



## 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



G La Rosa<sup>a,\*</sup>, P. Mancini<sup>a</sup>, G. Bonanno Ferraro<sup>a</sup>, C. Veneri<sup>a</sup>, M. Iaconelli<sup>a</sup>, L. Lucentini<sup>a</sup>, L. Bonadonna<sup>a</sup>, S. Brusaferrò<sup>b</sup>, D. Brandtner<sup>c</sup>, A. Fasanella<sup>d</sup>, L. Pace<sup>d</sup>, A. Parisi<sup>d</sup>, D. Galante<sup>d</sup>, E. Suffredini<sup>e</sup>

<sup>a</sup> Department of Environment and Health, Italian National Institute of Health, Rome, Italy

<sup>b</sup> Office of the President, Istituto Superiore di Sanità, Rome, Italy

<sup>c</sup> Independent Researcher, Rome, Italy

<sup>d</sup> Experimental Zooprophyllactic Institute of the Apulia and Basilicata Regions, Foggia, Italy

<sup>e</sup> Department of Food Safety, Nutrition and Veterinary Public Health, Italian National Institute of Health, Rome, Italy

### ARTICLE INFO

#### Article history:

Received 27 February 2021

Revised 28 March 2021

Accepted 30 March 2021

Available online 2 April 2021

#### Keywords:

SARS-CoV-2

Variant

Mutation

VOC

PCR

Sequencing

### ABSTRACT

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 nucleotide changes resulting in 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 20 J/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 GISAIID 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 31 clinical samples (7 fully-sequenced swab samples, and 24 uncharacterized ones) and 34 urban wastewater samples, some of which collected in areas where circulation of VOCs had been reported.

The long assay successfully amplified 29 of the 31 swabs (93%), allowing the correct identification of variants 20I/501Y.V1 and 20E.EU1 present in the panel of previously characterized samples. The Spanish variant was detected in 14/24 of the uncharacterized samples as well. 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 20 J/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.

© 2021 The Author(s). Published by Elsevier Ltd.  
This is an open access article under the CC BY-NC-ND license  
(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

## 1. Introduction

On 31 December 2019, the Wuhan Municipal Health and Health Commission reported a cluster of pneumonia cases of unknown etiology. 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, 2020).

One amino acid substitution 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., 2021).

In August/September 2020, a SARS-CoV-2 variant linked to infection among farmed minks, subsequently transmitted to humans, was identified in Denmark (Hammer et al., 2021). In all six countries (Denmark, the Netherlands, Spain, Sweden, Italy and the United States) reported cases of farmed mink infected with SARS-CoV-2 to the World organisation for animal health (OIE) (WHO, 2020b).

An additional novel variant - clade GV, lineage B.1.177 (GISAID) or 20E.EU1 (Nextstrain) - emerged in early summer of 2020 in Spain, and subsequently spread to multiple locations in Europe, possibly introduced into other countries by summer tourists (Hodcroft et al., 2020).

On 14 December 2020, British authorities announced that a new SARS-CoV-2 variant had been identified through viral genomic sequencing (WHO, 2020c). The new variant belongs to clade GR, lineage B.1.1.7 (GISAID), or clade 20I/501Y.V1 (Nextstrain) and was discovered following an unforeseen rise in COVID-19 cases in South East England, between October and December 2020 (WHO, 2020d). As of 19 January 2021, 23 EU/EEA countries, including Italy, and 37 additional countries have reported the new COVID-19 variant 20I/501Y.V1 (ECDC, 2021). The emergence of the new SARS-CoV-2 variant has prompted some countries to close their borders with the UK. A review by the UK's New and Emerging Respiratory Virus Threats Advisory Group (NERVTAG) advises that there is moderate confidence that 20I/501Y.V1 has a substantial increase in transmissibility compared to other variants (NERVTAG, 2020). Recent findings are consistent in indicating increased disease severity in people infected with VOC B.1.1.7 compared to people infected with non-VOC virus variants (NEVERTAG, 2021). As for antigenic response, to date, it is considered unlikely that 20I/501Y.V1 may hinder vaccine-induced immunity (Conti et al., 2020).

On 18 December 2020, national authorities in South Africa announced the detection of a new variant of SARS-CoV-2, belonging to clade GH, lineage B.1.351 (GISAID) or 20H/501Y.V2 (Nextstrain). 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, 2021). 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](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. There is no evidence that it is associated with a higher severity of infection (ECDC, 2020).

In January, another variant, clade GR lineage P.1 (GISAID) or 20 J/501Y.V3 (Nextstrain) was identified in Japan in four travellers who had arrived from Brazil (NIID, 2021; Galloway et al., 2021). This variant was later identified in the capital of Amazonas, Manaus, where an upsurge in COVID-19 cases was registered in January of 2021 (Faria 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, 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 (PHE, 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 amino acid substitution and the Y144del - a deletion potentially contributing to spike surface variation. The N501Y amino acid substitution is in the receptor binding domain (RBD), 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 amino acid substitution (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 amino acid substitutions (the most common are L18F, K417N, E484K and K417N) in the S protein, which are not usually present in the UK strain, though amino acid substitution 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, 20 J/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).

