RAPID SCREENING FOR SARS-CoV-2 VARIANTS OF CONCERN IN CLINICAL AND ENVIRONMENTAL SAMPLES USING NESTED RT-PCR ASSAYS TARGETING KEY MUTATIONS OF THE SPIKE PROTEIN

La Rosa G.^{1*}, Mancini P.¹, Bonanno Ferraro G.¹, Veneri C.¹, Iaconelli M.¹, Lucentini L.¹, Bonadonna L.¹, Brusaferrero S.², Brandtner D.³, Fasanella A.⁴, Pace L.⁴, Parisi A.⁴, Galante D.⁴, Suffredini E.⁵

¹ Department of Environment and Health, Italian National Institute of Health, Rome, Italy.

² Office of the President, Istituto Superiore di Sanità, Rome, Italy

³ Independent Researcher, Rome, Italy.

⁴ Experimental Zooprophyllactic Institute of the Apulia and Basilicata Regions, Foggia, Italy.

⁵ Department of Food Safety, Nutrition and Veterinary Public Health, Italian National Institute of Health, Rome, Italy.

***Corresponding author:**

Giuseppina La Rosa

Department of Environment and Health

Italian National Institute of Health, Rome, Italy

Phone: +390649902718

Email: giuseppina.larosa@iss.it

SUMMARY

New SARS-CoV-2 mutations are constantly emerging, raising concerns of increased transmissibility, virulence or escape from host immune response.

We describe a nested RT-PCR assay (~1500 bps) to detect multiple key spike protein mutations distinctive of the major known circulating SARS-CoV-2 variants, including the three Variants of Concern (VOCs) 20I/501Y.V1 (United Kingdom), 20H/501Y.V2 (South Africa), and 20J/501Y.V3 (Brazil), as well as the 20E.EU1 variant (Spain), the CAL.20C recently identified in California, and the mink-associated variant (GR, lineage B.1.1.298). Prior to application to field samples, the discriminatory potential of this PCR assay was explored using GISAID and Nextclade. To extend variant detection to challenging matrices such as sewage, where the amplification of long fragments is problematic, two short nested RT-PCR assays (~300 bps) were also designed, targeting portions of the region spanned by the long nested assay.

The three newly-designed assays were then tested on field samples, including 7 fully-sequenced viral isolates from swab samples and 34 urban wastewater samples, some of which collected in areas where circulation of VOCs had been reported.

The long assay successfully amplified all the previously characterized viral isolates, allowing the correct identification of variants 20I/501Y.V1 and 20E.EU1 present in the panel. The sequences obtained using the short assays were consistent with those obtained with the long assay. Mutations characteristic of VOCs (UK and Brazilian variant) and of other variant (Spanish) were detected in sewage samples. To our knowledge, this is the first evidence of the presence of sequences harboring key mutations of 20I/501Y.V1 and 20J/501Y.V3 in urban wastewaters, highlighting the potential contribution of wastewater surveillance to explore SARS-CoV-2 diversity.

The developed nested RT-PCR assays can be used as an initial rapid screening test to select clinical samples containing mutations of interest. This can speed up diagnosis and optimize resources since

it allows full genome sequencing to be done only on clinically relevant specimens. The assays can be also employed for a rapid and cost-effective detection of VOCs or other variants in sewage for the purposes of wastewater-based epidemiology. The approach proposed here can be used to better understand SARS-CoV-2 variant diversity, geographic distribution and impact worldwide.

Keywords: SARS-CoV-2, variant, mutation, VOC, PCR, sequencing

INTRODUCTION

On 31 December 2019, the Wuhan Municipal Health and Health Commission reported a cluster of pneumonia cases of unknown aetiology. Further investigations identified a novel coronavirus as the causative agent of the disease (China CDC, 2020). On 11 March 2020, following the spread of the outbreak to other parts of China and to a number of countries worldwide, the Director General of the WHO declared the COVID-19 outbreak a pandemic (WHO, 2020a). As of 23 February 2021, 111 million cases have been documented in 192 countries, with 2.4 million confirmed deaths (<https://coronavirus.jhu.edu/map.html>).

A growing number of SARS-CoV-2 variant sequences are being detected since the beginning of the pandemic, some of which considered of global concern. All of these variants have characteristic mutations, the most significant of which are located in the gene encoding the spike (S) protein, potentially affecting viral infectivity and antigenicity (Li et al., 2020; Pillay et al., 2020).

One mutation in the spike protein (D614G), in particular, emerged early in the epidemic and spread rapidly through Europe and North America (Korber et al., 2020; Volz et al., 2020).

68 In August/September 2020, a SARS-CoV-2 variant linked to infection among farmed minks,
69 subsequently transmitted to humans, was identified in Denmark (Hammer et al., 2020). In all, six
70 countries (Denmark, the Netherlands, Spain, Sweden, Italy and the United States) reported cases of
71 farmed mink infected with SARS-CoV-2 to the World Organisation for Animal Health (OIE)
72 (WHO, 2020b).

73 An additional novel variant - clade GV, lineage B.1.177 (GISAID) or 20E.EU1 (Nexstrain) -
74 emerged in early summer of 2020 in Spain, and subsequently spread to multiple locations in
75 Europe, possibly introduced into other countries by summer tourists (Hodcroft et al., 2020). The
76 introduction of the B.1.177 lineage into Ontario, Canada, by a traveller returning from Europe in
77 late September 2020 was also reported in a recent preprint (Guthrie et al., 2020).

78 On 14 December 2020, British authorities announced that a new SARS-CoV-2 variant had been
79 identified through viral genomic sequencing (WHO, 2020c). The variant was initially termed
80 SARS-CoV-2 VUI 202012/01 (Variant Under Investigation, year 2020, month 12, variant 01) and
81 then re-designated VOC (Variant Of Concern) 202012/01 on 18 December 2020 (Chand et al.,
82 2021). The new variant belongs to clade GR, lineage B.1.1.7 (GISAID), or clade 20I/501Y.V1
83 (Nextstrain) and was discovered following an unforeseen rise in COVID-19 cases in South East
84 England, with a more than three-fold increase in the period from 5 October to 13 December 2020
85 (WHO, 2020d). As of 19 January 2021, 23 EU/EEA countries, including Italy, have reported the
86 new COVID-19 variant 20I/501Y.V1. In the rest of the world, 37 additional countries reported it
87 (ECDC, 2021b). The emergence of the new SARS-CoV-2 variant, which could be 40-70% more
88 transmissible than previous strains circulating in the UK (WHO 2020b; ECDC, 2020), has
89 prompted some countries to close their borders with the UK. A review by the UK's New and
90 Emerging Respiratory Virus Threats Advisory Group (NERVTAG) advises that there is moderate
91 confidence that 20I/501Y.V1 has a substantial increase in transmissibility compared to other
92 variants (NERVTAG, 2020; Mahase 2020). However, there is not enough information at present to

determine whether this variant is associated with any change in clinical disease severity or increased mortality, though preliminary analyses in the United Kingdom, showed a risk ratio of death of 1.65 for B.1.1.7 cases compared to non-B.1.1.7 cases (PHE, 2021). As for antigenic response, to date, it is considered unlikely that 20I/501Y.V1 may hinder vaccine-induced immunity (Conti et al., 2021). On 18 December 2020, national authorities in South Africa announced the detection of a new variant of SARS-CoV-2, designated VOC-202012/02 (the second Variant Of Concern, December 2020). This variant, belonging to clade GH, lineage B.1.351 (GISAID) or 20H/501Y.V2 (Nextstrain), was reported to be largely replacing other coronavirus strains in South Africa since at least November 2020. As of 19 January 2021, 10 EU/EEA countries (Germany, France, Ireland, Austria, Belgium, Finland, the Netherlands, Norway and Sweden) and 13 countries/territories outside of the EU/EEA have documented the presence of this new SARS-CoV-2 variant (ECDC, 2021b). One case of the South African variant was detected in Italy, in a man who landed at Milan's Malpensa Airport from Africa at the end of January (https://www.ansa.it/lombardia/notizie/2021/02/03/a-varese-primo-caso-variante-sudafricana-in-italia_41b99448-367c-4000-ac88-9b3ae130d839.html). Preliminary results indicate that it may have increased transmissibility. Here too, as with the VOC 202012/01 variant, there is no evidence that it is associated with a higher severity of infection (ECDC, 2021b).

