B.1.1.7 and an increase of between 0.4 and 0.7 in the reproduction number with respect to previous lineages (2).

Other VOCs have been identified recently, notably lineages B.1.351 and P.1, known to be circulating in South Africa since early October 2020 (11) and in Brazil since early December 2020 (12), respectively, and also containing complex mutation constellations, including an N501Y amino acid change (Fig. 1). B.1.351 and P.1 variants have two more spike mutations of biological significance: E484K and K417N/T, both located in the RBD and both important for ACE-2 binding and antibody recognition. Mutations at spike amino acid 484 have been shown to reduce binding to convalescent-phase serum antibodies, with neutralization by some sera reduced >10-fold (13). Amino acid substitutions at residue 417, K417N in B.1.351 and K417T in P.1, appear to improve evasion from antibodies in combination with N501Y and E484K (14). More recently, mutation E484K has been identified in a small number of B.1.1.7 viruses in England (15), raising increased concerns about this variant. The effects of mutations found in the VOCs detailed above suggest that VOCs could be a threat to the protective efficacy of current vaccines by reducing vaccine-induced immunity against them. Recent research has shown that while B.1.1.7 is modestly resistant to sera from vaccinees, B.1.351 is markedly more resistant to neutralization by such sera, and some COVID-19 vaccines appear to show reduced efficacy against this variant (16–20). Although further research is required to fully understand the effect of spike mutations on vaccine efficacy, updating vaccine strains might be required in the future in a similar manner to what is common practice for flu vaccines.

The appearance and rapid growth of VOCs shows the need for enhanced genomic and epidemiological surveillance worldwide to ensure that any new changes in the spike protein that might arise are rapidly detected and their biological effects investigated. Environmental surveillance (ES) has proven to be a sensitive method for the detection and monitoring of SARS-CoV-2 circulation, as described in multiple reports from many countries (reviewed in reference 21). Since ES can produce a real-time snapshot of virus transmission at a given time point, including that from asymptomatic infections, it might provide precise information on the prevalence of VOCs in the population. We were able to detect SARS-CoV-2 viral RNA in sewage samples collected in London during the first wave of the pandemic with variation in viral RNA levels



**FIG 1** Molecular cartoon diagram of the 3D structure of SARS-CoV-2 spike protein monomer in the open form (cyan) in complex with receptor ACE-2 (green). The image was generated using PyMOL Molecular Graphics System version 1.7.0.3 software (Schrödinger, LLC) using cryo-EM data (Protein Data Bank accession number 7DF4 [36]). Amino acid substitutions found in B.1.1.7 (red), B.1.35.1 (orange), and P.1 (blue) variants are shown.

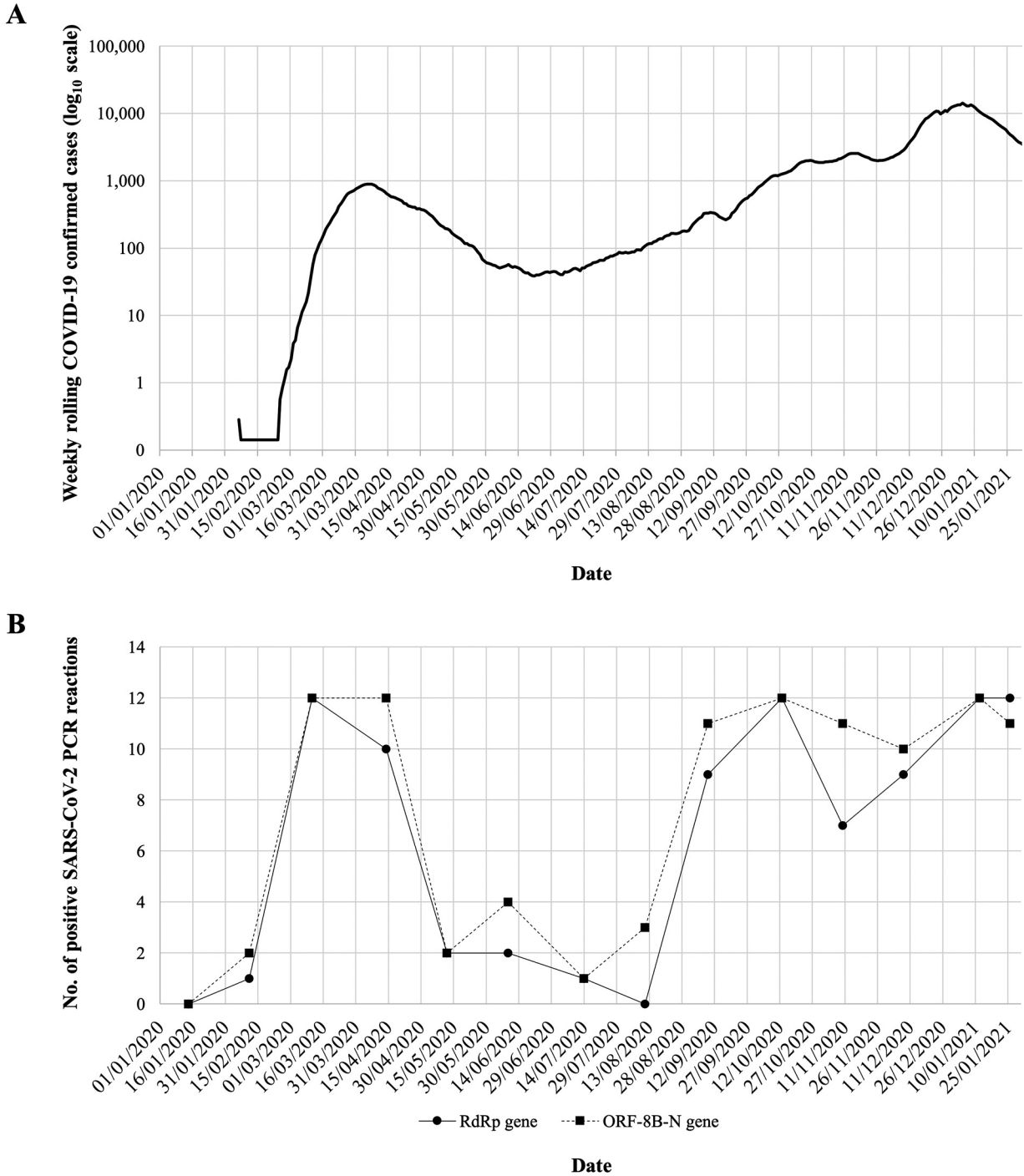
comparing to that of numbers of confirmed COVID-19 cases (22). In addition, using next-generation sequencing (NGS) analysis of PCR amplicons generated from sewage concentrates, we were able to track changes in variant predominance during the first stages of the pandemic detecting variants that were particularly prevalent in the UK and the rapid expansion of D614G variant reaching nearly complete dominance in May 2020 (22). This variant contains a mutation from A to G at nucleotide 23403, resulting in amino acid change from aspartic acid to glycine at residue 614 of the spike protein, which appeared to increase viral infectivity and transmissibility. Variant D614G was first described in late February 2020 in Italy, rapidly expanding and becoming the dominant SARS-CoV-2 variant globally a few months later (23). Similar work by other groups has described the detection of cocirculating SARS-CoV-2 variants in sewage samples from different locations, including a recent report describing the identification of B.1.1.7 lineage in sewage samples from Switzerland (24–28).

In order to further evaluate the value of ES for SARS-CoV-2 detection, we analyzed sewage samples collected in London between 14 January 2020 and 26 January 2021 for the presence of SARS-CoV-2 by using a semiquantitative nested reverse transcription-PCR (nPCR) system targeting two different genomic regions since we have found this system to be more sensitive and consistent than quantitative RT-PCR assays (22). More importantly, we have designed new nPCR reactions to generate DNA amplicons and analyze them by NGS in order to specifically detect and quantify the presence of key mutations that discriminate lineages B.1.1.7, B.1.35.1, and P.1 from ancestral isolates and between themselves.