\* Corresponding author: Giuseppina La Rosa, Department of Environment and Health, Italian National Institute of Health, Rome, Italy.  
E-mail address: [giuseppina.larosa@iss.it](mailto:giuseppina.larosa@iss.it) (G. La Rosa).

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 and Halden, 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 (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.

## 2. Materials and methods

### 2.1. 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 Fig. 1) were designed using the Primer 3 software (<http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi>).

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

Variant	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							L452R														
CAL.20C (GH, B.1.429)						W152C																								
Mink cluster V (GR, B.1.1.298)	H69 del	V70 del																												
PCR ID	<b>PCR 973 (a)</b>										<b>PCR 980 (b)</b>										<b>PCR 975 (c)</b>									

<sup>a</sup>Amino acid position 58 to 150 of the spike protein (primers excluded).

<sup>b</sup>Amino acid position 58 to 573 of the spike protein (primers excluded).

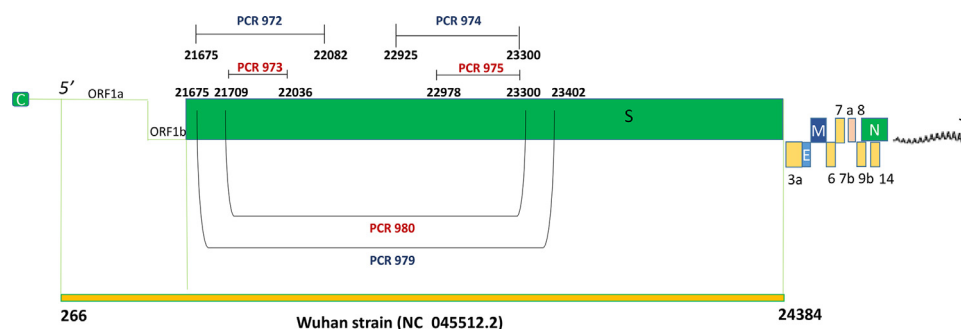
<sup>c</sup>Amino acid position 479 to 573 of the spike protein (primers excluded).

**Table 2**  
Primers and PCRs used in this study.

PCR ID	Primer ID and sequence (5'–3')	Positions *	Amplicon (bps)
972	2319 ACCCTGACAAAGTTTTTCAGATCCT	21,675–21,698	1st cycle
	2320 GCTGAGAGACATATTCAAAAGTGCA	22,082–22,058	399** to 408 bps
973	2321 TTCAACTCAGGACTTGTCTTACC	21,709–21,732	nested
	2322 TCTGAACTCACTTTCATCCAA	22,036–22,015	319** to 327 bps
974	2323 TTGTTTAGGAAGTCTAATCTCAAACC	22,925–22,950	1st cycle
	2326 GTGGATCACGGACAGCATC	23,300–23,282	375 bps
975	2325 CTATCAGGCCGCTAGCACAC	22,978–22,997	hemi-nested
	2326 GTGGATCACGGACAGCATC	23,300–23,282	323 bps
979	2319 ACCCTGACAAAGTTTTTCAGATCCT	21,675–21,698	1st cycle
	2324 CCTGATAAAGAACAGCAACCTG	23,402–23,381	1719** to 1728 bps
980	2321 TTCAACTCAGGACTTGTCTTACC	21,709–21,732	Nested
	2326 GTGGATCACGGACAGCATC	23,300–23,282	1583** to 1592 bps

\* Reference genome NC45512.2 (Wuhan).

\*\* Expected size for the 'UK variant'.



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

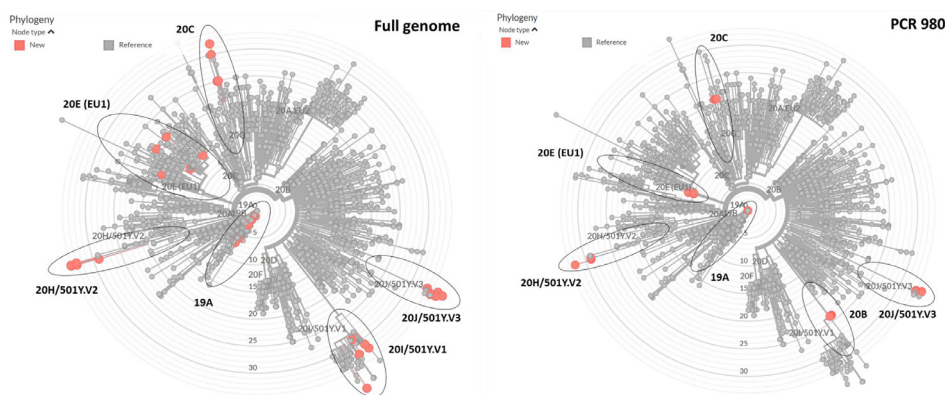
Ultimately, three assays were designed with different aims:

- 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.
- 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).
- 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 amino acid substitution N501Y shared by different VOCs, which is combined with substitution A570D in the 20I/501Y.V1 variant, and with E484K in both the 20H/501Y.V2 and 20J/501Y.V3 variants.