In January, another variant, clade GR lineage P.1 (GISAID) or 20J/501Y.V3 (Nextstrain) was identified in Japan in four travellers who had arrived from Brazil (NIID, 2021; Galloway et al., 2021). Whole genome sequences for all four isolates were deposited in GISAID EpiCoV on 10 January 2021. On 13 January 2021, researchers from Fiocruz Amazônia announced the identification of this variant in the capital of Amazonas, Manaus, where an upsurge in COVID-19 cases was registered in January of 2021 (Farias et al., 2021). To date, the P.1 variant has been identified either in Brazil, or in travellers from Brazil (mostly from the State of Amazonas), including one returning to Italy from Brazil (Maggi et al., 2021).

Lastly, on 17 January 2021, public health officials announced that the SARS-CoV-2 GH, lineage B.1.429 (GISAID) – also termed CAL.20C – has become increasingly common in multiple Californian counties and has been responsible for several large COVID-19 outbreaks in that state. All of these variants have characteristic mutations, the most significant of which are located in the gene encoding the spike (S) protein (Li et al., 2020; Pillay et al., 2020). The so-called ‘UK variant’ 20I/501Y.V1 is characterized by 23 mutations: 14 nonsynonymous mutations, 3 deletions and 6 synonymous mutations affecting ORF1ab, S, ORF8, and N (Chand et al, 2021). Mutations in the gene that encodes for the S protein, which is associated with viral entry into cells, has attracted the most attention. Ten mutations are documented in the S protein of the UK variant. Two of these, the N501Y (an asparagine to tyrosine amino acid substitution in the receptor binding domain) and the HV69-70del (a 6-base deletion), had already been circulating globally before merging into this new variant. The most relevant changes in the 20I/501Y.V1 variant are likely to be the N501Y mutation and the Y144del - a deletion potentially contributing to spike surface variation. The N501Y mutation is in the receptor binding domain (RDB), important to ACE2 binding and antibody recognition. It is likely to enhance the transmissibility of the virus by improving the S protein's affinity for the host's ACE2 receptor. *In vitro* (Starr et al., 2020) and animal experiments (Gu et al., 2020) seem to confirm this hypothesis. The 20H/501Y.V2 variant, first detected in South Africa, displays the same N501Y mutation (as well as the D614G) as the UK variant, but this mutation is thought to have arisen independently in the two lineages (WHO 2020b). It also carries additional mutations (the most common are L18F, K417N, E484K and K417N) in the S protein, which are not usually present in the UK strain, though mutation E484K has been recently detected in 21 B.1.1.7 sequences in UK (PHE, 2021). Furthermore, 501Y.V2 does not have the HV69-70del characterizing the UK variant, but presents in about 85% of sequences another deletion in the S region, that of amino acids 242 to 244 (LAL242-244del).

The new variant from Brazil, 20J/501Y.V3, harbors 12 mutations in the S protein, some of which it shares with either the ‘UK variant’ (N501Y and D614G) or the ‘South African variant’ (L18F, E484K, N501Y, D614G).

The Spanish variant 20E.EU1 is characterized by substitution A222V in the A domain of the S protein, a domain which is not considered to contribute directly to receptor binding or membrane fusion for SARS-CoV-2 (Hodcroft et al., 2020).

The rapid emergence of variants worldwide highlights the importance of genetic surveillance of the SARS-CoV-2 pandemic. The risk of introduction and community spread of variants of concern in the EU/EEA countries is considered very high (ECDC, 2021b). WHO advises all countries to increase the routine sequencing of SARS-CoV-2 viruses where possible, and to share sequence data internationally through open-source platforms (WHO 2021). ECDC also recommends urgent analysis and sequencing of virus isolates to identify cases of the new variants in a timely manner (ECDC, 2021a). Currently, whole genome sequencing (WGS) of human isolates has resulted in over 591,000 complete genomes available in GISAID as of 23 February, 2021.

A full genome approach (WGS), however, is time-consuming and costly, and requires specialized computational infrastructure and technical skills. Furthermore, WGS cannot be effectively performed in conditions of low virus titres, reduced sequence integrity and simultaneous occurrence of more than one genome.

We developed a long PCR nested RT-PCR assay (~1500 bp), followed by conventional Sanger sequencing, to quickly screen for the different circulating variants or detect the introduction of possible novel variants. The assay amplifies a key region of the S gene, the region containing the major mutations of the UK, South African and Brazilian variants, as well as crucial mutations of other variants of clinical interest, like the 20E.EU1 that emerged in Spain, the mink-associated variant, and the CAL.20C variant identified in California (see Table 1). Preliminary studies on sequences available in GISAID and Nextclade were performed before application on field samples.

The first, aimed at verifying whether specific mutation combinations are sufficiently informative to screen for specific SARS-CoV-2 variants, and the second, a comparative study of full-length sequences and their corresponding ~1500 bp fragments to ascertain whether the different variants would be correctly assigned to their respective clades on the sole basis of the genome region amplified by the long nested PCR assay.

We subsequently enhanced our method to allow variant detection in highly challenging matrices such as wastewater, where the amplification of long RNA fragments is difficult, by designing two additional short nested RT-PCRs targeting portions of the region spanned by the long assay.

SARS-CoV-2 is excreted by both symptomatic and asymptomatic individuals that are infected with the novel coronavirus, and ends up in wastewater systems (Foladori et al., 2020). By analysing wastewater, it is therefore possible to obtain information on viruses circulating in the population served by a given Wastewater Treatment Plant (WTP). Previous studies have detected SARS-CoV-2 RNA in wastewater samples worldwide, highlighting environmental surveillance as a valuable tool potentially allowing us to detect peaks in virus circulation and estimate the proportion of individuals infected with SARS-CoV-2 in a given catchment area (Hart et al., 2020), thus contributing to inform public health interventions.

All the assays were successfully tested on a 20I/501Y.V1 strain detected in Italy at the end of December 2020 (see Supplementary material), and on a Wuhan strain (clade 19A). Finally, the three assays were tested on a panel of field samples of clinical (viral isolates from swabs) and environmental (sewage) origin.

The methods proposed could be used to perform targeted sequencing of clinical and environmental samples for screening and early detection of SARS-CoV-2 variants.

MATERIALS AND METHODS

Primer design and preliminary in silico studies

Primers were designed to capture key mutations of different variants. Such mutations (Table 1), potentially affecting viral infectivity and antigenicity, are located in the gene encoding the spike (S) protein. Primers (Table 2 and Figure 1) were designed using the Primer 3 software (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>).

Ultimately, three assays were designed with different aims:

a) A long nested RT-PCR (first cycle - PCR ID 979; nested assay - PCR ID 980, 1583-1592 bps) designed to screen for the presence of mutations characteristic of the three VOCs (20I/501Y.V1, 20H/501Y.V2, and 20J/501Y.V3), as well as other variants of interest such as 20E.EU1,CAL 20C, and mink variant, as shown in Table 1.

b) A short nested RT-PCR (319-327 bps) amplifying a portion of the region targeted by the long assay (first cycle - PCR ID 972; nested assay - PCR ID 973) designed to reveal the presence of the mutations HV69-70del and Y144del (UK variant), D80A (South African variant), and D138Y (Brazilian variant).

c) Another short nested RT-PCR (323 bps) amplifying a different portion of the region targeted by the long assay (first cycle - PCR ID 974; hemi-nested assay - PCR ID 975) designed to reveal the presence of the key mutation N501Y shared by different variants of concern. This amplicon can also reveal the presence of amino acid substitution A570D (combined with N501Y) of the 20I/501Y.V1 variant, and E484K (combined with the N501Y) of either the 20H/501Y.V2 or the 20J/501Y.V3 variant. It can also detect L452R of the Californian variant and Y453F of the mink variant.

Before application to field samples, a preliminary in silico study was performed to examine the discriminatory potential of the specific mutations covered by our long assay, for SARS-CoV-2

variant identification. This was done using GISAID (<https://www.gisaid.org/>) and Nextclade (<https://nextstrain.org/>). The GISAID study was performed on 20I/501Y.V1 (14 January 2021) and on 20H/501Y.V2 (17 January 2021) using complete genomes (see Supplementary Materials). Overall, considering the characteristic mutations of 20I/501Y.V1 present in the ~ 1500 bp fragment of the S gene chosen for the long assay (HV69-70del, Y144del, N501Y and A570D), the results showed that among the complete genomes belonging to lineage GR B.1.1.7 any combination of two mutations was present at a frequency ranging from 97.6% to 99.4%. Furthermore, cluster analysis attributed all of the sequences displaying at least two of the aforementioned mutations to the GR B.1.1.7 lineage. As for the 20H/501Y.V2 variant, characteristic mutations such as D80A, E484K, N501Y, LAL242-244del were present at a frequency ranging from 88.6 to 91.4%. Any combination of two of these mutations was present in clade GH lineage B.1.351 at a frequency exceeding 84.2%, and sequences displaying at least two of the mutations belonged to lineage B.1.351 in >99% of cases. One significant exception was the E484K/N501Y combination, which belonged to lineage GH B.1.351 in 96.8% of cases, due to its presence in the 20J/501Y.V3 (Brazilian) variant as well. Based on this preliminary analysis, it was concluded that, for these two variants, the detection of a combination of at least two mutations in the selected ~ 1500 bp fragment provided a strong indication of the presence of the variant in question.