## RESULTS

### nPCR amplifications to estimate SARS-CoV-2 RNA levels in sewage samples.

Twelve RNA replicates from each sewage concentrate were used to generate nPCR products with primer combinations A and B (see Table S1 in the supplemental



**FIG 2** (A) Number of daily confirmed COVID-19 cases in London. Values are represented as 7-day rolling averages. Nationwide lockdown periods are indicated as horizontal lines. The source for the COVID-19 case data is available online (<https://coronavirus.data.gov.uk/>). (B) Detection of SARS-CoV-2 RNA in sewage samples collected in London between 14 January 2020 and 26 January 2021. The numbers of RNA replicates, out of a total of 12, producing positive results in nPCR A (circles) and nPCR B (squares) reactions targeting RdRP and ORF8b genomic regions, respectively, are represented for each sample date.

material) targeting RNA-dependent RNA polymerase (RdRP) and ORF8b gene regions, respectively. Fourteen sewage samples collected between 14 January 2020 and 26 January 2021 were analyzed. The proportion of positive PCR results for each sample was used as a semiquantitative comparison of RNA levels present in sewage samples. The results are shown in Fig. 2 in the context of epidemiological data. As we have shown before, a sample from 14 January 2020 was negative, and only low levels of viral

**TABLE 1** Nucleotide and amino acid differences between VOCs and ancestral isolates<sup>a</sup>

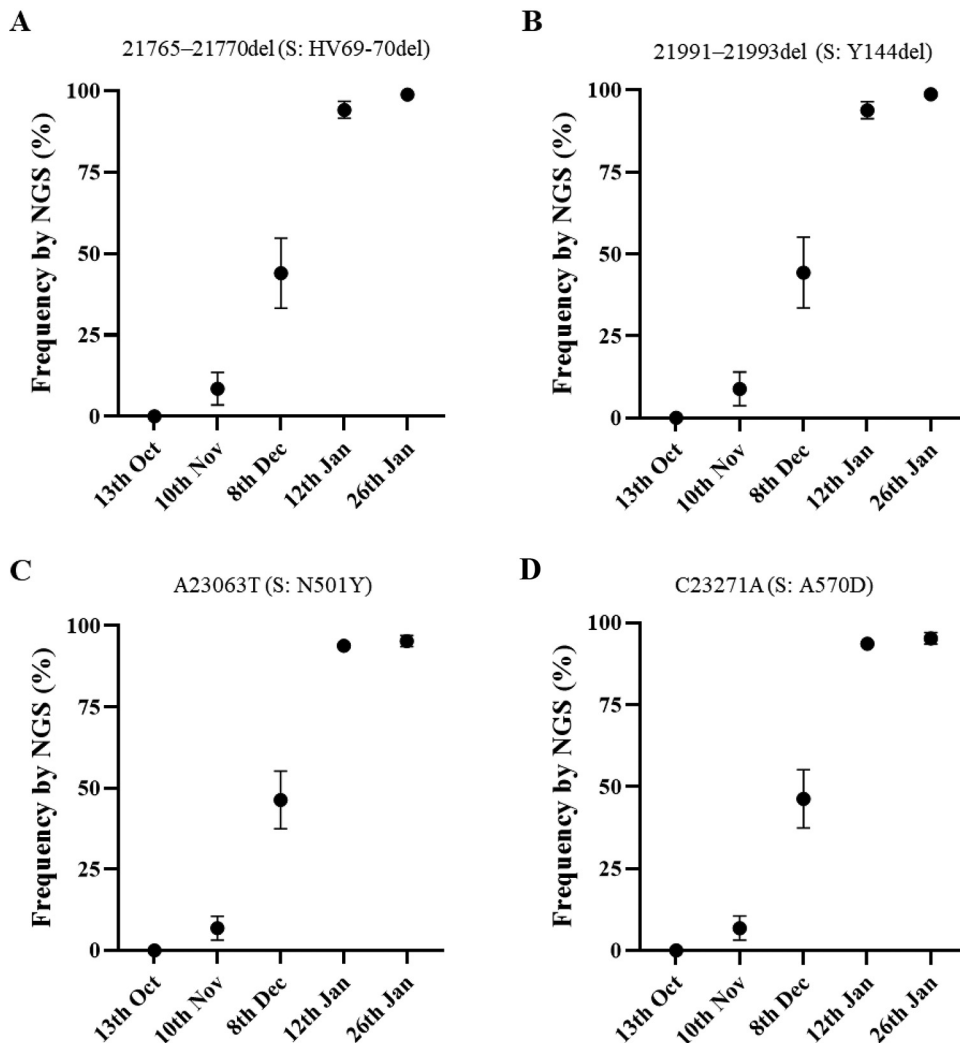
Reaction	Genome position		nt				aa			
	nt	aa	WT	B.1.1.7	B.1.351	P.1	WT	B.1.1.7	B.1.351	P.1
nPCR C	21765–21770	69–70	ACATGT	<i>del</i>	ACATGT	ACATGT	HV	<i>del</i>	HV	HV
	21801	80	A	A	<u>C</u>	A	D	D	<u>A</u>	D
	21974	138	G	G	G	<u>T</u>	D	D	D	<u>Y</u>
	21991–21993	144	TTA	<i>del</i>	TTA	TTA	Y	<i>del</i>	Y	Y
nPCR D	23012	484	G	G	<u>A</u>	<u>A</u>	E	E	<u>K</u>	<u>K</u>
	23063	501	A	<u>T</u>	<u>T</u>	<u>T</u>	N	<u>Y</u>	<u>Y</u>	<u>Y</u>
	23271	570	C	<u>A</u>	C	C	A	<u>D</u>	A	<u>A</u>

<sup>a</sup>WT, Wuhan-Hu-1 strain (GenBank accession number [MN908947](https://www.ncbi.nlm.nih.gov/nuccore/MN908947)). This strain was first detected in the UK (B.1.1.7), South Africa (B.1.351), or Brazil (P.1). Single-letter amino acid codes are used. Changes from the WT sequence are italicized and underlined. Primers for nPCR reactions are shown in Table S1 in the supplemental material. Abbreviations: nt, nucleotides; aa, amino acids.

RNA were detected in the sample from 11 February 2020 (22), 3 days before the first COVID-19 cases were confirmed in the area. The number of RNA replicates producing positive SARS-CoV-2 PCR results increased in March and April, likely reflecting an increase in viral concentration in sewage during the first wave of the pandemic and sharply decreased in May, in agreement with the reduction in COVID-19 cases as a result of nationwide lockdown measures introduced from 23 March. After a period of about 4 months with low-level virus transmission, COVID-19 confirmed cases increased again from early September and so did the number of RNA replicates from sewage samples showing positive SARS-CoV-2 PCR results.

**NGS analysis of nPCR products for the detection of SARS-CoV-2 B.1.1.7, B1.351, and P.1 lineages.** Twelve additional RNA replicates from each sewage concentrate were used to generate nPCR products with primer combinations C and D (see Table S1), mapping in the spike protein gene. All amplicons from positive nPCR reactions were analyzed by NGS with an aim to detect and quantify nucleotide sequence variations specifically found in B.1.1.7, B1.351, and P.1 lineages (Table 1). Nucleotide sequences characteristic of lineage B.1.1.7 were detected in sewage samples from 10 November 2020 onward. Mean sequence frequency values at key nucleotide positions, calculated from all sequenced replicates for each monthly sewage sample, are shown in Fig. 3. The results show that mutations discriminating the B.1.1.7 lineage from ancestral isolates, which include relevant mutations coding for amino acid changes and amino acid deletions (Table 1), clearly increased in frequency between November 2020 and January 2021 from 6.8 to 8.9% on 10 November to 44.0 to 46.4% on 8 December 2020, 93.6 to 94.1% on 12 January 2021, and 95.2 to 98.8% on 26 January 2021. It is worth noting that all four mutations analyzed showed very similar frequencies between them at any given time, reinforcing the conclusion that they all correspond to an increment of B.1.1.7 lineage through time. An example of NGS results showing an increasing proportion of a nucleotide 21765-21770 deletion (deletion of amino acids HV69-70) in viral RNAs from sequential sewage samples is shown in Fig. S1. It is important to emphasize that when analyzing surveillance data in real time, clinical surveillance has a potential limitation with respect to environmental surveillance since only a fraction of clinical samples is subjected to whole-genome sequence analysis. If 5% or fewer positive clinical samples are sequenced and the prevalence of B.1.1.7 or any other variant of interest in the population is 5%, at least 400 samples would need to be sequenced to find one such virus by chance. The time lapse between infection and symptoms developing, as well as samples taken and being tested, would add further delays compared to the expected immediacy of results from sewage surveillance.