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 full genomes 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 in complete genomes 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](https://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/>



**Fig. 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.

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, 20 J/501Y.V3 and 20E.EU1. Only the 'UK variant' was placed at the base of the 20I/501Y.V1 branch (Fig. 2).

## 2.2. 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 prototype 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.)/ $\mu\text{L}$  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.

## 2.3. Field samples

- i) **Clinical samples:** Thirty-one RNA samples, including the above-mentioned sample of a 20I/501Y.V1 variant, provided by the Experimental Zooprophyllactic Institute of Apulia and Basilicata, were used for primer validation (Supplementary Materials). These were randomly selected among samples from routine COVID-19 testing.

Of these, seven SARS-CoV-2 RNA samples originating from nasopharyngeal swabs, had previously been extracted from viruses isolated on Vero E6 cells, 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./ $\mu\text{L}$ . The samples were subsequently diluted to an approximate concentration of  $10^4$  g.c./ $\mu\text{L}$  before testing by nested PCR.

Twenty-four additional RNA samples originated from nasopharyngeal swabs collected in Apulia and Basilicata between November 23, 2020 and January 18, 2021. These tested positive for SARS-CoV-2 by real-time RT-PCR (GeneFinder COVID-19 Plus RealAmp kit, OSANG Healthcare, Anyang-si, Korea), but had not been characterized by WGS. Viral RNA concentrations in these swab samples ranged from  $3.2 \times 10^1$  to  $1.1 \times 10^6$  g.c./ $\mu\text{L}$ . The samples were tested undiluted to assess the performance of the designed PCRs on normal clinical sample concentrations.

The performed activities on nasopharyngeal swabs were conducted as part of the legislated mandate of the Health Promotion and Public Health Department of Apulia. All procedures were carried out in accordance with the Declaration of Helsinki, as revised in 2013.

- 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 NucleiSens extraction system (bioMerieux, Marcy-l'Etoile, France). The eluted RNA (100  $\mu\text{L}$ ) was aliquoted and stored at  $-80^\circ\text{C}$  until molecular analysis. Viral concentrations in these samples ranged from 2.0 to 37.8 g.c./ $\mu\text{L}$  of RNA (corresponding to  $1.6 \times 10^3$  to  $3.0 \times 10^4$  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 Guardiagrele in Abruzzo (central Italy), where a cluster linked to the UK variant had been reported ([https://www.ansa.it/abruzzo/notizie/2021/01/18/covid-a-guardiagrele-36-casi-da-variante-inglese\\_fbfc1c88-649b-4f32-9042-d1f0648b2bbb.html](https://www.ansa.it/abruzzo/notizie/2021/01/18/covid-a-guardiagrele-36-casi-da-variante-inglese_fbfc1c88-649b-4f32-9042-d1f0648b2bbb.html)).

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/18/25,027,633/Ordinanza\\_n.14del6febbraio2021.pdf](https://www.regione.umbria.it/documents/18/25,027,633/Ordinanza_n.14del6febbraio2021.pdf)). The aim was to investigate whether the S mutations associated with these variants were detectable in the local sewage using the newly designed assays.

#### 2.4. 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 s, 58°C for 10 s, 72 °C for either 30 s (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  $\mu$ L of each primer (10  $\mu$ M; Table 1) and either 4  $\mu$ L of RNA in a final volume of 25  $\mu$ L (clinical samples) or 8  $\mu$ L of RNA in a final volume of 50  $\mu$ L (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  $\mu$ L of the first PCR product, 1  $\mu$ L of each primer (10  $\mu$ M) and the following conditions: initial denaturation at 98°C for 30 s, followed by 45 cycles of 98 °C for 10 s, 62 °C for 30 s, 72 °C for either 30 s (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. RNA from SARS-CoV-2 Wuhan prototype strain (provided by the Robert Koch-Institute under the RefBio project) and molecular grade water were included in each PCR run as positive and negative controls, respectively. 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 MW660619 to MW660743.

#### 2.5. 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, 20 G, 20I/501Y.V1, 20H/501Y.V2, 20 J/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.

The flowchart methodology of this research, created with Coggle (<https://coggle.it/flowcharts>), is available in Fig. 3.

### 3. Results

#### 3.1. Primer validation and assay specificity

All three newly designed PCR assays successfully amplified the UK variant and the Wuhan prototype strain used for primer assessment.

Sanger sequencing of the amplicon 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 (Fig. 2). As for the Wuhan sequence, 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.

#### 3.2. 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).

Twenty-two of the 24 SARS-CoV-2-positive nasopharyngeal swabs for which genome characterization was not available were successfully amplified by the long nested assay. The swabs which failed to be amplified were those with the lowest viral concentrations (below 71 g.c./ $\mu$ L of RNA). Of these 22 uncharacterized swabs, 14 harbored the nucleotide change responsible for the A222V amino acid substitution characteristic of the Spanish variant. In one swab (ID 18) this amino acid substitution was associated with two more amino acid changes, A262S and P272L, occurring in 1.17% and 0.87%, respectively, of all sequences deposited in GISAID. In another swab (ID 24) A222V was associated with a rare amino acid substitution, A411S (0.01% occurrence). No other variants of clinical interest were found in the uncharacterized samples, although other mutations were detected occasionally (e.g. S477N in swab IDs 2, 3, 13, 16, 23; S98F in swab 19; D215H in swab 26).