To further explore the discriminatory potential of our newly-designed long nested PCR for the different SARS-CoV-2 variants, we compared the ~1500 bp fragments to their corresponding full-length sequences in terms of how they would be classified by Nextclade web tool. We therefore created two FASTA files: one containing 30 full-length genomes, comprising five GISAID strains for each of the following clades/variants: 19A (Wuhan, China), 20I/501Y.V1 (UK), 20H/501Y.V2 (South Africa), 20J/501Y.V3 (Brazil), 20E.EU1 (Spain), CAL.20C (California) as described in the Supplementary Materials; and the other, derived from the first, containing the corresponding ~1500 bp fragments amplified by our long nested PCR assay. Both files were then analyzed using the

Nextclade web tool (nextstrain.org) in order to determine whether our nested assay (PCR ID 980) would yield similar results to those given by a full sequence analysis. Nextclade web tool assigned all the variants correctly on the sole basis of the genome region amplified by the PCR ID 980. The only significant exception was the 20I/501Y.V1 variant, which was incorrectly assigned to clade 20B. This misclassification, however, was due to a limitation of the Nextclade web tool, which, for the purposes of phylogenetic placement and clade assignment, uses only mutations and no gaps (<https://github.com/nextstrain/nextclade/issues/311#issuecomment-765904855>), and was therefore unable to take into consideration the deletions HV69-70del and Y144del for a correct assignment of our partial sequences of the S gene. Similarly, the two trees built with Nextclade, one based on full-length genomes and the other based on ~1500 bp fragments, both correctly assigned all of the variants to the branches corresponding to clades 19A, 20H/501Y.V2, 20J/501Y.V3 and 20E.EU1. Only the ‘UK variant’ was placed at the base of the 20I/501Y.V1 branch (Figure 2).

Primer validation on field samples and assay specificity

The RNA of the 20I/501Y.V1 variant was obtained from a nasopharyngeal swab of an Apulian patient who had returned from the United Kingdom. The sample had previously been isolated on Vero E6 cells, fully sequenced (Experimental Zooprophyllactic Institute of Puglia and Basilicata), and the sequence submitted to GISAID under reference hCoV-19/Italy/APU-IZSPB-399PT/2020. RNA from the Wuhan strain (kindly provided by the Robert Koch-Institute under the RefBio project) was included as control.

Before being tested with the three nested PCR assays, both RNAs were standardized at a concentration ranging between 10^2 and 10^3 genome copies (g.c.)/ μ L using a previously published real-time RT-qPCR (La Rosa et al., 2021).

The European Virus Archive global (EVAg) Coronavirus RNA specificity panel (kindly provided by the Erasmus University Medical Center, Rotterdam, The Netherlands), which includes both

Alfa- and Beta-coronaviruses (HCoV-NL63, HCoV-229E, HCoV-OC43, MERS-CoV, SARS-CoV and SARS-CoV-2), was used to assess the specificity of our newly-designed primers.

Field samples

i) Clinical samples: Seven SARS-CoV-2 RNA samples had previously been extracted from viruses originating from nasopharyngeal swabs following isolation on Vero E6 cells, and had been characterized by WGS and deposited in GISAID as belonging to clades 20A, 20B, 20E.EU1 (two samples each), and 20I/501Y.V1 (one sample). Viral concentrations in these samples were calculated as previously described (La Rosa et al., 2021) and found to be in excess of 10^6 g.c./ μ L. The samples were subsequently diluted to an approximate concentration of 10^4 g.c./ μ L before testing by nested PCR.

ii) Environmental samples: a total of 34 sewage samples were included in the study. Twenty samples, collected in Wastewater Treatment Plants (WTPs) in Rome, Italy, between September and December 2020 (see Supplementary Materials), had been previously analyzed and found positive for SARS-CoV-2 following the procedure described in La Rosa et al. (2021). Briefly, concentration of sewage samples (250 ml) was performed by the two-phase polyethylene glycol-dextran separation method (WHO 2003) and nucleic acids were extracted using the NucliSens extraction system (bioMerieux, Marcy-l'Etoile, France). The eluted RNA (100 μ l) was aliquoted and stored at -80°C until molecular analysis. Viral concentrations in these samples ranged from 2.0 to 37.8 g.c./ μ L of RNA (corresponding to 1.6×10^3 to 3.0×10^4 g.c./L of raw sewage).

Eight sewage samples (4 composite and 4 grab) were collected between 21 and 25 January 2021 in 4 WTPs in the small town of Guardiareale in Abruzzo (central Italy), where a cluster linked to the UK variant had been reported, to investigate whether the S mutations associated with 20I/501Y.V1 were detectable in the local sewage.

Furthermore, six sewage samples were collected between 5 and 8 February 2021 in the city of Perugia, Umbria (central Italy), where the UK variant had been on the rise and the Brazilian variant had been identified as well (<https://www.regione.umbria.it/documents/18/25027633/Ordinanza+n.14+del+6+febbraio+2021.pdf>).

Nested RT-PCR conditions

Following standardization of the protocol, SuperScript IV One-Step RT-PCR System with Platinum SuperFi RT-PCR Master mix (Invitrogen, Carlsbad, CA, USA) was used for the RT-PCR. The amplification conditions for the RT-PCR were as follows: reverse transcription at 45 °C for 10 min, denaturation at 98 °C for 2 min, followed by 35 cycles of 98 °C for 10 sec, 58 °C for 10 sec, 72 °C for either 30 sec (PCR ID 972 and 974) or 1 min (PCR ID 979), and a final extension at 72 °C for 5 min. Each of the three PCR reactions was performed using 1 µL of each primer (10 µM; Table 1) and either 4 µL of RNA in a final volume of 25 µL (clinical samples) or 8 µL of RNA in a final volume of 50 µL (environmental samples). After the first round of PCR, nested PCR was performed using the Phusion Hot Start II DNA Polymerase with GC buffer (Thermo Fisher Scientific, Waltham, MA, USA), 2 µL of the first PCR product, 1 µL of each primer (10 µM) and the following conditions: initial denaturation at 98 °C for 30 sec, followed by 45 cycles of 98 °C for 10 sec, 62 °C for 30 sec, 72 °C for either 30 sec (PCR ID 973 and 975) or 1 min (PCR ID 980), and a final extension at 72 °C for 10 min. Standard precautions were taken to avoid laboratory contamination. PCR products were observed by gel electrophoresis (2% agarose gel, stained with GelRed, Biotium; Fremont, CA, USA). To confirm amplifications, all products were purified from the PCR reaction or from gel using either Montage PCRm96 Micro-well Filter Plate (Millipore, Burlington, MA, USA) or GRS PCR & Gel Band Purification Kit (GRISP, Porto, Portugal) and

were sequenced (see Supplementary Materials) on both strands (Bio-Fab Research, Rome, Italy and Eurofins Genomics, Ebersberg, Germany).

Sequences were submitted to GenBank under the accession numbers [to be assigned; submission IDs 4852387, 4852507, and 4852527].

Bioinformatics analysis

The public database GISAID (Shu and McCauley, 2017) was used for BLAST searches and for mutation analysis. The Nextclade web tool v.012.0 (Hadfield et al., 2018) was used to compare study sequences to SARS-CoV-2 reference sequences, assign them to clades, and determine their position within the SARS-CoV-2 phylogenetic tree. As of 21 January 2021, the database includes 11 major clades (19A, 19B, 20A, 20B, 20C, 20D, 20E(EU1), 20F, 20G, 20I/501Y.V1, 20H/501Y.V2, 20J/501Y-V3). Auspice v2.0, an open-source interactive tool, was used to visualize phylogenetic trees.