Estimates of B.1.1.7 prevalence based on NGS data were in very good agreement with those from whole-genome sequencing and SGTF data from clinical samples from London and England (1, 2); according to the GISAID sequence database (10), the frequency rates of B.1.1.7 sequences from England during the week before and the week after sewage sample collection were 3.4 and 7.3% for 10 November, 25.2 and 42.7% for 8 December, 81.1 and 83.4% for 12 January, and 89.6 and 94.3% for 26 January, which is very much in line with what was observed in sewage (Table 2). No significant



**FIG 3** Mean frequency values of nucleotide mutations in the spike gene distinctive of SARS-CoV-2 B.1.1.7 lineage identified in nPCR C and nPCR D products synthesized from sewage concentrates and measured by NGS analysis. (A) Deletion of nucleotides 21765 to 21770 (deletion of amino acids HV69-70); (B) deletion of nucleotides 21991 to 21993 (deletion of amino acid Y144); (C) mutation A23063T (amino acid change N501Y); (D) mutation C23271A (amino acid change A570D). Error bars indicate standard errors of the mean. Mean numeric values are shown in Table 2. Statistical values associated with these data are shown in Table S2.

differences were found between B.1.1.7 frequency estimates obtained from sewage using data from any of the four mutations analyzed for any of the five sampling dates ( $t$  test,  $P > 0.05$ ) (see Table S3). A good correlation between B.1.1.7 frequency estimates from mutation analysis of viral RNAs from sewage and clinical data was observed throughout the period of study, with B.1.1.7 frequency increasing from  $<0.1\%$  in October to  $>95\%$  in late January and  $r^2$  values ranging between 0.968 and 0.997 (Table 2). There was better agreement between sewage and clinical data when using data from clinical samples corresponding to the week after sewage sample collection than to the week before, with sewage and clinical data superimposing almost perfectly when represented in the same graph (Fig. 4) and small differences between B.1.1.7 frequency estimates from sewage and clinical data representing only between 1.04 and 17.30% of the actual sewage frequency values versus between 5.92 and 61.46% when using clinical data from the week before (Table 2). In any case, this interpretation has to be taken with caution because sewage data corresponding to London, with a catchment area of approximately  $4.0 \times 10^6$  people, were compared to sequencing data from the whole of England, since no sequence data specific for London are publicly

**TABLE 2** Frequency of SARS-CoV-2 B.1.1.7 lineage in sewage (ES) and clinical samples<sup>a</sup>

			% B.1.1.7 mutations [SEM] (dif) <sup>b</sup>			
Sample		Date (day.mo.yr)	21765–21770 deletion	21991–21993 deletion	A23063T	C23271A
Oct	ES	13.10.20	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]	0.00 [0.00]
	Clinical.1	06–12.10.20	0.08 (0.08)	0.08 (0.08)	0.08 (0.08)	0.08 (0.08)
	Clinical.2	13–19.10.20	0.06 (0.06)	0.06 (0.06)	0.06 (0.06)	0.06 (0.06)
Nov	ES	10.11.20	8.53 [5.01]	8.88 [5.10]	6.88 [3.70]	6.84 [3.69]
	Clinical.1	03–09.11.20	3.42 (5.11)	3.42 (5.46)	3.42 (3.46)	3.42 (3.42)
	Clinical.2	10–16.11.20	7.34 (1.19)	7.34 (1.54)	7.34 (0.47)	7.34 (0.50)
Dec	ES	8.12.20	43.99 [10.77]	44.34 [10.81]	46.36 [8.86]	46.28 [8.85]
	Clinical.1	01–07.12.20	25.19 (18.80)	25.19 (19.15)	25.19 (21.17)	25.19 (21.09)
	Clinical.2	08–14.12.20	42.67 (1.32)	42.67 (1.67)	42.67 (3.69)	42.67 (3.61)
Jan.1	ES	12.01.21	94.13 [2.59]	93.9 [2.58]	93.86 [0.89]	93.6 [0.89]
	Clinical.1	05–11.01.21	81.08 (13.05)	81.08 (12.82)	81.08 (12.78)	81.08 (12.52)
	Clinical.2	12–18.01.21	83.44 (10.69)	83.44 (10.46)	83.44 (10.42)	83.44 (10.16)
Jan.2	ES	26.01.21	98.83 [0.45]	98.78 [0.44]	95.3 [1.71]	95.25 [1.73]
	Clinical.1	19–25.01.21	89.61 (9.22)	89.61 (9.17)	89.61 (5.69)	89.61 (5.64)
	Clinical.2	26.01.21 to 01.02.21	94.26 (4.57)	94.26 (4.52)	94.26 (1.04)	94.26 (0.99)
r <sup>2c</sup>	ES vs Clinical.1		0.980	0.979	0.968	0.968
	ES vs Clinical.2		0.996	0.997	0.994	0.994

<sup>a</sup>Data were obtained from the GISAID sequence database (10), consulted on 13 February 2021.

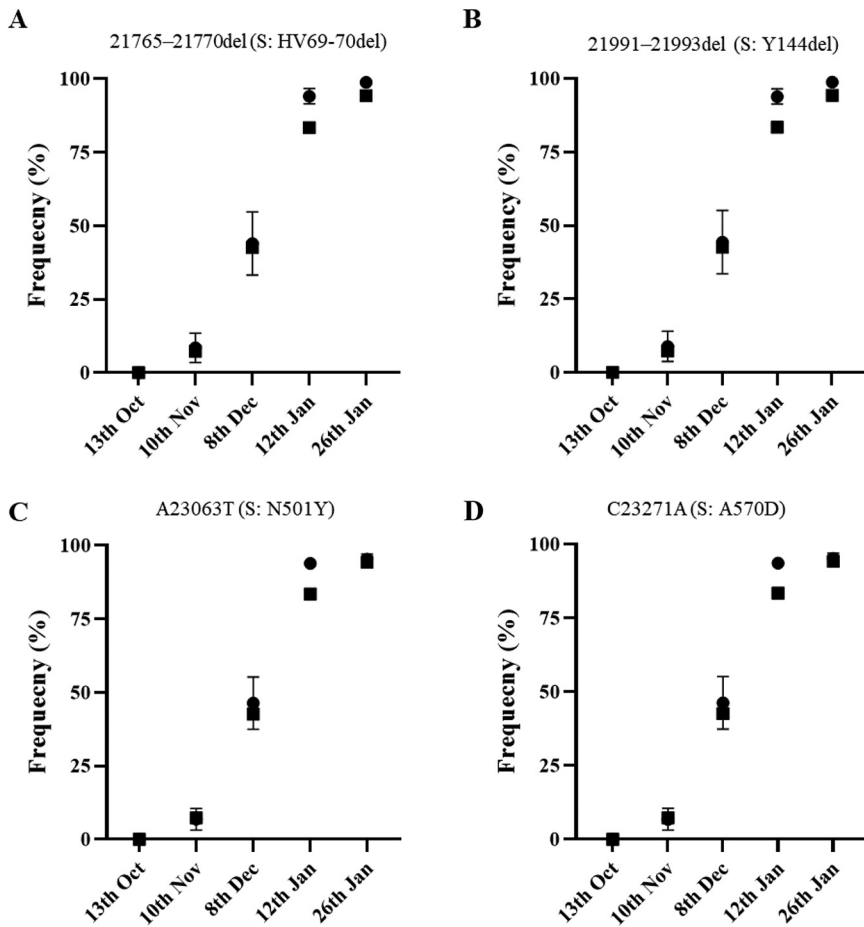
<sup>b</sup>Mean frequency values of B.1.1.7 mutations with associated standard errors of the mean (indicated in brackets) are presented. “dif” indicates the difference between B.1.1.7 mutation frequencies determined in sewage versus the clinical data in absolute number. dif values are given in parentheses.