All clinical samples 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. The two swab samples (IDs 21 and 22) which could not be amplified by the long nested PCR, were also successfully amplified and sequenced with both of the short nested assays. Since no mutations were found in the two regions, however, these samples could not be reliably classified.

#### 3.3. Environmental samples

A total of 34 wastewater samples were tested with the three nested PCR assays. All PCR positive and negative controls gave the expected results. Sequencing results are summarized in Table 4.

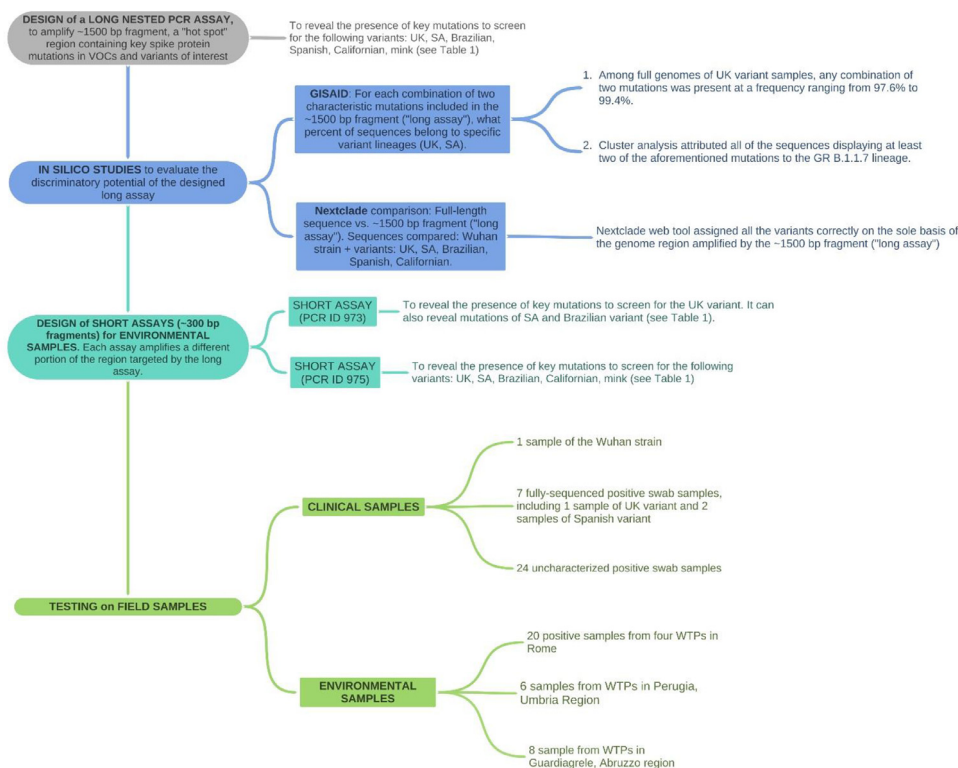


Fig. 3. Flowchart methodology of this research.

The long nested RT-PCR did not amplify any of the 20 SARS-CoV-2-positive urban sewage samples collected from WTPs in Rome. 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 and one of the six samples collected in Perugia, Umbria. The first two samples showed nucleotide changes corresponding to the amino acid substitution A222V characteristic of the Spanish variant, while the other carried the nucleotide changes responsible for amino acid substitutions characteristic of the Brazilian variant (a panel of five substitutions: 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 a nucleotide change corresponding to an amino acid substitution characteristic of the Brazilian variant (D138Y), in three of the six samples collected in Perugia. In these samples, the second short assay, PCR 975, showed the nucleotide changes corresponding to amino acid substitutions N501Y and E484K, a combination present in the Brazilian and the South African variant (see Table 1). Another sample from Perugia carried nucleotide changes resulting in amino acid substitutions N501Y and A570D, which, combined, are a key signature of the UK variant.

#### 4. 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, 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

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

Sample	Identification based on WGS	Mutation map <sup>1</sup>	GISAID blast	Nextclade
swab_1	–	A222V	GV_B.1.177	20E (EU1)
swab_2	–	S477N	GH_B.1.160	20A
swab_3	–	S477N	GH_B.1.160	20A
swab_4	–	A222V	GV_B.1.177	20E (EU1)
swab_5	–	A222V	GV_B.1.177	20E (EU1)
swab_6	–	A222V	GV_B.1.177	20E (EU1)
swab_7	–	A222V	GV_B.1.177	20E (EU1)
swab_8	–	A222V	GV_B.1.177	20E (EU1)
swab_9	–	A222V	GV_B.1.177	20E (EU1)
swab_10	–	A222V	GV_B.1.177	20E (EU1)
swab_11	–	A222V	GV_B.1.177	20E (EU1)
swab_12	–	A222V	GV_B.1.177	20E (EU1)
swab_13	–	S477N	GH_B.1.160	20A
swab_14	–	–	Not assignable <sup>2</sup>	20A
swab_15	–	–	Not assignable <sup>3</sup>	20A
swab_16	–	S477N	GH_B.1.160	20A
swab_17	–	A222V	GV_B.1.177	20E (EU1)
swab_18	–	A222V, A262S, P272L	GV_B.1.177 <sup>4</sup>	20E (EU1)
swab_19	–	S98F	G_B.1.221	20A
swab_20	–	A222V	GV_B.1.177	20E (EU1)
swab_21	–	n.d.	n.d.	n.d.
swab_22	–	n.d.	n.d.	n.d.
swab_23	–	S477N	GH_B.1.160	20A
swab_24	–	A222V, A411S	GV_B.1.177	20E (EU1)
swab_25	G B.1 / 20A	–	Not assignable <sup>5</sup>	19A
swab_26	G B.1 / 20A	D215H	Not assignable <sup>6</sup>	19A
swab_27	GR B.1.1.316 / 20B	–	Not assignable <sup>5</sup>	19A
swab_28	GR B.1.1.229 / 20B	–	Not assignable <sup>5</sup>	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