All SARS-CoV-2 mutations found in field samples were compared with the GISAID reference strain hCoV-19/Wuhan/WIV04/2019 using the CoVsurver mutation analysis tool (<https://www.gisaid.org/epiflu-applications/covsurver-mutations-app/>), which provides geographic and temporal distributions of SARS-CoV-2 mutations.

RESULTS

Primer validation and assay specificity

All three newly designed PCR assays successfully amplified the two strains (the UK variant and the Wuhan strain) used for primer assessment.

Amplicon sequencing of the long fragment of the UK variant revealed the presence of mutations HV69-70del, Y144del, N501Y and A570D characteristic of 20I/501Y.V1 (Table 1). Sequencing of the short amplicons obtained using the other two assays confirmed this result. In the Nextclade tree constructed using the ~1500 bps sequence, this isolate fell at the base of the branch corresponding to clade 20I/501Y.V1, as predicted by the preliminary study we conducted with the UK variant in Nextrain prior to primer design (Figure 2). As for the Wuhan strain, although amplification was achieved with all three assays, the long PCR yielded a low amount of product, probably due to RNA fragmentation, allowing only partial sequencing of the fragment. Identity with GenBank reference sequence NC_045512 was confirmed for amplicons provided by the two short assays. Specificity of the three designed assays was confirmed on the EVAg Coronavirus RNA panel. Amplicons of the expected size were obtained only for SARS-CoV-2 (sequencing confirmed 100% identity to reference sequence NC_045512) while no amplification was present for HCoV-NL63, HCoV-229E, HCoV-OC43, MERS-CoV, and SARS-CoV RNAs.

Clinical samples

All of the seven SARS-CoV-2 RNAs extracted from viruses cultured on Vero E6 cells and previously characterized by WGS were successfully amplified by the three newly developed assays. Sequencing results of amplicons generated by the long PCR are summarized in Table 3. The geographic and temporal occurrence of the mutations detected in the field samples, obtained using the GISAID CoVsurver: Mutation Analysis tool, is summarized in Supplementary Materials. Sequences obtained by the long PCR confirmed the presence of the two variants in swab samples (20E.EU1 in swab IDs 29 and 30, and 20I/501Y.V1 in swab ID 31). The UK variant, however, could only be assigned to its correct lineage using the GISAID BLAST tool, given the aforementioned limitation of the Nextclade tool. Samples free of mutations or carrying mutations not specific to any VOC in the analyzed region (swab IDs 25 to 28) could not be differentiated and

assigned. Indeed, they showed 100% nt identity with different clades/lineages in GISAID, and were all classified as 19A by Nextclade (Table 3).

All clinical samples also tested positive using the two short nested PCR assays. Amplicon sequencing of the short PCR fragments confirmed the sequences obtained using the long nested PCR in all samples.

Environmental samples

A total of 34 wastewater samples were tested with the three nested PCR assays. PCR and sequencing results are summarized in Table 4. The long nested RT-PCR did not amplify any of the 20 SARS-CoV-2-positive urban sewage samples collected from WTPs in Rome (IDs 3702 to 3815). Amplification, however, was achieved with PCR IDs 973 and 975 (8 and 10 samples, respectively, with an overlap of 5 samples). None of these samples displayed mutations indicative of VOCs.

With regard to the 14 sewage samples collected from locations experiencing outbreaks of SARS-CoV-2 variants, the long nested PCR successfully amplified two of the eight samples collected in Guardiagrele, Abruzzo (IDs 3863 and 3865) and one of the six samples collected in Perugia, Umbria (ID 3947). The first two samples showed mutations characteristic of the Spanish variant (A222V), while the sample ID 3947 carried the mutation characteristic of the Brazilian variant (D138Y, R190S, K417T, E484K, and N501Y). The short assays - PCR ID 973, and PCR ID 975 - successfully amplified 7 and 6 of the 14 sewage samples, respectively (with an overlap of 6 samples, including the three samples testing positive also by the long assay).

Sequencing of fragments obtained by PCR 973 showed no mutations in the PCR amplicons originating from Guardiagrele. On the contrary, this short assay did find the presence of mutation D138Y, characteristic of the Brazilian variant, in three of the six samples collected in Perugia: ID 3944, 3945, and 3947. Another sample from Perugia (ID 3949), showed the aa substitution E96G, while the remaining two amplifications did not display any mutation.

The second short assay, PCR 975, confirmed the results obtained by PCR 973 on the amplified samples from Guardiagrele (IDs 3862 and 3863), namely, the absence of mutations indicative of VOCs. In the four samples collected in Perugia and identified as carrying mutations, on the other hand, this short assay (PCR 975) showed the key mutation N501Y. This mutation was combined with E484K in three of them (IDs 3944, 3945, and 3947), a combination present in both the Brazilian and the South African variants (see Table 1). The results of PCR 973, where all three samples showed the aa substitution D138Y, and those of the long assay showing one of them (ID 3847) to carry all 5 aa substitutions typical of the Brazilian variant, pointed to the presence of this VOC. The fourth sample from Perugia (ID3949) carried mutation N501Y along with A570D, a key signature of the UK variant.

DISCUSSION

SARS-CoV-2 sequence diversity was initially thought to be low, but an increasing number of variant sequences have since been detected, indicating that new mutations are continually emerging. Major variants discovered thus far include the UK variant (clade GR, lineage B.1.1.7), the South African variant (GH, lineage B.1.351), the Brazilian variant (clade GR, lineage P1), the variant from California (GH, B.1.429), the Spanish variant, the mink variant, and others. All these SARS-CoV-2 variants carry characteristic mutations (deletions and/or missense mutations) in different parts of the genome, some in important regions in terms of potential clinical significance, such as the gene coding for the spike protein (the S gene). The spike protein appears to be critical for cellular entry, and allows the virus to attach to the host cell (Pillar et al., 2020), which is why this protein is also a target for vaccine development and antigen testing.

The analysis of SARS-CoV-2 virus genomes can complement and support strategies to reduce the burden of COVID-19 (WHO, 2021). Whole genome sequencing has proven to be a powerful tool capable of helping us understand outbreak transmission dynamics and spill-over events and to screen for mutations with a potential impact on transmissibility, pathogenicity or countermeasures (e.g. diagnostics, antiviral drugs and vaccines) (WHO, 2020d). For this reason, the WHO recommends that routine sequencing of SARS-CoV-2 viruses be increased where possible, and sequence data be shared through open-source platforms (WHO 2020c). The ECDC similarly encourages EU/EEA states to sequence at least 500 randomly selected samples each week, and to prioritize samples from large outbreaks, or samples collected from travellers and individuals with an infection detected 2 weeks after vaccination (ECDC, 2021b).

Whole genome sequencing demands substantial investment in terms of time and financial resources, however, and the acquisition of sufficient, high quality RNA to maximize sequencing yield and the validity of sequence data. In addition, bioinformatic analysis necessitates high-performance computational power and trained staff. To address some of these problems, we developed a rapid, inexpensive and sensitive long nested PCR assay to detect key changes in the S protein of SARS-CoV-2, indicative of the presence of VOCs 20I/501Y.V1, 20H/501Y.V2 and 20J/501Y.V3, as well as of other variants of clinical interest. This assay can be used as a first screening tool to select samples of interest for WGS characterization. Resources for genome sequencing would thus be optimized by allocating them to clinically relevant specimens. The region amplified by this long assay was specially chosen to reveal key mutations distinctive of main variants, as shown in Table 1.