<sup>c</sup>r<sup>2</sup> values were obtained from linear regression analysis.

available. Because different regions showed different timelines of B.1.1.7 introduction and growth, the comparison presented above might not be totally accurate. However, results from recent reports show that whole-genome sequencing and SGTF data from clinical samples from London resemble those for the whole of England very closely (1, 2).

In contrast, none of the nucleotide substitutions unique to B.1.351 and P.1 lineages (Table 1) were detected in any of the replicate nPCR products from any of the sewage samples. Additional nucleotide sequence variations were found at different sites in amplicons from sewage samples (Table 3), but none followed the same pattern as that observed with B.1.1.7 mutations, which were repeatedly detected in replicate viral RNA sequences and consistently found in sequential samples in increasing proportions. Interestingly, additional sequence variations most frequently found in viral RNA replicates from the same or different sampling dates mapped at locations adjacent to mutations characteristic of variant B.1.1.7 in the spike protein three-dimensional (3D) structure (Fig. 5): mutation S98F is located next to amino acid 69, which is deleted in variant B.1.1.7; substitution D574Y maps next to amino acid 570, which changes from alanine (A) to aspartic acid (D) in B.1.1.7 viruses; and mutations T478A, F490S, and S494P locate in the RBD, near amino acid 501, which changes from asparagine (N) to tyrosine (Y) in B.1.1.7. Most of these mutations have been found in clinical isolates from England, some of them in relatively large numbers, such as S98F and S494P (Table 3).

**Sanger sequence analysis of nPCR products.** Selected PCR products from nPCR reactions A and B were analyzed by Sanger sequencing to confirm the SARS-CoV-2 genome sequence. nPCR product B contains codon GAT at nucleotide position 28280, which changes to CTA in B.1.1.7 lineage viruses, resulting in the amino acid substitution aspartic acid to leucine at residue 3 of the nucleocapsid (N) protein. Hence, nPCR product B can also be used to identify and quantify B.1.1.7 sequences in sewage samples. The Sanger sequence analysis revealed that B.1.1.7 codon CTA was not present in replicate viral sequences from sewage concentrates from September (0/6) and October 2020 (0/4) but was detected in replicate samples from November 2020 (1/4), December 2020 (2/3), and January 2021 (2/2) in agreement with the results of NGS analysis of nPCR products C and D described above.



**FIG 4** Frequency rates of B.1.1.7 whole-genome sequences among clinical samples from England collected during the week immediately after sewage sample collection are shown as black squares superimposed to mean frequency values of B.1.1.7 mutations found in sewage samples measured by NGS (black circles with error bars indicating standard error of the mean, as shown in Fig. 3). Values in this figure represent data in Table 2. Statistical values associated with these data are shown in Table 2 and Table S3. Panels A to D are as defined for Fig. 3.

**DISCUSSION**

We were able to detect SARS-CoV-2 RNA in sewage samples from London throughout the COVID-19 pandemic, with changes in viral RNA levels shown to be in good agreement with those of confirmed COVID-19 cases reflecting national restrictions (Fig. 2). The second lockdown period lasting from 5 November to 2 December appear to only have had a limited effect in reducing virus transmission which, together with the emergence of the apparently highly transmissible B.1.1.7 variant, may explain the high levels of viral RNA found in sewage and the high number of SARS-CoV-2 infections reported in England in December 2020 and January 2021. In England 1 in 50 people (1 in 30 in London) were estimated to have had SARS-CoV-2 infection in the week ending 2 January according to an infection survey by the Office for National Statistics (29). The third nationwide lockdown initiated on 4 January 2021 is having a clear effect at reducing virus transmission, as judged by the observed drop in COVID-19 confirmed cases, but as of 26 January 2021, SARS-CoV-2 RNA levels remained high in sewage (Fig. 2), reflecting the extent of virus transmission occurring during the second phase of the second wave of the COVID-19 epidemic. Although detailed quantification of SARS-CoV-2 RNA in sewage and its correlation with COVID-19 infection prevalence was outside the scope of the present study and would require more frequent and widespread sampling, our results suggest that high viral RNA concentration levels could be