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

<sup>2</sup> 100% identity with sequences assigned to the following clades: G\_B.1.241; G\_B.1.78; G\_B.1.416; GR\_B.1.1.216; G\_B.1; GR\_B.1.1.297; GRB.1.1.8; GR\_B.1.1.105; GH\_B.1.2.

<sup>3</sup> 100% identity with sequences assigned to the following clades: GH\_B.1.2; G\_B.1; GR\_B.1.1.37; GR\_B.1.1.306.

<sup>4</sup> 100% identity with sequences assigned to the following clades: GV\_B.1.177 or GV\_B.1.177.2.

<sup>5</sup> 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.

<sup>6</sup> 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.

changes in the S protein of SARS-CoV-2, indicative of the presence of VOCs 20I/501Y.V1, 20H/501Y.V2 and 20 J/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 SARS-CoV-2 RNA and the discriminatory power of the amplified fragments were then tested on field samples, consisting of SARS-CoV-2 positive swabs, some of which had already been fully sequenced. All of the clinical samples were successfully amplified with the long PCR, with the exception of two samples having low viral concentrations. Among the fully-sequenced samples, the 20I/501Y.V1 and 20E.EU1 variants were correctly identified by the long assay. In the uncharacterized clinical samples, the only variant of interest identified was the Spanish variant. It was detected in 14 of 24 samples, 12 harboring only the A222V key amino acid substitution discriminating the variant, one also carrying the A262S and P272L amino acid substitutions, and one displaying the A411S as well. Since these samples were randomly selected from

among those collected between August 2020 and mid-January 2021 by the Experimental Zooprophyllactic Institute of Apulia and Basilicata for routine COVID-19 testing, this finding is consistent with a significant circulation of the Spanish variant in Italy, also supported by a recent preprint article suggesting that this variant has been rapidly spreading in the Lazio region of Italy (Bartolini et al., 2020). Moreover, judging by the number of Italian sequences submitted to GISAID, the 20E.EU1 variant has been circulating to a considerable extent in Italy (1792 sequences as of 23 February 2021), more than the UK variant (659 submitted sequences). Other variants of concern, like the Brazilian or the South African variants, were not available to us as field samples. Indeed, only one case of the South African variant had been 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](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 EPI\_ISL\_1,012,924), we verified that it harbors the spike amino acid substitutions D80A, D215G, L242del, L244del, K417N, E484K, and N501Y - all of which would have been detectable by our long PCR assay. Similarly, only three cases of the Brazilian variant had been identified and fully sequenced (GISAID EPI\_ISL\_875,566 to ID EPI\_ISL\_875,568) thus far in Italy. All these isolates carry the characteristic spike amino acid substitutions D138Y, R190S E484K, K417T, and N501Y included in our assay.

Once we had confirmed that the developed PCR assay could be successfully used to screen for variants in clinical samples, we



**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	variant likely present	result	mutations	variant likely present	result	mutations	variant likely present
<i>Archive samples</i>										
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	
<i>Ad-hoc samples from outbreak locations</i>										
3862	Guardiagrele	-			+	none		+	none	
3863	Guardiagrele	+	A222V	20E (EU1)	+	none		+	none	
3865	Guardiagrele	+	A222V	20E (EU1)	+	none		-		
3944	Perugia	-			+	D138Y	20 J/501Y.V3	+	E484K, N501Y	20 J/501Y.V3
3945	Perugia	-			+	D138Y	20 J/501Y.V3	+	E484K, N501Y	20 J/501Y.V3
3947	Perugia	+	D138Y, R190S, K417T, E484K, and N501Y	20 J/501Y.V3	+	D138Y	20 J/501Y.V3	+	E484K, N501Y	20 J/501Y.V3
3949	Perugia	-			+	E96G		+	N501Y, A570D	20I/501Y.V1

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.

went on to investigate whether this assay could also be used for the screening of SARS-CoV-2 variants in complex matrices, such as urban wastewaters. Since the amplification of long fragments poses specific challenges in such matrices, we also designed two short nested RT-PCR assays targeting portions of the region spanned by the long assay, to be used in case of long PCR failure. A number of studies have demonstrated the added value of regular surveillance of wastewaters in combination with other indicators for the management of the pandemic and a recent document by the European Commission highlights the importance of wastewater surveillance to allow screening among large population groups (European Commission, 2021). This can help public health authorities identify where more detailed analysis is needed and speed up the detection of variants as part of an increased genomic and epidemiological surveillance.