The discriminatory potential of the long PCR assay was initially explored in silico using GISAID and Nextclade. Phylogenetic trees predicted the ~1500 bps sequence of the S protein to have the same discriminatory power as the entire genome in assigning sequences with the S mutations characteristic of the VOCs to their corresponding clusters. The ability of primer pairs to amplify

441 SARS-CoV-2 RNA and the discriminatory power of the amplified fragments were then tested on
 442 field samples, consisting of SARS-CoV-2 viral isolates, already been fully sequenced. All of the
 443 clinical samples were successfully amplified with the long PCR. The 20I/501Y.V1 and 20E.EU1
 444 variants present in the panel were correctly identified by the long assay. Other variants of concern,
 445 like the Brazilian or the South African variants, were not available to us as field samples. Indeed,
 446 only one case of the South African variant had been detected in Italy, in a man who landed at
 447 Milan's Malpensa Airport from Africa at the end of January
 448 ([https://www.ansa.it/lombardia/notizie/2021/02/03/a-varese-primo-caso-variante-sudafricana-in-](https://www.ansa.it/lombardia/notizie/2021/02/03/a-varese-primo-caso-variante-sudafricana-in-italia_41b99448-367c-4000-ac88-9b3ae130d839.html)
 449 [italia_41b99448-367c-4000-ac88-9b3ae130d839.html](https://www.ansa.it/lombardia/notizie/2021/02/03/a-varese-primo-caso-variante-sudafricana-in-italia_41b99448-367c-4000-ac88-9b3ae130d839.html)). Having downloaded the sequence (GISAID
 450 EPI_ISL_1012924), we verified that it harbors the Spike mutations D80A, D215G, L242del,
 451 L244del, K417N, E484K, and N501Y - all of which would have been detectable by our long PCR
 452 assay. Similarly, only three cases of the Brazilian variant had been identified and fully sequenced
 453 (GISAID EPI_ISL_875566 to ID EPI_ISL_875568) thus far in Italy. All these isolates carry the
 454 characteristic Spike mutations D138Y, R190S E484K, K417T, and N501Y included in our assay.
 455 Once we had confirmed that the developed PCR assay could be successfully used to screen for
 456 variants in clinical samples, we went on to investigate whether this assay could also be used for the
 457 screening of SARS-CoV-2 variants in complex matrices, such as urban wastewaters. Since the
 458 amplification of long fragments poses specific challenges in such matrices, we also designed two
 459 short nested RT-PCR assays targeting portions of the region spanned by the long assay, to be used
 460 in case of long PCR failure. A number of studies have demonstrated the added value of regular
 461 surveillance of wastewaters in combination with other indicators for the management of the
 462 pandemic and a recent document by the European Commission highlights the importance of
 463 wastewater surveillance to allow screening among large population groups (European Commission,
 464 2021). This can help public health authorities identify where more detailed analysis is needed and
 465 speed up the detection of variants as part of an increased genomic and epidemiological surveillance.

Two recent papers have demonstrated that environmental surveillance can be used to study SARS-CoV-2 diversity (Crits-Christoph et al., 2021; Martin et al., 2020). Crits-Christoph and co-workers sequenced SARS-CoV-2 RNA directly from sewage collected in the San Francisco Bay Area to generate complete and near-complete SARS-CoV-2 genomes, and found that the major consensus genotypes detected in the sewage were identical to clinical genomes from the region. Moreover, additional variants not found in clinical samples, were identified in wastewaters, suggesting that wastewater sequencing can provide evidence for recent introductions of viral lineages before they are detected by local clinical sequencing (Crits-Christoph et al., 2021). Metagenomic analysis of environmental matrices is a complex task, however, with high operational costs and specific requirements in terms of bioinformatic skills.

Martin and co-workers demonstrated changes in virus variant predominance during the COVID-19 pandemic by tracking SARS-CoV-2 in sewage using a nested RT-PCR approach targeting five different regions of the viral genome (Martin et al., 2020). Here, we adopted a similar approach, but concentrated on a unique region of the SARS-CoV-2 genome - the spike protein gene, which harbors several key mutations associated with known VOCs and other variants of interest.

The newly-designed assays were therefore used on 34 sewage samples, including 20 SARS-CoV-2-positive samples collected from WTPs in Rome between September and December 2020, eight sewage sample collected from a small town in Abruzzo where an outbreak of the UK variant was reported in mid-January (https://www.ansa.it/abruzzo/notizie/2021/01/18/covid-a-guardiagrele-36-casi-da-variante-inglese_fbfc1c88-649b-4f32-9042-d1f0648b2bbb.html), and six sewage samples collected in Perugia, Umbria, where, at the beginning of February, both UK and Brazilian variants have been documented by public health authorities (<https://www.regione.umbria.it/documents/18/25027633/Ordinanza+n.14+del+6+febbraio+2021.pdf>).

As expected, only three of the samples could be amplified and successfully sequenced with the long PCR assay, one from Umbria, the others from Abruzzo. Wastewater is indeed a complex matrix

where the amplification of long fragments is seldom achieved, due to the presence of inhibitors, nucleic acids of different origin, and fragmented target genomes, often in low amounts. The sequences obtained by the long assay from samples collected in January in Abruzzo showed mutations characteristic of the Spanish variant, which was indeed the most common variant in the EU till December 2020 (ECDC, 2021a). The sequence obtained from Perugia displayed instead the mutations indicative of 20J/501Y.V3. Interestingly, this sample was collected on 6 February 2021 from a WTP serving the outskirts of Perugia, including the local hospital, which had reportedly been experiencing nosocomial clusters of this variant, as per press releases issued by public health officials (https://www.ansa.it/umbria/notizie/2021/02/05/rezza-in-umbria-variante-brasiliana-in-cluster-ospedalieri_d50f6dce-9b78-443c-a0d8-0d23e16058a8.html; <https://tg24.sky.it/salute-e-benessere/2021/02/08/covid-umbria>).

As expected, the two short nested RT-PCR assays were more sensitive than the long one, successfully amplifying about half of the tested sewage samples. No mutations of interest were detected in the samples collected in Rome and in Guardiagrele, Abruzzo, using the short assays, as the mutation A222V, characteristic of the Spanish variant, is not included in the short fragments. Conversely, sequences obtained from the sewage samples collected in Perugia, Umbria, showed the presence of mutations characteristic of the UK variant (one sample) and of the Brazilian variant (three samples, one of which was confirmed by the long assay). Both of these variants had been circulating in the area according to statements by public health authorities, but no genome sequences were available at the time of the study.

Comparing results obtained by the two nested PCRs for the same sample, concordant results were obtained for the Brazilian variant (presence of mutation D138Y in the fragment obtained by PCR 973, and mutations E484K/N501Y in the amplicon of PCR 975). In contrast, the sample harboring mutations of the UK variant in the region amplified by PCR 975 (N501Y and A570D) did not show the expected mutations (HV69-70del and Y144del) in the region amplified by PCR 973. This

exposes a limit of the combined use of two short PCR assays. Since different SARS-CoV-2 strains may coexist in a single sewage sample, the use of separate tests may result in each of the tests amplifying a sequence belonging to a different strain. However, our GISAID in silico study showed the combination of the N501Y and A570D mutations to strongly suggest that the sequence carrying them belongs to lineage GR B.1.1.7. Furthermore, we only performed classical Sanger sequencing on PCR amplicons, and this approach may underestimate the existence of some, possibly less common, sequences. Our next goal is therefore to combine the protocol with NGS on PCR amplicons for a more in-depth analysis of sequences, as successfully applied to enteric viruses (Suffredini et al., 2018; Iaconelli et al, 2017).

To our knowledge, this is the first evidence of the presence of sequences harboring key mutations of 20I/501Y.V1 and 20J/501Y.V3, as well as of the Spanish variant in urban wastewaters, highlighting the potential contribution of wastewater surveillance to explore SARS-CoV-2 diversity. Moreover, we demonstrated the potential of this approach, based on nested amplification followed by conventional Sanger sequencing, to allow the rapid identification of specific VOCs in the catchment of any given WTP to quickly identify areas where the intensification of clinical surveillance and/or targeted preventive intervention is required.

In conclusion, the method we propose can be used to rapidly screen SARS-CoV-2-positive samples for VOCs and other variants of clinical interest. Compared with WGS, conventional RT-PCR followed by Sanger sequencing can be performed routinely at low cost and does not require specialized expertise to interpret the results. Through the targeting of key mutations, this approach could enable the rapid screening of samples and the identification of samples to submit to WGS. Moreover, the long nested RT-PCR described here can provide significant, albeit partial, molecular information for samples in which low amounts or poor quality SARS-CoV-2 RNA hamper the successful application of WGS. The approach can also be applied to environmental samples for the purposes of SARS-CoV-2 wastewater-based epidemiology, to explore the possible presence of

541 VOCs or other variants of clinical interest, therefore highlighting geographical areas where an
542 enhancement of clinical monitoring could be beneficial.

543 Due to an increase in the prevalence of new SARS-CoV2 variants with possible clinical
544 implications, monitoring the spread of different variants in the general population in a timely, cost-
545 effective manner is of utmost importance. The proposed method can contribute to achieve this goal.

546

547

548 **Declaration of interests**

549 We declare no competing interests.

Acknowledgments

This publication was supported by the European Virus Archive Global (EVA-GLOBAL) project, which provided the coronavirus RNA panel for the assessment of our in-house PCR. The EVA-GLOBAL project, in turn, received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 871029.