**TABLE 3** Additional missense nucleotide variations identified in viral RNAs from sewage samples

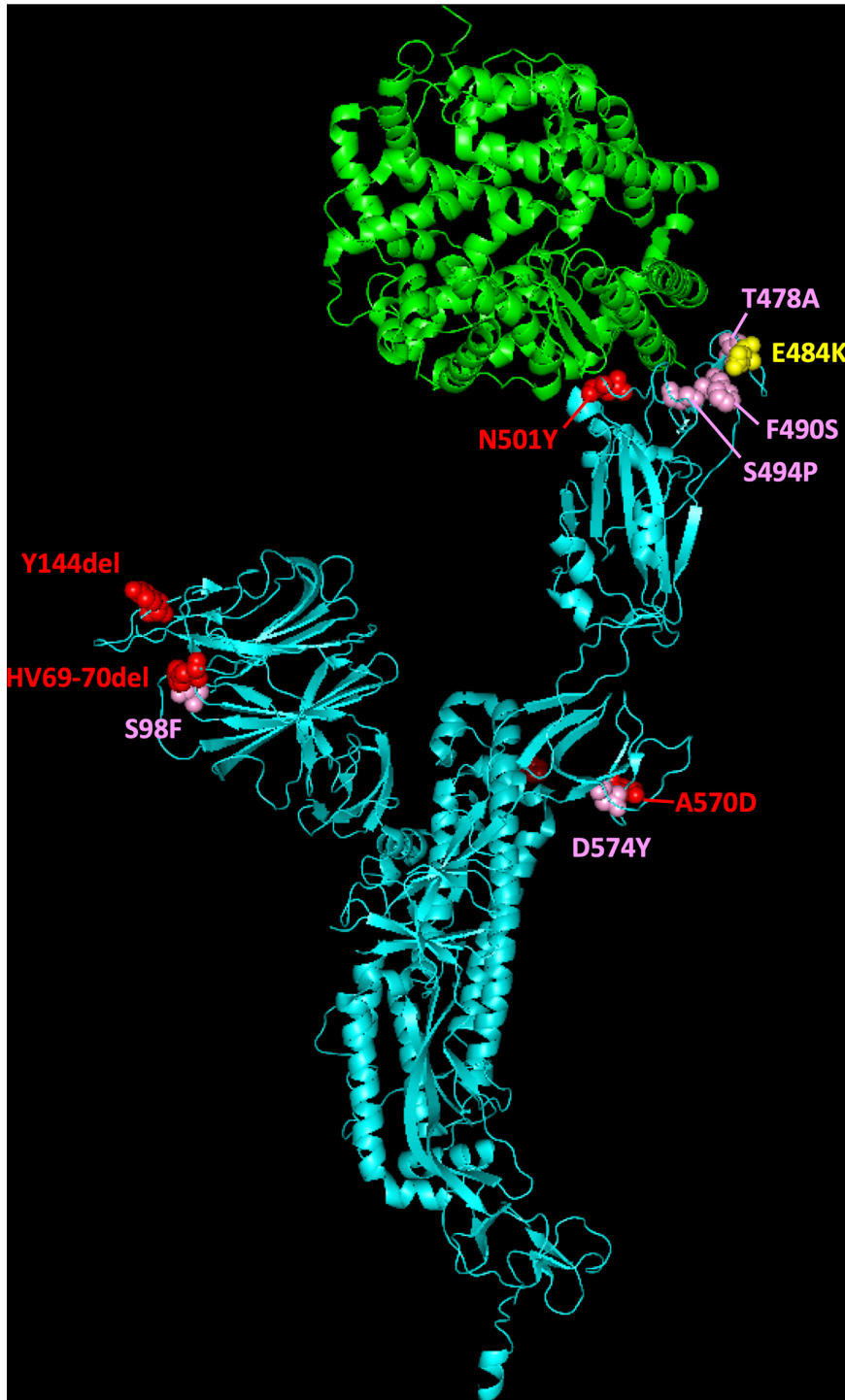
Sample	nt change	aa change	SNP (%) <sup>a</sup>	Clinical sequences <sup>b</sup>		
				Total UK (no.)	Last 28 days England (no.)	Date of first detection
13-Oct-20	A22994G	<u>T478A</u>	7.8, 7.6			
	A23201G	T547A	7.8	1	0	23-10-20
	G23282T	<u>D574Y</u>	99.4, 44.4	400	5	30-03-20
10-Nov-20	T21710C	S50P	8.4			
	T21765C	I68T	12.1	19	0	10-04-20
	C21846T	T95I	7.8	456	141	22-03-20
	C21855T	<u>S98F</u>	70.1, 14.5	1,672	281	22-03-20
	C21867A	R102K	6.5			
	C22995G	T478R	18.3	22	0	19-10-20
	T23105C	F515L	6.5	1	0	01-01-21
	G23282T	<u>D574Y</u>	33.1, 14.2	400	5	30-03-20
8-Dec-20	C21727T	S56F	10.0			
	A21894G	D111G	11.8			
	T21990C	V143A	5.1			
	T23031C	<u>F490S</u>	7.3	100	50	05-04-20
	T23075C	Y505H	8.2	9	4	02-11-20
12-Jan-21	A21683G	K41E	18.2	2	0	27-11-21
	T21723C	L54S	5.1	1	0	29-01-21
	G21761T	A67S	99.6	149	28	19-04-20
	A21779G	T73A	16.3	4	2	16-10-20
	C21855T	<u>S98F</u>	6.0, 48.0	1,672	281	22-03-20
	T21861C	I100T	31.9	15	0	27-03-20
26-Jan-21	T21831C	V90A	13.2			
	C21855T	<u>S98F</u>	6.3, 5.9, 16.4	1,672	281	22-03-20
	A21975G	D138G	13.2	25	18	13-10-20
	T23031C	<u>F490S</u>	7.0	100	50	05-04-20
	T23042C	<u>S494P</u>	5.1, 5.7, 6.8	740	162	20-03-20
	C23202A	T547I	7.8	87	20	07-04-20

<sup>a</sup>The percentages of mutations found in 12 individual replicate amplicon sequences per sewage sample are presented. Results are shown for each replicate in which the mutation was detected.

<sup>b</sup>Numbers and dates for clinical isolates showing the mutation are presented. The results were obtained using the COVID-19 Genomics (COG-UK) Consortium/Mutation Explorer from sequencing data generated on 28 February 2021 (<http://sars2.cvr.gla.ac.uk/cog-uk/>). Variations found in more than one replicate from the same or different collection dates are underlined.

detected preceding peaks in COVID-19 confirmed cases, showing the potential predictive value of ES, as has been shown in other studies (21).

NGS analysis of DNA amplicons directly synthesized from SARS-CoV-2 RNA present in sewage concentrates from London allowed us to readily identify viruses belonging to B.1.1.7 lineage before widespread transmission of this variant was obvious. We were able to detect B.1.1.7 variant at a frequency of 6.8 to 8.9% in a sample collected on 10 November 2020, a few weeks before it was first noticed through clinical surveillance. The B.1.1.7 variant first became apparent in late November 2020 when PHE was investigating why infection rates in Kent remained high despite national restrictions. A concerning cluster of virus isolates showing an unusual constellation of mutations was then found that could explain the persistent high infection rates (3, 15). Retrospective analysis showed that the two earliest genomes belonging to B.1.1.7 lineage were collected on 20 September and 21 September 2020 in Kent and London, respectively. B.1.1.7 is believed to have circulated at very low levels in the population until mid-November; it then rapidly spread across the country. The frequency rates of B.1.1.7 sequences observed in sewage between November 2020 and January 2021 were very similar to those of COVID-19 cases associated with infection by B.1.1.7 viruses at around the time of sewage sample collection (Table 2). B.1.1.7 was not detected in the sewage sample from 13 October 2020, consistent with the fact that the frequency of B.1.1.7 viruses identified in clinical samples collected in the weeks before and after 13



**FIG 5** Molecular cartoon diagram of the 3D structure of SARS-CoV-2 spike protein monomer in the open form (cyan) in complex with receptor ACE-2 (green). The image was generated using PyMOL Molecular Graphics System version 1.7.0.3 software (Schrödinger, LLC) using cryo-EM data (Protein Data Bank accession number 7DF4 [36]). The location of selected amino acid substitutions found in SARS-CoV-2 B.1.1.7 variant are shown in red; mutations found at low frequency in viral RNAs from sewage are indicated in pink, and the position of mutation E484K present in B.1.351 and P.1 VOCs is shown in yellow.

October 2020 was <0.08% in the whole of England. The proportion of B.1.1.7 sequences found in sewage increased rapidly from early November, reaching >90% 8 weeks after the first detection, in agreement with increasing infections associated with B.1.1.7. No evidence of the presence of variants B.1.351 and P.1 was found in any of

the sewage samples, which was expected since only a small number of B.1.351 isolates had been found in England at that time, not necessarily in London, and no P.1 viruses had been detected in the UK yet. The G23012A nucleotide substitution present in both B.1.351 and P.1 lineages and responsible for spike amino acid change E484K, thought to play a critical role at changing virus antigenicity and reducing human immune response to the virus and vaccine efficacy (13, 16–20), was not detected in viral genomes from sewage. This mutation has recently been found in a small number of B.1.1.7 isolates in England but mostly in locations away from London (15). We expect that the methods shown here can help to rapidly identify such mutants, as well as B.1.351 and P.1 lineages, in the same way that we have identified B.1.1.7 viruses (Table 1). We did find additional mutations in low frequency at amino acid sites of possible biological relevance located close to mutations characteristic of variant B.1.1.7 in the spike protein 3D structure; most of these are known to be present in clinical isolates from England. Mutations T478A, F490S, and S494P, in particular, are located very close to amino acid 484, known to have an effect on virus antigenicity, as discussed above. Similarly, mutations at these three sites have been shown to be associated with weaker neutralization to the virus by sera from convalescent people who have been infected with SARS-CoV-2 and/or monoclonal antibodies that could be used to patients with COVID-19 (30–34). It is important to monitor the appearance and to quantify the prevalence of such additional mutations since they might represent a challenge for vaccine efficacy. As shown in this study, ES appears to be a suitable system for such purpose. An early assessment of the biological effect of mutations of concern found to be prevalent in successive samples would be critical.

These methods can also be used to sequence SARS-CoV-2 in clinical samples using traditional Sanger sequencing for the rapid identification of VOCs in settings where whole-genome sequencing capabilities are limited. New PCRs can be quickly designed to track down new VOCs that might arise anywhere and detect importations. Furthermore, systematic sequencing of the spike gene of viral RNAs present in sewage samples collected regularly would help in proactively identifying potential new VOCs based on mutations and trends encountered. Eventually, if vaccine strains need to be regularly updated, as is current practice with flu vaccines, ES could be a useful resource to help vaccine strain selection based on prevalent SARS-CoV-2 variants circulating around the world.