A few recent papers have demonstrated that environmental surveillance can be used to study SARS-CoV-2 diversity. 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.

A study by Graber and collaborators, published as a preprint article, describes a primer extension strategy to distinguish B.1.1.7 lineage in wastewater by allele-specific RT-qPCR assays (Graber et al., 2021).

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 sewage samples, including samples collected in areas where both the UK and the Brazilian variants had been documented by public health authorities.

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 20 J/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-ospedaliari\\_d50f6dce-9b78-443c-a0d8-0d23e16058a8.html](https://www.ansa.it/umbria/notizie/2021/02/05/rezza-in-umbria-variante-brasiliana-in-cluster-ospedaliari_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 amino acid substitution 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 amino acid substitution D138Y in the fragment obtained by PCR 973, and substitutions 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 variant. However, our GISAID in silico study showed the combination of the N501Y and A570D amino acid substitutions 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). The method can be used to discriminate the currently known SARS-CoV-2 variants, and may not be as useful with new variants possibly emerging in the future. Nevertheless, the recently described variants from Nigeria (G, B.1.525), and from India (GH, B.1.36), also have key mutations in the region targeted by the described assays (A67V, H69del, H70del, Y144del, E484K for G, B.1.525, and N440K for GH, B.1.36) (data not shown), confirming that this is a hot spot region for mutations.

To our knowledge, this is the first evidence of the presence of sequences harboring key mutations of 20I/501Y.V1 and 20 J/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.

## 5. Conclusions

- 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 be applied to environmental samples for the purposes of SARS-CoV-2 wastewater-based epidemiology, to ex-

plore the possible presence of VOCs or other variants of clinical interest, therefore highlighting geographical areas where an enhancement of clinical monitoring could be beneficial.

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

### Declaration of Competing Interest

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

### 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 program 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 Petrucci 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.

### Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.watres.2021.117104](https://doi.org/10.1016/j.watres.2021.117104).

### References

- Bartolini, B., Rueca, M., Ernesto, C., Gruber, M., Messina, F., Giombini, E., Ippolito, G., Capobianchi, M.R., Di Caro, A., 2020. The newly introduced SARS-CoV-2 variant A222V is rapidly spreading in Lazio region, Italy. *medRxiv* 11 (28), 20237016. doi:10.1101/2020.11.28.20237016.
- CDC, 2020. Interim: implications of the emerging SARS-CoV-2 Variant VOC 202012/01. Interim: implications of the Emerging SARS-CoV-2 Variant VOC 202012/01 | CDC, accessed on 12 January 2021.
- Conti, P., Caraffa, A., Gallenga, C.E., Kritas, S.K., Frydas, I., Younes, A., Di Emidio, P., Tetè, G., Pregliasco, F., Ronconi, G., 2020. The British variant of the new coronavirus-19 (Sars-Cov-2) should not create a vaccine problem. *J. Biol. Regul. Homeost. Agents* 35 (1). doi:10.23812/21-3-E, 10.23812/21-3-E. Advance online publication.
- Crits-Christoph, A., Kantor, R.S., Olm, M.R., Whitney, O.N., Al-Shayeb, B., Lou, Y.C., Flamholz, A., Kennedy, L.C., Greenwald, H., Hinkle, A., Hetzel, J., Spitzer, S., Koble, J., Tan, A., Hyde, F., Schroth, G., Kuersten, S., Banfield, J.F., Nelson, K.L., 2021. Genome sequencing of sewage detects regionally prevalent SARS-CoV-2 variants. *MBio* 12 (1), e02703–e02720. doi:10.1128/mBio.02703-20.
- ECDC, 2020. European Centre for Disease Prevention and Control, 2020. Risk related to spread of new SARS-CoV-2 variants of concern in the EU/EEA 29 December 2020.
- ECDC, 2021. European Centre for Disease prevention and control, 2021. Risk of spread of new SARS-CoV-2 variants of concern in the EU/EEA - first update (europa.eu) ECDC