Our thanks go to "Project RefBio - Germany's contribution to strengthen the bio-analytical reference laboratories in the UNSGM (2017-2021), EQAE on SARS-CoV-2 (2020), Centre for Biological Threats and Special Pathogens (ZBS 1-3) at RKI (Project Coordinator), Berlin, Germany" for providing the Wuhan strain.

We also thank Stefano Fontana for his insightful comments.

We thank Luigina Serrecchia, Viviana Manzulli, Valeria Rondinone, Ines della Rovere, Fiorenza Petruzzi and Mr. Guido Asturio for their technical support.

We acknowledge the support of Utilitalia, Italian federation of energy, water and environmental services, and ACEA ATO2 for wastewater sampling in Rome. We also thank Tommaso Pagliani and Fabio Campitelli of SASI SpA (Società Abruzzese per il Servizio Idrico Integrato) and the Municipal Authorities for allowing wastewater collection in Guardiagrele (CH), Abruzzo, and Tiziana Buonfiglio, Andrea Vitali, Paolo Meniconi, Massimo Lorenzetti Renzetti and Marino Burini, Umbra Acque SpA, for allowing wastewater collection in Perugia, Umbria.

571 References

- 572 Bartolini B, Rueca M, Ernesto C, Gruber M, Messina F, Giombini E, Ippolito G, Capobianchi
573 M R, Di Caro A. The newly introduced SARS-CoV-2 variant A222V is rapidly spreading in Lazio
574 region, Italy. medRxiv 2020.11.28.20237016; doi: <https://doi.org/10.1101/2020.11.28.20237016>
- 575 CDC, 2020. Interim: Implications of the Emerging SARS-CoV-2 Variant VOC 202012/01. Interim:
576 Implications of the Emerging SARS-CoV-2 Variant VOC 202012/01 | CDC, accessed on 12
577 January, 2021.
- 578 Chand, Meera; Hopkins, Susan; Dabrera, Gavin; Achison, Christina; Barclay, Wendy; Ferguson,
579 Neil; Volz, Erik; Loman, Nick; Rambaut, Andrew; Barrett, Jeff (21 December 2020). Investigation
580 of novel SARS-COV-2 variant: Variant of Concern 202012/01 (PDF)(Report). Public Health
581 England. Retrieved 12 january 2021.
- 582 China CDC Weekly. The Epidemiological Characteristics of an Outbreak of 2019 Novel
583 Coronavirus Diseases (COVID-19) — China, 2020, COVID-19.pdf.pdf (onb.it)
- 584 Cleemput S, Dumon W, Fonseca V, Karim WA, Giovanetti M, Alcantara LC, Deforche K, de
585 Oliveira T. Genome Detective Coronavirus Typing Tool for rapid identification and
586 characterization of novel coronavirus genomes. bioRxiv [Preprint]. 2020 Feb 2:2020.01.31.928796.
587 doi: 10.1101/2020.01.31.928796. Update in: Bioinformatics. 2020 Jun 1;36(11):3552-3555. PMID:
588 32511309; PMCID: PMC7217295.
- 589 Conti P, Caraffa A, Gallenga CE, Kritas SK, Frydas I, Younes A, Di Emidio P, Tetè G, Pregliasco
590 F, Ronconi G. The British variant of the new coronavirus-19 (Sars-Cov-2) should not create a
591 vaccine problem. J Biol Regul Homeost Agents. 2021 Feb 24;35(1). doi: 10.23812/21-3-E. Epub
592 ahead of print. PMID: 33377359.

593 Crits-Christoph A, Kantor RS, Olm MR, Whitney ON, Al-Shayeb B, Lou YC, Flamholz A,
594 Kennedy LC, Greenwald H, Hinkle A, Hetzel J, Spitzer S, Koble J, Tan A, Hyde F, Schroth G,
595 Kuersten S, Banfield JF, Nelson KL. Genome Sequencing of Sewage Detects Regionally Prevalent
596 SARS-CoV-2 Variants. mBio. 2021 Jan 19;12(1):e02703-20. doi: 10.1128/mBio.02703-20. PMID:
597 33468686.

598 ECDC, 2020. European Centre for Disease Prevention and Control, 2020. Threat Assessment Brief:
599 Rapid increase of a SARS-CoV-2 variant with multiple spike protein mutations observed in the
600 United Kingdom. Threat Assessment Brief: Rapid increase of a SARS-CoV-2 variant with multiple
601 spike protein mutations observed in the United Kingdom (europa.eu), accessed on 12 January,
602 2021.

603 ECDC, 2021a. European Centre for Disease Prevention and Control, 2020. Risk related to spread of
604 new SARSCoV-2 variants of concern in the EU/EEA 29 December 2020

605 ECDC, 2021b. European Centre for Disease Prevention and Control, 2021. Risk of spread of new
606 SARS-CoV-2 variants of concern in the EU/EEA - first update (europa.eu) ECDC

607 EUROPEAN COMMISSION, 2021. HERA Incubator: Anticipating together the threat of COVID-
608 19 variants [https://ec.europa.eu/info/sites/info/files/communication-hera-incubator-anticipating-](https://ec.europa.eu/info/sites/info/files/communication-hera-incubator-anticipating-threat-covid-19-variants_en.pdf)
609 [threat-covid-19-variants_en.pdf](https://ec.europa.eu/info/sites/info/files/communication-hera-incubator-anticipating-threat-covid-19-variants_en.pdf)

610 Foladori P, Cutrupi F, Segata N, Manara S, Pinto F, Malpei F, Bruni L, La Rosa G. SARS-CoV-2
611 from faeces to wastewater treatment: What do we know? A review. Sci Total Environ. 2020 Nov
612 15;743:140444. doi: 10.1016/j.scitotenv.2020.140444. Epub 2020 Jun 24. PMID: 32649988;
613 PMCID: PMC7311891. GISAID 2020, UK reports new variant, termed VUI 202012/01. GISAID -
614 UK reports new variant, termed VUI 202012/01, accessed on 12 January, 2021.

615 Galloway SE, Paul P, MacCannell DR, et al. Emergence of SARS-CoV-2 B.1.1.7 Lineage —
616 United States, December 29, 2020–January 12, 2021. *MMWR Morb Mortal Wkly Rep*. ePub: 15
617 January 2021. DOI: <http://dx.doi.org/10.15585/mmwr.mm7003e2external icon>.

618 Groves DC, Rowland-Jones SL, Angyal A. The D614G mutations in the SARS-CoV-2 spike
619 protein: Implications for viral infectivity, disease severity and vaccine design [published online
620 ahead of print, 2020 Nov 5]. *Biochem Biophys Res Commun*. 2020;S0006-291X(20)32038-6.
621 doi:10.1016/j.bbrc.2020.10.109

622 Gu, H. et al. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* 369
623 (2020).

624 Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher
625 RA. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics*. 2018 Dec 1;34(23):4121-
626 4123. doi: 10.1093/bioinformatics/bty407. PMID: 29790939; PMCID: PMC6247931.

627 Hammer AS, Quaade ML, Rasmussen TB, Fonager J, Rasmussen M, Mundbjerg K, Lohse L,
628 Strandbygaard B, Jørgensen CS, Alfaro-Núñez A, Rosenstjerne MW, Boklund A, Halasa T,
629 Fomsgaard A, Belsham GJ, Bøtner A. SARS-CoV-2 Transmission between Mink (*Neovison vison*)
630 and Humans, Denmark. *Emerg Infect Dis*. 2020 Nov 18;27(2). doi: 10.3201/eid2702.203794. Epub
631 ahead of print. PMID: 33207152.

632 Hart OE, Halden RU. Computational analysis of SARS-CoV-2/COVID-19 surveillance by
633 wastewater-based epidemiology locally and globally: Feasibility, economy, opportunities and
634 challenges. *Sci Total Environ*. 2020 Aug 15;730:138875. doi: 10.1016/j.scitotenv.2020.138875.
635 Epub 2020 Apr 22. PMID: 32371231; PMCID: PMC7175865.

636 Hodcroft EB, Zuber M, Nadeau S, Crawford KHD, Bloom JD, Veessler D, Vaughan TG, Comas I,
637 Candelas FG, Stadler T, Neher RA. Emergence and spread of a SARS-CoV-2 variant through
638 Europe in the summer of 2020. medRxiv [Preprint]. 2020 Nov 27:2020.10.25.20219063. doi:
639 10.1101/2020.10.25.20219063. PMID: 33269368; PMCID: PMC7709189.