As shown here, ES has a potential advantage for the early detection of VOCs versus clinical surveillance because it provides an immediate snapshot of virus transmission events representing thousands/millions of people, including asymptomatic infections. The methods shown here can provide sequencing results within 1 week of sewage samples arriving in the laboratory. According to the GISAID sequence database (10), only 33 B.1.1.7 sequences had been identified on 16 November 2020 in England, 1 week after the collection date of our first positive sewage sample. Even in the context of the phenomenal SARS-CoV-2 sequencing program established in England, clinical surveillance requires some time to build up enough data representative of current transmission patterns that is comparable to that provided by ES. In addition, since most clinical testing is based on samples from symptomatic infections, a delay of several days is expected between initial infection and viral sequence data availability due to the time required for symptoms to develop, samples to be collected, and PCR/sequencing analyses to be completed. Clinical surveillance is highly dependable on rapid and widespread PCR testing, as well as an extensive whole-genome sequencing program. This is currently only happening in a few countries so, in most locations, ES could be a major asset for monitoring SARS-CoV-2 virus circulation and identifying VOCs. After preparation of this manuscript, we recently detected a clear decline in the proportion of B.1.1.7 lineage in a sewage sample from April 2021 for the first time since it became dominant in late January 2020 and a concurrent increase in the presence of non-B.1.1.7 VOCs such as B.1.351 and B.1.525 (further research is in progress). On the other hand, identification of locally circulating VOCs could be delayed using ES at sites that cover very large populations, and therefore community-based ES targeting smaller

populations would increase the sensitivity for detection of such VOCs, e.g., current efforts to track down transmission of B.1.1.7 viruses harboring mutation E484K in England could be helped by community-based ES in areas where isolation of such viruses are known to have occurred.

## MATERIALS AND METHODS

**Wastewater sample collection and processing.** One-liter inlet wastewater composite samples were collected during a 24-h period at a sewage plant in London with a catchment area of approximately  $4.0 \times 10^6$  people. Monthly samples collected every second Tuesday of the month were collected between 14 January 2020 and 12 January 2021, with an additional sample collected on 26 January 2021. Each sample was processed using a filtration-centrifugation method described before (22). Briefly, following removal of solids by centrifugation at  $3,000 \times g$ , raw sewage was filtered through a  $0.45\text{-}\mu\text{m}$  filter (Nalgene 500-ml Rapid-Flow) and concentrated down from 70 ml to 200 to  $400\ \mu\text{l}$  using Merck Millipore Centricon Plus-70 centrifugal filter units with a 10-kDa molecular weight cutoff (Merck) according to the manufacturer's instructions. Samples from January to May 2020 reported previously (22) were reanalyzed in this study.

**nPCR amplifications.** Viral RNA was purified from sewage concentrates using the High-Pure viral RNA kit (Roche). RT-PCR fragments were amplified from purified viral RNAs using an nPCR system described before consisting of a one-step RT-PCR, followed by a second PCR using the first PCR product as the template (22). The genome location and nucleotide sequences of primer sets used for the nPCR reactions are shown in Table S1. Primer sets A and B targeting RdRP and ORF8b gene regions were used for diagnostic nPCRs. Primer sets C and D targeting spike gene regions were used for the detection of B.1.1.7, B.1.351, and P.1 lineages. As shown in Table 1, sequence analysis of nPCR C and D products can discriminate between B.1.1.7, B.1.351, and P.1 lineages and between them and ancestral isolates. Primers were tested using serial dilutions of purified RNA from the National Institute for Biological Standards and Control (NIBSC)'s virus reagent 19/304 containing noninfectious synthetic SARS-CoV-2 RNA packaged within a lentiviral vector ([https://www.nibsc.org/products/brm\\_product\\_catalogue/detail\\_page.aspx?catid=19/304](https://www.nibsc.org/products/brm_product_catalogue/detail_page.aspx?catid=19/304)). Twelve RNA aliquots from each sewage sample were used for nPCR amplifications for each nPCR reaction. Good laboratory practices were observed in all assays to reduce the possibility of cross-contamination, using different laboratory locations for sample processing, preparation of reaction mixtures, template addition, and postprocessing analysis. RNA extraction and negative template controls were included in every assay. Selected purified DNA amplicon products were sequenced by the Sanger method using an ABI Prism 3130 genetic analyzer (Applied Biosystems).

**NGS analysis of nPCR products.** Sequencing libraries were constructed by A/T adapter ligation using the KAPA HyperPrep kit (Roche, Switzerland) and dual-indexed using IDT TruSeq DNA unique dual indexes (Illumina, USA) with five PCR cycles for library amplification. These libraries were pooled in equimolar concentrations according to manufacturer's instructions and sequenced with 250-bp paired-end reads on MiSeq v2 (500 cycles) kits (Illumina). Initial demultiplexing was performed on-board by the MiSeq Reporter software. FASTQ sequencing data were adapter and quality trimmed by Cutadapt v2.10 (35) for a minimum Phred score of Q30, minimal read length of 75 bp, and 0 ambiguous nucleotides.

**Generation of SARS-CoV-2 sequence contigs and identification of single nucleotide polymorphisms.** Further processing and analysis of NGS data were performed using Geneious 10.2.3 software as described before (22). Filtered reads were imported into Geneious 10.2.3, paired-end reads were combined and merged, and sequence contigs were built by reference-guided assembly. Reads were mapped to references with a minimum 50-base overlap, minimum overlap identity of 90%, maximum 10% mismatches per read, allowing up to 15% gaps, and index word length of 12. Single nucleotide polymorphisms (SNPs) were identified using Geneious 10.2.3 default settings. The original SARS-CoV-2 Wuhan-Hu-1 strain (GenBank accession number [MN908947](https://www.ncbi.nlm.nih.gov/nuccore/MN908947)) was used as a reference. Variants with coverage  $<500$ , average quality  $<30$ , variant frequency  $<5\%$ , and strand-bias  $P > 10^{-6}$  were excluded. The analysis of several amplicon replicates per sample is critical to reduce sampling effects and increase the accuracy of SNP determinations for viral genomes present in a given sewage sample. If we set up a conservative limit of 5% for any SNP found in any replicate amplicon sequence to be accepted as genuinely present in viral RNAs, and given that we sequenced 12 independent RNA replicates per sample and nPCR reaction, the limit of detection for detecting a sequence variant would be approximately 0.42%, which seems reasonably low for the early detection of the spread of SARS-CoV-2 variants of potential concern. If we consider sequencing 5% of all positive clinical samples as a minimum requirement, which is currently only achieved in a few countries, a 0.42% prevalence of a variant would mean finding 21 to 105 viruses among 10,000 to 50,000 new daily positive cases, respectively. However, for clinical surveillance to be representative and have comparable sensitivity to ES, widespread and consistent clinical sampling across the sewage catchment area would be needed.

**SARS-CoV-2 nucleotide sequence analysis and estimates of the prevalence of B.1.1.7 lineage among whole-genome sequences from clinical samples.** SARS-CoV-2 sequences obtained in this study were compared to those available in the Global Initiative on Sharing All Influenza Data (GISAID) SARS-CoV-2 sequence database (10). The COVID-19 Genomics (COG-UK) Consortium/Mutation Explorer with sequencing data generated by 28 February 2021 (<http://sars2.cvr.gla.ac.uk/cog-uk/>) was also used to examine the presence and proportion of selected viral amino acid variations in clinical samples. Geneious version 10.2.3 software (Biomatters) was used for these analyses. Estimates of the proportion

of viruses belonging to B.1.1.7 lineage at different time points in England were obtained using the GISAID SARS-CoV-2 sequence database (10) with results available on 1 March 2021.