- EUROPEAN COMMISSION, 2021. HERA Incubator: anticipating together the threat of COVID-19 variants [https://ec.europa.eu/info/sites/info/files/communication-hera-incubator-anticipating-threat-covid-19-variants\\_en.pdf](https://ec.europa.eu/info/sites/info/files/communication-hera-incubator-anticipating-threat-covid-19-variants_en.pdf)
- Faria, Nuno R., Mellan, Thomas A., Whittaker, Charles, Claro, Ingra M., Candido, Darlan da S., Mishra, Swapnil, Crispim, Myuki A. E., Sales, Flavia C., Hawryluk, Iwona, McCrone, John T., Hulsmit, Ruben J. G., Franco, Lucas A. M., Ramundo, Mariana S., de Jesus, Jaqueline G., Andrade, Pamela S., Coletti, Thais M., Ferreira, Giulia M., Silva, Camila A. M., Manuli, Erika R., Pereira, Rafael H. M., Peixoto, Pedro S., Kraemer, Moritz U., Gaburo Jr., Nelson, Camilo, Cecilia da C., Hoeltgebaum, Henrique, Souza, William M., Rocha, Esmeria C., de Souza, Leandro M., de Pinho, Mariana C., Araujo, Leonardo J. T., Malta, Frederico S. V., de Lima, Aline B., Silva, Joice do P., Zauli, Danielle A. G., Ferreira, Alessandro C. de S., Schnekenberg, Ricardo P., Laydon, Daniel J., Walker, Patrick G. T., Schlüter, Hannah M., dos Santos, Ana L. P., Vidal, Maria S., Del Caro, Valentina S., Filho, Rosinaldo M. F., dos Santos, Helem M., Aguiar, Renato S., Modena, José L. P., Nelson, Bruce, Hay, James A., Monod, Melodie, Miscouridou, Xenia, Coupland, Helen, Sonabend, Raphael, Vollmer, Michaela, Gandy, Axel, Suchard, Marc A., Bowden, Thomas A., Pond, Sergei L. K., Wu, Chieh-Hsi, Ratmann, Oliver, Ferguson, Neil M., Dye, Christopher, Loman, Nick J., Lemey, Philippe, Rambaut, Andrew, Fraij, Nelson A., Carvalho, Maria do P. S. S., Pybus, Oliver G., Flaxman, Seth, Bhatt, Samir, Sabino, Ester C., Genomics and epidemiology of a novel SARS-CoV-2 lineage in Manaus, Brazil. *medRxiv* 2021.02.26.21252554; doi:10.1101/2021.02.26.21252554
- 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. *The science of the total environment*, 743, 140444. doi:10.1016/j.scitotenv.2020.140444.
- Galloway, S.E., Paul, P., MacCannell, D.R., Johansson, M.A., Brooks, J.T., MacNeil, A., Slayton, R.B., Tong, S., Silk, B.J., Armstrong, G.L., Biggerstaff, M., Dugan, V.G., 2021. Emergence of SARS-CoV-2 B.1.1.7 Lineage - United States, December 29, 2020-January 12, 2021. *MMWR Morb. Mortal. Wkly. Rep.* 70 (3), 95–99. doi:10.15585/mmwr.mm7003e2.
- Graber T.E., Mercier E., D'Aouist P.M., Hoang H.D., Tian X., Tasneem S., Bhatnagar K., Delatolla R. An allele-specific primer extension assay to quantify the proportion of B.1.1.7-specific SARS-CoV-2 RNA in wastewater. 2021, medRxiv preprint doi:10.1101/2021.02.22.21252041
- Gu, H., Chen, Q., Yang, G., He, L., Fan, H., Deng, Y.Q., Wang, Y., Teng, Y., Zhao, Z., Cui, Y., Li, Y., Li, X.F., Li, J., Zhang, N.N., Yang, X., Chen, S., Guo, Y., Zhao, G., Wang, X., Luo, D.Y., Zhou, Y., 2020. Adaptation of SARS-CoV-2 in BALB/c mice for testing vaccine efficacy. *Science* 369 (6511), 1603–1607. doi:10.1126/science.abc4730.
- Hadfield, J., Megill, C., Bell, S.M., Huddleston, J., Potter, B., Callender, C., Sagulenko, P., Bedford, T., Neher, R.A., 2018. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 34 (23), 4121–4123. doi:10.1093/bioinformatics/bty407.
- Hammer, A.S., Quaaed, M.L., Rasmussen, T.B., Fonager, J., Rasmussen, M., Mundbjerg, K., Lohse, L., Strandbygaard, B., Jørgensen, C.S., Alfaro-Núñez, A., Rosentierne, M.W., Boklund, A., Halasa, T., Fomsgaard, A., Belsham, G.J., Bøtner, A., 2021. SARS-CoV-2 Transmission between Mink (Neovison vison) and Humans. *Denmark. Emerging infect. diseases* 27 (2), 547–551. doi:10.3201/eid2702.203794.
- Hart, O.E., Halden, R.U., 2020. Computational analysis of SARS-CoV-2/COVID-19 surveillance by wastewater-based epidemiology locally and globally: feasibility, economy, opportunities and challenges. *Sci. Total Environ.* 730, 138875. doi:10.1016/j.scitotenv.2020.138875.
- Hodcroft, E.B., Zuber, M., Nadeau, S., Crawford, K., Bloom, J.D., Veessler, D., Vaughan, T.G., Comas, I., Candelas, F.G., Stadler, T., Neher, R.A., 2020. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv: the preprint server for health sciences* doi:10.1101/2020.10.25.20219063, 2020.10.25.20219063.
- Iaconelli, M., Valdazo-González, B., Equestre, M., Ciccaglione, A.R., Marcantonio, C., Della Libera, S., La Rosa, G., 2017. Molecular characterization of human adenoviruses in urban wastewaters using next generation and Sanger sequencing. *Water Res.* 121, 240–247. doi:10.1016/j.watres.2017.05.039.
- Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., Hastie, K.M., Parker, M.D., Partridge, D.G., Evans, C.M., Freeman, T.M., de Silva, T.I., McDanal, C., Perez, L.G., Tang, H., ... Montefiori, D.C. Sheffield COVID-19 Genomics Group, 2020. Tracking Changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 Virus. *Cell* 182 (4), 812–827. doi:10.1016/j.cell.2020.06.043, e19.
- La Rosa, G., Mancini, P., Bonanno Ferraro, G., Veneri, C., Iaconelli, M., Bonadonna, L., Lucentini, L., Suffredini, E., 2021. SARS-CoV-2 has been circulating in northern Italy since december 2019: evidence from environmental monitoring. *Sci. Total Environ.* 750, 141711. doi:10.1016/j.scitotenv.2020.141711.
- 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 (5), 1284–1294. doi:10.1016/j.cell.2020.07.012, e9.
- Maggi, F., Novazzi, F., Genoni, A., Baj, A., Spezia, P.G., Focosi, D., Zago, C., Colombo, A., Cassani, G., Pasciuta, R., Tamborini, A., Rossi, A., Prestia, M., Capuano, R., Azzi, L., Donadini, A., Catanoso, G., Grossi, P.A., Maffioli, L., Bonelli, G., 2021. Imported SARS-CoV-2 Variant P.1 detected in traveler returning from Brazil to Italy. *Emerging Infect. Dis.* (4) 27. doi:10.3201/eid2704.210183.
- Martin, J., Klapka, 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 (10), 1144. doi:10.3390/v12101144.