640 Iaconelli M, Valdazo-González B, Equestre M, Ciccaglione AR, Marcantonio C, Della Libera S, La
641 Rosa G. Molecular characterization of human adenoviruses in urban wastewaters using next
642 generation and Sanger sequencing. Water Res. 2017 Sep 15;121:240-247. doi:
643 10.1016/j.watres.2017.05.039. Epub 2017 May 19. PMID: 28550812.

644 . Guthrie JL, Teatero S, Zittermann S, Chen Y, Sullivan A, Rilkoff H, Joshi E, Sivaraman K, de
645 Borja R, Sundaravadanam Y, Laszloffy M, Heisler L, Allen VG., Simpson JT., Fittipaldi N.
646 Detection of the Novel SARS-CoV-2 European Lineage B.1.177 in Ontario, Canada.
647 medRxiv 2020.11.30.20241265; doi: <https://doi.org/10.1101/2020.11.30.20241265>

648 Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, Hengartner N, Giorgi EE,
649 Bhattacharya T, Foley B, Hastie KM, Parker MD, Partridge DG, Evans CM, Freeman TM, de Silva
650 TI; Sheffield COVID-19 Genomics Group, McDanal C, Perez LG, Tang H, Moon-Walker A,
651 Whelan SP, LaBranche CC, Sapphire EO, Montefiori DC. Tracking Changes in SARS-CoV-2 Spike:
652 Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell. 2020 Aug 20;182(4):812-
653 827.e19. doi: 10.1016/j.cell.2020.06.043. Epub 2020 Jul 3. PMID: 32697968; PMCID:
654 PMC7332439.

655 La Rosa G, Mancini P, Bonanno Ferraro G, Veneri C, Iaconelli M, Bonadonna L, Lucentini L,
656 Suffredini E. SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence
657 from environmental monitoring. Sci Total Environ. 2021 Jan 1;750:141711. doi:
658 10.1016/j.scitotenv.2020.141711. Epub 2020 Aug 15. PMID: 32835962; PMCID: PMC7428442.

659 Leung K, Shum MH, Leung GM, Lam TT, Wu JT. Early transmissibility assessment of the N501Y
660 mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Euro Surveill.*
661 2021 Jan;26(1):2002106. doi: 10.2807/1560-7917.ES.2020.26.1.2002106. PMID: 33413740;
662 PMCID: PMC7791602.

663 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,
664 Li X, Sun Q, Liu J, Zhang L, Li X, Huang W, Wang Y. The Impact of Mutations in SARS-CoV-2
665 Spike on Viral Infectivity and Antigenicity. *Cell.* 2020 Sep 3;182(5):1284-1294.e9. doi:
666 10.1016/j.cell.2020.07.012. Epub 2020 Jul 17. PMID: 32730807; PMCID: PMC7366990.

667 Maggi F, Novazzi F, Genoni A, Baj A, Spezia PG, Focosi D, Zago C, Colombo A, Cassani G,
668 Pasciuta R, Tamborini A, Rossi A, Prestia M, Capuano R, Azzi L, Donadini A, Catanoso G, Grossi
669 PA, Maffioli L, Bonelli G. Imported SARS-COV-2 Variant P.1 Detected in Traveler Returning
670 from Brazil to Italy. *Emerg Infect Dis.* 2021 Feb 10;27(4). doi: 10.3201/eid2704.210183. Epub
671 ahead of print. PMID: 33567246.

672 Mahase, 2020. Covid-19: What have we learnt about the new variant in the UK?
673 *BMJ* 2020; 371 doi: <https://doi.org/10.1136/bmj.m4944> (Published 23 December 2020)

674 Martin J, Klapsa D, Wilton T, Zambon M, Bentley E, Bujaki E, Fritzsche M, Mate R, Majumdar M.
675 Tracking SARS-CoV-2 in Sewage: Evidence of Changes in Virus Variant Predominance during
676 COVID-19 Pandemic. *Viruses.* 2020 Oct 9;12(10):1144. doi: 10.3390/v12101144. PMID:
677 33050264; PMCID: PMC7601348.

678 Martin J, Klapsa D, Wilton T, Zambon M, Bentley E, Bujaki E, Fritzsche M, Mate R, Majumdar M.
679 Tracking SARS-CoV-2 in Sewage: Evidence of Changes in Virus Variant Predominance during
680 COVID-19 Pandemic. *Viruses.* 2020 Oct 9;12(10):1144. doi: 10.3390/v12101144. PMID:
681 33050264; PMCID: PMC7601348.

682 Martínez-Puchol S, Rusiñol M, Fernández-Cassi X, Timoneda N, Itarte M, Andrés C, Antón A,
683 Abril JF, Girones R, Bofill-Mas S. Characterisation of the sewage virome: comparison of NGS
684 tools and occurrence of significant pathogens. Sci Total Environ. 2020 Apr 15;713:136604. doi:
685 10.1016/j.scitotenv.2020.136604. Epub 2020 Jan 8. PMID: 31955099.

686 NERVTAG, 2020. New and Emerging Respiratory Virus Threats Advisory Group. NERVTAG
687 meeting on SARS-CoV-2 variant under investigation VUI-202012/01, 18 December 2020.
688 NERVTAG meeting minutes on SARS-CoV-2 variant under investigation. 18 December
689 (khub.net), accessed on 12 January, 2021.

690 NIID, 2021. Brief report: New Variant Strain of SARS-CoV-2 Identified in Travelers from Brazil.
691 PUBLISHED: 12 JANUARY 2021. Brief report: New Variant Strain of SARS-CoV-2 Identified in
692 Travelers from Brazil (niid.go.jp). Accessed on 16 January 2021.

693

694 Public Health England (PHE) 2021. Investigation of novel SARS-CoV-2 variant: Variant of
695 Concern 202012/01, Technical briefing 5. 2020 [updated January 2021]. Available from:
696 [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/9](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957504/Variant_of_Concern_VOC_202012_01_Technical_Briefing_5_England.pdf)
697 [57504/Variant of Concern VOC 202012 01 Technical Briefing 5 England.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/957504/Variant_of_Concern_VOC_202012_01_Technical_Briefing_5_England.pdf)

698 Pillay TS. Gene of the month: the 2019-nCoV/SARS-CoV-2 novel coronavirus spike protein. J Clin
699 Pathol. 2020 Jul;73(7):366-369. doi: 10.1136/jclinpath-2020-206658. Epub 2020 May 6. PMID:
700 32376714.

701 Ester C Sabino, Lewis F Buss, Maria P S Carvalho, Carlos A Prete Jr, Myuki A E Crispim, Nelson
702 A Fraiji, Rafael H M Pereira, Kris V Parag, Pedro da Silva Peixoto, Moritz U G Kraemer, Marcio K
703 Oikawa, Tassila Salomon, Zulma M Cucunuba, Márcia C Castro, Andreza Aruska de Souza Santos,

704 Vítor H Nascimento, Henrique S Pereira, Neil M Ferguson, Oliver G Pybus, Adam Kucharski,
705 Michael P Busch, Christopher Dye, Nuno R Faria. Resurgence of COVID-19 in Manaus, Brazil,
706 despite high seroprevalence. The Lancet, Published: January 27,
707 2021 DOI: [https://doi.org/10.1016/S0140-6736\(21\)00183-5](https://doi.org/10.1016/S0140-6736(21)00183-5)

708 Starr, Tyler N., Allison J. Greaney, Sarah K. Hilton, Daniel Ellis, Katharine H. D. Crawford, Adam
709 S. Dingens, Mary Jane Navarro, et al. 2020. "Deep Mutational Scanning of SARS-CoV-2 Receptor
710 Binding Domain Reveals Constraints on Folding and ACE2 Binding." Cell 182 (5): 1295–
711 1310.e20.

712 Statens Serum Institut, Copenhagen S, DENMARK, 2020. Preliminary rapport on SARS-CoV-2
713 spike mutations arising in Danish mink, 2 their spread to humans and neutralization data. SARS-
714 CoV-2 spike mutations arising in Danish mink and their spread to 5 humans

715 Suffredini E, Iaconelli M, Equestre M, Valdazo-González B, Ciccaglione AR, Marcantonio C, Della
716 Libera S, Bignami F, La Rosa G. Genetic Diversity Among Genogroup II Noroviruses and
717 Progressive Emergence of GII.17 in Wastewaters in Italy (2011-2016) Revealed by Next-
718 Generation and Sanger Sequencing. Food Environ Virol. 2018 Jun;10(2):141-150. doi:
719 10.1007/s12560-017-9328-y. Epub 2017 Nov 28. Erratum in: Food Environ Virol. 2018 May 4;:
720 PMID: 29185203.