**Statistical analysis.** Statistical analysis and graphical representation of variant sequence frequency data were conducted using GraphPad Prism version 8.1.2 software. Statistical significance was determined using the Holm-Sidak method, with  $\alpha = 0.05$ .

**Data availability.** FASTQ files used in this study are available from the NCBI Short Read Archive under project code [PRJNA712959](https://www.ncbi.nlm.nih.gov/short-read-archive/projects/PRJNA712959).

## SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

**FIG S1**, PDF file, 0.6 MB.

**TABLE S1**, PDF file, 0.05 MB.

**TABLE S2**, PDF file, 0.1 MB.

**TABLE S3**, PDF file, 0.1 MB.

## ACKNOWLEDGMENTS

We thank Mihaela Cirdei, Ola Miloszewska, and Julian Hand from Public Health England for arranging the collection and transport of raw sewage samples.

This study is based on independent research commissioned and funded by the NIHR Policy Research Program (NIBSC Regulatory Science Research Unit). The views expressed in the publication are those of the author(s) and not necessarily those of the NHS, the NIHR, the Department of Health, “arms” length bodies, or other government departments.

## REFERENCES

- Chand M, Hopkins S, Dabrera G, Achison C, Christina B, Barclay W, Ferguson N, Volz E, Loman N, Rambaut A, Barrett J. 2020. Investigation of novel SARS-CoV-2 variant: variant of concern 202012/01. Public Health England, London, United Kingdom. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/959438/Technical\\_Briefing\\_VOC\\_SH\\_NJL2\\_SH2.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959438/Technical_Briefing_VOC_SH_NJL2_SH2.pdf). Accessed 9 February 2021.
- Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, Hinsley WR, Laydon DJ, Dabrera G, O'Toole A, Amato R, Ragonnet-Cronin M, Harrison I, Jackson B, Ariani CV, Boyd O, Loman NJ, McCrone JT, Gonçalves S, Jorgensen D, Myers R, Hill V, Jackson DK, Gaythorpe K, Groves N, Sillitoe J, Kwiatkowski DP, Flaxman S, Ratmann O, Bhatt S, Hopkins S, Gandy A, Rambaut A, Ferguson NM, The COVID-19 Genomics UK (COG-UK) Consortium. 2021. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* 593:266–269. <https://doi.org/10.1038/s41586-021-03470-x>.
- Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, Connor T, Peacock T, Robertson DL, Volz E. 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>. Accessed 9 February 2021.
- Chan KK, Tan TJC, Narayanan KK, Procko E. 2021. An engineered decoy receptor for SARS-CoV-2 broadly binds protein S sequence variants. *Sci Adv* 17:eabf1738.
- Starr TN, Greaney AJ, Hilton SK, Ellis D, Crawford KHD, Dingens AS, Navarro MJ, Bowen JE, Tortorici MA, Walls AC, King NP, Veesele D, Bloom JD. 2020. Deep mutational scanning of SARS-CoV-2 receptor binding domain reveals constraints on folding and ACE2 binding. *Cell* 182:1295–1310. <https://doi.org/10.1016/j.cell.2020.08.012>.
- Gu H, Chen Q, Yang G, He L, Fan H, Deng YQ, Wang Y, Teng Y, Zhao Z, Cui Y, Li Y, Li XF, Li J, Zhang NN, Yang X, Chen S, Guo Y, Zhao G, Wang X, Luo DY, Wang H, Yang X, Li Y, Han G, He Y, Zhou X, Geng S, Sheng X, Jiang S, Sun S, Qin CF, Zhou Y. 2020. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* 369:1603–1607. <https://doi.org/10.1126/science.abc4730>.
- Hoffmann M, Kleine-Weber H, Pohlmann S. 2020. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Mol Cell* 78:779–784. <https://doi.org/10.1016/j.molcel.2020.04.022>.
- Johnson BA, Xie X, Bailey AL, Kalveram B, Lokugamage KG, Muruato A, Zou J, Zhang X, Juelich T, Smith JK, Zhang L, Bopp N, Schindewolf C, Vu M, Vanderheiden A, Winkler ES, Swetnam D, Plante JA, Aguilar P, Plante KS, Popov V, Lee B, Weaver SC, Suthar MS, Routh AL, Ren P, Ku Z, An Z, Debbink K, Diamond MS, Shi PY, Freiberg AN, Menachery VD. 2021. Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. *Nature* 591:293–299. <https://doi.org/10.1038/s41586-021-03237-4>.
- Kemp S, Harvey W, Lytras S, Carabelli A, Robertson D, Gupta R. 2021. Recurrent emergence and transmission of a SARS-CoV-2 spike deletion H69/V70. *bioRxiv* <https://www.biorxiv.org/content/10.1101/2020.12.14.422555v6>.
- 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>.
- Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, Doolabh D, Pillay S, San EJ, Msomi N, Mlisana K, von Gottberg A, Walaza S, Allam M, Ismail A, Mohale T, Glass AJ, Engelbrecht S, Van Zyl G, Preiser W, Petruccione F, Sigal A, Hardie D, Marais G, Hsiao NY, Korsman S, Davies MA, Tyers L, Mudau I, York D, Maslo C, Goedhals D, Abrahams S, Laguda-Akingba O, Alisoltani-Dehkordi A, Godzik A, Wibmer CK, Sewell BT, Lourenco J, Alcantara LCJ, Kosakovsky Pond SL, Weaver S, Martin D, Lessells RJ, Bhiman JN, 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>.
- Faria NR, Claro IM, Candido D, Franco LAM, Andrade PS, Coletti TM, Silva CAM, Sales FC, Manuli ER, Aguiar RS, Gaburo N, Camilo CC, Frajji NA, Crispim MAE, Carvalho MPSS, Rambaut A, Loman N, Pybus OG, Sabino EC. 2021. Genomic characterization of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. <https://virological.org/t/genomic-characterization-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586>. Accessed 9 February 2021.
- Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, Bloom JD. 2021. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* 29:463–476 e6. <https://doi.org/10.1016/j.chom.2021.02.003>.
- Cheng MH, Krieger JM, Kaynak B, Arditi M, Bahar I. 2021. Impact of South African 501.V2 variant on SARS-CoV-2 spike infectivity and neutralization: a structure-based computational assessment. *bioRxiv* <https://www.biorxiv.org/content/10.1101/2021.01.10.426143v1>.
- Public Health England. 2021. Investigation of novel SARS-CoV-2 variant of concern 202012/01: technical briefing 5. Public Health England, London, United Kingdom. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/959426/Variant\\_of\\_Concern\\_VOC\\_202012\\_01\\_Technical\\_Briefing\\_5.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959426/Variant_of_Concern_VOC_202012_01_Technical_Briefing_5.pdf). Accessed 9 February 2021.
- Wu K, Werner AP, Moliva JI, Koch M, Choi A, Stewart-Jones GBE, Bennett H, Boyoglu-Barnum S, Shi W, Graham BS, Carfi A, Corbett KS, Seder RA, Edwards DK. 2021. mRNA-1273 vaccine induces neutralizing antibodies