- NERVTAG, 2020. New and Emerging Respiratory Virus threats advisory group. NERV-TAG meeting on SARS-CoV-2 variant under investigation VUI-202012/01, 18 December 2020.
- NERVTAG meeting minutes on SARS-CoV-2 variant under investigation. 18 December (khub.net), accessed on 12 January 2021.
- National Institute of Infectious Diseases (NIID), 2021. Brief report: new Variant Strain of SARS-CoV-2 Identified in Travelers from Brazil. PUBLISHED: 12 JANUARY 2021. brief report: new Variant Strain of SARS-CoV-2 Identified in Travelers from Brazil ([niid.go.jp](http://niid.go.jp)). Accessed on 16 January 2021.
- Public Health England (PHE) 2021. Investigation of novel SARS-CoV-2 variant: variant of Concern 202012/01, Technical briefing 5. 2020 [updated January 2021]. Available from: [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](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).
- Pillay, T.S., 2020. Gene of the month: the 2019-nCoV/SARS-CoV-2 novel coronavirus spike protein. *J. Clin. Pathol.* 73 (7), 366–369. doi:10.1136/jclinpath-2020-206658.
- Starr, T.N., Greaney, A.J., Hilton, S.K., Ellis, D., Crawford, K., Dingens, A.S., Navarro, M.J., Bowen, J.E., Tortorici, M.A., Walls, A.C., King, N.P., Veesler, D., Bloom, J.D., 2020. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on folding and ACE2 Binding. *Cell* 182 (5), 1295–1310. doi:10.1016/j.cell.2020.08.012, e20.
- Volz, E., Hill, V., McCrone, J.T., Price, A., Jorgensen, D., O'Toole, Á., Southgate, J., Johnson, R., Jackson, B., Nascimento, F.F., Rey, S.M., Nicholls, S.M., Colquhoun, R.M., da Silva Filipe, A., Shepherd, J., Pascall, D.J., Shah, R., Jesudason, N., Li, K., Jarrett, R., Connor, T.R., 2021. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. *Cell* 184 (1), 64–75. doi:10.1016/j.cell.2020.11.020, e11.
- Shu, Y., McCauley, J., 2017. GISAID: global initiative on sharing all influenza data - from vision to reality. *Eurosurveillance* 22 (13), 30494. doi:10.2807/1560-7917.ES.2017.22.13.30494.
- Suffredini, E., Iaconelli, M., Equestre, M., Valdazo-González, B., Ciccaglione, A.R., Marcantonio, C., Della Libera, S., Bignami, F., La Rosa, G., 2018. Genetic Diversity among Genogroup II Noroviruses and Progressive Emergence of GII.17 in Wastewaters in Italy (2011–2016) revealed by Next-Generation and Sanger Sequencing. *Food Environ. Virol.* 10 (2), 141–150. doi:10.1007/s12560-017-9328-y.
- WHO, 2020a. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. 11 March 2020
- WHO, 2020b. Emergencies preparedness, response. SARS-CoV-2 mink-associated variant strain - Denmark Disease Outbreak News.6 November 2020. WHO | SARS-CoV-2 mink-associated variant strain - Denmark
- WHO 2020c. Emergencies preparedness, response. SARS-CoV-2 Variant, WHO | SARS-CoV-2 Variants
- WHO, 2020d. SARS-CoV-2 Variant. United Kingdom of Great Britain and Northern Ireland -.
- WHO, 2021. Genomic sequencing of SARS-CoV-2 A guide to implementation for maximum impact on public health 8 January 2021. [9789240018440-eng.pdf](https://www.who.int/publications/i/item/9789240018440-eng-pdf)