721 Volz E, Hill V, McCrone JT, Price A, Jorgensen D, O'Toole Á, Southgate J, Johnson R, Jackson B,
722 Nascimento FF, Rey SM, Nicholls SM, Colquhoun RM, da Silva Filipe A, Shepherd J, Pascall DJ,
723 Shah R, Jesudason N, Li K, Jarrett R, Pacchiarini N, Bull M, Geidelberg L, Siveroni I; COG-UK
724 Consortium, Goodfellow I, Loman NJ, Pybus OG, Robertson DL, Thomson EC, Rambaut A,
725 Connor TR. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and

726 Pathogenicity. Cell. 2021 Jan 7;184(1):64-75.e11. doi: 10.1016/j.cell.2020.11.020. Epub 2020 Nov
727 19. PMID: 33275900; PMCID: PMC7674007.

728 Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality.
729 Euro Surveill. 2017 Mar 30;22(13):30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494. PMID:
730 28382917; PMCID: PMC5388101.

731 Suffredini E, Iaconelli M, Equestre M, Valdazo-González B, Ciccaglione AR, Marcantonio C, Della
732 Libera S, Bignami F, La Rosa G. Genetic Diversity Among Genogroup II Noroviruses and
733 Progressive Emergence of GII.17 in Wastewaters in Italy (2011-2016) Revealed by Next-
734 Generation and Sanger Sequencing. Food Environ Virol. 2018 Jun;10(2):141-150. doi:
735 10.1007/s12560-017-9328-y. Epub 2017 Nov 28. Erratum in: Food Environ Virol. 2018 May 4;:
736 PMID: 29185203.

737 WHO, 2020a. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11
738 March 2020. 11 March 2020

739 WHO, 2020b. Emergencies preparedness, response. SARS-CoV-2 mink-associated variant strain –
740 Denmark Disease Outbreak News.6 November 2020. WHO | SARS-CoV-2 mink-associated variant
741 strain – Denmark

742 WHO 2020c. Emergencies preparedness, response. SARS-CoV-2 Variant, WHO | SARS-CoV-2
743 Variants

744 WHO 2020d. SARS-CoV-2 Variant – United Kingdom of Great Britain and Northern Ireland

745 Disease Outbreak News, 21 December 2020 WHO | SARS-CoV-2 Variant – United Kingdom of
746 Great Britain and Northern Ireland

747 WHO, 2021. Genomic sequencing of SARS-CoV-2 A guide to implementation for maximum
748 impact on public health 8 January 2021. [9789240018440-eng.pdf](#)

749

750

Figure 1. Positions of the primers in the newly designed PCR assays in the Spike gene of SARS-CoV-2

Figure 2. Nextclade web tool phylogenetic trees for representative strains of the different variants using the full genome or the Spike partial gene fragment amplified with PCR ID 980 - nextstrain.org with ad hoc modifications.

Table 1. Mutations indicative of the main SARS-CoV-2 VOC detectable by the newly designed PCR assays

Variant	Nucleotide position in the S gene																			
	207	210	240	414	432	456	570	645	666	726	729	732	735	738	1251	1356	1359	1452	1503	1710
20I/501Y.V1 (GR, B.1.1.7)	H69 del	V70 del			Y144 del														N501Y	A570D
20H/501Y.V2 (GH, B.1.351)			D80A					D215G		L242H	A243 del	L244 del	H245 del	R246I	K417N			E484K	N501Y	
20J/501Y.V3 (GR, P1)				D138Y			R190S								K417T			E484K	N501Y	
20E.EU1 (GV, B.1.177)									A222V											
CAL 20C (GH, B.1.429)						W152C										L452R				
Mink cluster V (GR, B.1.1.298)	H69 del	V70 del															Y453F			
PCR ID	PCR 973 (171-452)																			
(nt position in the S gene)																PCR 975 (1436-1719)				
	PCR 980 (171-1719)																			

Table 2: Primers and PCRs used in this study

PCR ID	Primer ID and sequence (5'-3')	Positions *	Amplicon (bps)
972	2319 ACCCTGACAAAGTTTTCAGATCCT	21675-21698	1 st cycle
	2320 GCTGAGAGACATATTCAAAAGTGCA	22082-22058	399 to 408 bps
973	2321 TTCAACTCAGGACTTGTTCTTACC	21709-21732	nested
	2322 TCTGAACTCACTTTCCATCCAA	22036-22015	319 to 327 bps
974	2323 TTGTTTAGGAAGTCTAATCTCAAACC	22925-22950	1 st cycle
	2326 GTGGATCACGGACAGCATC	23300-23282	375 bps
975	2325 CTATCAGGCCGGTAGCACAC	22978-22997	hemi-nested
	2326 GTGGATCACGGACAGCATC	23300-23282	323 bps
979	2319 ACCCTGACAAAGTTTTCAGATCCT	21675-21698	1 st cycle
	2324 CCTGATAAAGAACAGCAACCTG	23402-23381	1719**-1728 bps
980	2321 TTCAACTCAGGACTTGTTCTTACC	21709-21732	nested
	2326 GTGGATCACGGACAGCATC	23300-23282	1583**-1592 bps

* Reference genome NC45512.2 (Wuhan)

** Expected size for the 'UK variant'

Table 3: Variant analysis of clinical samples obtained using PCR ID 980 and different bioinformatics tools

Sample ID	Identification based on WGS	Mutation map ¹	GISAID blast	Nextclade
swab_25	G B.1 / 20A	-	Not assignable ²	19A
swab_26	G B.1 / 20A	D215H	Not assignable ³	19A
swab_27	GR B.1.1.316 / 20B	-	Not assignable ²	19A
swab_28	GR B.1.1.229 / 20B	-	Not assignable ²	19A
swab_29	GV B.1.177 / 20E EU1	A222V	GV_B.1.177	20E (EU1)
swab_30	GV B.1.177 / 20E EU1	A222V	GV_B.1.177	20E (EU1)
swab_31	GR B.1.1.7 / 20I/501Y.V1	H69del, V70del, Y144del, N501Y, A570D	GR/501Y.V1 (B.1.1.7)	20B

¹ Analysis with GISAID CoVsurver: Mutation Analysis (<https://www.gisaid.org/epiflu-applications/covsurver-mutations-app/>)

² 100% identity with sequences assigned to the following clades: GR_B.1.1.222; GR_B.1.1.70; GR_B.1.1.119; GR_B.1.1.28; GR_B.1.1.316; GR_B.1.1.33; GR_B.1.1.103; GH_B.1.2; GH_B.1.425; G_B.1; G_B.1.91; GH_B.1.4703

³ 100% identity with sequences assigned to the following clades: G_B.1; GH_B.1.2; GR_B.1.1.47; GR_B.1.1.304; GR_B.1.1.127; GR_B.1.1.284; GR_B.1.1.119; O_B-13

n.d. not done due to absence of PCR amplification

Table 4. Variant analysis of environmental samples obtained using PCR IDs 980, 973, and 975

Sample ID	Sampling location	PCR ID 980		PCR ID 973		PCR ID 975	
		result	mutations	result	mutations	result	mutations
Archive samples							
3716	Rome East	-		-		+	none
3732	Rome North	-		+	none	+	none
3737	Rome East	-		-		+	none
3738	Rome East	-		+	none	+	none
3740	Rome East	-		-		+	none
3742	Rome Ostia	-		+	none	+	none
3750	Rome North	-		-		+	none
3798	Rome East	-		+	none	+	none
3801	Rome East	-		+	none	-	
3802	Rome East	-		-		+	none
3805	Rome East	-		+	none	-	
3808	Rome South	-		+	none	-	
3809	Rome Ostia	-		+	none	+	none
Ad-hoc samples from outbreak locations							
3862	Guardiagrele	-		+	none	+	none
3863	Guardiagrele	+	A222V	+	none	+	none
3865	Guardiagrele	+	A222V	+	none	-	
3944	Perugia	-		+	D138Y	+	E484K, N501Y
3945	Perugia	-		+	D138Y	+	E484K, N501Y
3947	Perugia	+	D138Y, R190S, K417T, E484K, and N501Y	+	D138Y	+	E484K, N501Y
3949	Perugia	-		+	E96G	+	N501Y, A570D

Samples IDs 3702, 3714, 3720, 3739, 3744, 3797, 3815, 3864, 3868, 3869, 3946 and 3948 did not provide amplification in any of the assays and were not included in the table.