- against spike mutants from global SARS-CoV-2 variants. *bioRxiv* <https://www.biorxiv.org/content/10.1101/2021.01.25.427948v1>.
17. Wang Z, Schmidt F, Weisblum Y, Muecksch F, Barnes CO, Finklin S, Schaefer-Babajew D, Cipolla M, Gaebler C, Lieberman JA, Oliveira TY, Yang Z, Abernathy ME, Huey-Tubman KE, Hurley A, Turroja M, West KA, Gordon K, Millard KG, Ramos V, Da Silva J, Xu J, Colbert RA, Patel R, Dizon J, Unson-O'Brien C, Shimeliovich I, Gazumyan A, Caskey M, Bjorkman PJ, Casellas R, Hatzioannou T, Bieniasz PD, Nussenzweig MC. 2021. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *Nature* 592:616–622. <https://doi.org/10.1038/s41586-021-03324-6>.
  18. Ho D, Wang P, Liu L, Iketani S, Luo Y, Guo Y, Wang M, Yu J, Zhang B, Kwong P, Graham B, Mascola J, Chang J, Yin M, Sobieszczyk M, Kyratsous C, Shapiro L, Sheng Z, Nair M, Huang Y. 2021. Increased resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7 to antibody neutralization. *bioRxiv* <https://www.biorxiv.org/content/10.1101/2021.01.25.428137v2>.
  19. Xie X, Liu Y, Liu J, Zhang X, Zou J, Fontes-Garfias CR, Xia H, Swanson KA, Cutler M, Cooper D, Menachery VD, Weaver SC, Dormitzer PR, Shi PY. 2021. Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nat Med* 27:620–621. <https://doi.org/10.1038/s41591-021-01270-4>.
  20. Muik A, Wallisch A-K, Sanger B, Swanson KA, Muhl J, Chen W, Cai H, Maurus D, Sarkar R, Tureci ˆO, Dormitzer PR, ˆsahin U. 2021. Neutralization of SARS-CoV-2 lineage B.1.1.7 pseudovirus by BNT162b2 vaccine-elicited human sera. *Science* 371:1152–1153. <https://doi.org/10.1126/science.abg6105>.
  21. 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>.
  22. Martin J, Klapsa D, Wilton T, Zambon M, Bentley E, Bujaki E, Fritzsche M, Mate R, Majumdar M. 2020. Tracking SARS-CoV-2 in sewage: evidence of changes in virus variant predominance during COVID-19 pandemic. *Viruses* 12:1144. <https://doi.org/10.3390/v12101144>.
  23. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE, Bhattacharya T, Foley B, Hastie KM, Parker MD, Partridge DG, Evans CM, Freeman TM, de Silva TI, McDanal C, Perez LG, Tang H, Moon-Walker A, Whelan SP, LaBranche CC, Saphire EO, Montefiori DC, Angyal A, Brown RL, Carrilero L, Green LR, Groves DC, Johnson KJ, Keeley AJ, Lindsey BB, Parsons PJ, Raza M, Rowland-Jones S, Smith N, Tucker RM, Wang D, Wyles MD. 2020. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 182:812–827. <https://doi.org/10.1016/j.cell.2020.06.043>.
  24. Fontenele RS, Kraberger S, Hadfield J, Driver EM, Bowes D, Holland LA, Faleye TOC, Adhikari S, Kumar R, Inchausti R, Holmes WK, Deitrick S, Brown P, Duty D, Smith T, Bhatnagar A, Yeager RA, Holm RH, Hoogesteijn von Reitzenstein N, Wheeler E, Dixon K, Constantine T, Wilson MA, Lim ES, Jiang X, Halden RU, Scotch M, Varsani A. 2021. High-throughput sequencing of SARS-CoV-2 in wastewater provides insights into circulating variants. *medRxiv* <https://www.medrxiv.org/content/10.1101/2021.01.22.21250320v1>.
  25. Nemudryi A, Nemudraia A, Wiegand T, Surya K, Buyukyoruk M, Cicha C, Vanderwood KK, Wilkinson R, Wiedenheft B. 2020. Temporal detection and phylogenetic assessment of SARS-CoV-2 in municipal wastewater. *Cell Rep Med* 1:100098. <https://doi.org/10.1016/j.xcrm.2020.100098>.
  26. Crits-Christoph A, Kantor RS, Olm MR, Whitney ON, Al-Shayeb B, Lou YC, Flamholz A, Kennedy LC, Greenwald H, Hinkle A, Hetzel J, Spitzer S, Koble J, Tan A, Hyde F, Schroth G, Kuersten S, Banfield JF, Nelson KL. 2021. Genome sequencing of sewage detects regionally prevalent SARS-CoV-2 variants. *mBio* 12:e02703–20. <https://doi.org/10.1128/mBio.02703-20>.
  27. Jahn K, Dreifuss D, Topolsky I, Kull A, Ganesanandamoorthy P, Fernandez-Cassi X, Banziger C, Stachler E, Fuhrmann L, Jablonski KP, Chen C, Aquino C, Stadler T, Ort C, Kohn T, Julian TR, Beerenwinkel N. 2021. Detection of SARS-CoV-2 variants in Switzerland by genomic analysis of wastewater samples. *medRxiv* <https://www.medrxiv.org/content/10.1101/2021.01.08.21249379v1>.
  28. Izquierdo-Lara R, Elsinga G, Heijnen L, Oude Munnink BB, Schapendonk CME, Nieuwenhuijse D, Kon M, Lu L, Aarestrup FM, Lycett S, Medema G, Koopmans MPG, de Graaf M. 2020. Monitoring SARS-CoV-2 circulation and diversity through community wastewater sequencing. *medRxiv* <https://www.medrxiv.org/content/10.1101/2020.09.21.20198838v1>.
  29. Office for National Statistics. 2021. Coronavirus (COVID-19) Infection Survey, UK: 8 January 2021. Office for National Statistics, London, United Kingdom. <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/coronaviruscovid19infectionsurveysurvey/8january2021>. Accessed 9 February 2021.
  30. Baum A, Fulton BO, Wloga E, Copin R, Pascal KE, Russo V, Giordano S, Lanza K, Negron N, Ni M, Wei Y, Atwal GS, Murphy AJ, Stahl N, Yancopoulos GD, Kyratsous CA. 2020. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* 369:1014–1018. <https://doi.org/10.1126/science.abd0831>.
  31. Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, Zhao C, Zhang Q, Liu H, Nie L, Qin H, Wang M, Lu Q, Li X, Sun Q, Liu J, Zhang L, Li X, Huang W, Wang Y. 2020. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* 182:1284–1294 e9. <https://doi.org/10.1016/j.cell.2020.07.012>.
  32. Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JC, Muecksch F, Rutkowska M, Hoffmann HH, Michailidis E, Gaebler C, Agudelo M, Cho A, Wang Z, Gazumyan A, Cipolla M, Luchsinger L, Hillyer CD, Caskey M, Robbiani DF, Rice CM, Nussenzweig MC, Hatzioannou T, Bieniasz PD. 2020. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife* 9:e61312. <https://doi.org/10.7554/eLife.61312>.
  33. Rappazzo CG, Tse LV, Kaku CI, Wrapp D, Sakharkar M, Huang D, Deveau LM, Yockachonis TJ, Herbert AS, Battles MB, O'Brien CM, Brown ME, Geoghegan JC, Belk J, Peng L, Yang L, Scobey TD, Burton DR, Nemazee D, Dye JM, Voss JE, Gunn BM, McLellan JS, Baric RS, Gralinski LE, Walker LM. 2020. An engineered antibody with broad protective efficacy in murine models of SARS and COVID-19. *bioRxiv* <https://www.biorxiv.org/content/10.1101/2020.11.17.385500v1>.
  34. Liu Z, VanBlargan LA, Bloyet LM, Rothlauf PW, Chen RE, Stumpf S, Zhao H, Errico JM, Theel ES, Liebeskind MJ, Alford B, Buchser WJ, Ellebedy AH, Fremont DH, Diamond MS, Whelan SPJ. 2020. Landscape analysis of escape variants identifies SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *bioRxiv* <https://www.biorxiv.org/content/10.1101/2020.11.06.372037v2>.
  35. Didion JP, Martin M, Collins FS. 2017. Atropos: specific, sensitive, and speedy trimming of sequencing reads. *PeerJ* 5:e3720. <https://doi.org/10.7717/peerj.3720>.
  36. Xu C, Wang Y, Liu C, Zhang C, Han W, Hong X, Wang Y, Hong Q, Wang S, Zhao Q, Wang Y, Yang Y, Chen K, Zheng W, Kong L, Wang F, Zuo Q, Huang Z, Cong Y. 2021. Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM. *Sci Adv* 7:eabe5575. <https://doi.org/10.1126/sciadv.abe5575>.