Coronavirus Pandemic

A straightforward molecular strategy to retrospectively investigate the spread of SARS-CoV-2 VOC202012/01 B.1.1.7 variant

Gabriele Ibba[1,2], Rosangela Sau[1], Flavia Angioj[2], Marcello Abbondio[1], Salvatore Rubino[1,2], Sergio Uzzau[1,2]

[1] Department of Biomedical Sciences, Division of Microbiology and Virology, University of Sassari, 07100 Sassari, Italy
[2] Microbiologia e Virologia, Azienda Ospedaliera Universitaria, 07100 Sassari, Italy

Abstract

The spread of new SARS-CoV-2 variants represents a serious threat worldwide, thus rapid and cost-effective methods are required for their identification. Since November 2020, the TaqPath COVID-19 assay (Thermo Fisher Scientific) has been used to identify viral strains of the new lineage B.1.1.7, since it fails to detect the S-gene with the Δ69/70 deletion. Here, we proposed S-gene mutations screening with the Allplex SARS-CoV-2 assay (Seegene), another widely used RT-PCR test that targets Sarbecovirus E, SARS-CoV-2 N, and RdRp/S genes. Accordingly, we evaluated the S gene amplification curve pattern compared to those of the other genes. Exploiting an Allplex assay-generated dataset, we screened 663 RT-PCR digital records, including all SARS-CoV-2 respiratory samples tested in our laboratory with the Allplex assay between January 1st and February 25th, 2021. This approach enabled us to detect 64 samples with peculiar non-sigmoidal amplification curves. Sequencing a selected group of 4 RNA viral genomes demonstrated that those curves were associated with B.1.1.7 variant strains. Our results strongly suggest that B.1.1.7 variant spread has begun in this area at least since January and imply the potential of these analytical methods to track and characterize the spread of B.1.1.7 strains in those areas where Allplex SARS-CoV-2 datasets have been previously recorded.

Key words: SARS-CoV-2 B.1.1.7 variant; RT-PCR; molecular diagnostics.

Introduction

Multiple variants of the SARS-CoV-2 virus have been reported globally during the current COVID-19 pandemic. The spread of new viral strains depends on the emergence of specific genetic variants conferring a selective advantage, including enhanced transmissibility. Hence, these variants represent a major health concern, since their capability to spread more quickly and easily increases, in turn, the toll of hospitalizations and deaths.

In late September 2020, a new lineage called B.1.1.7 emerged in the UK. This strain, also designated as SARS-CoV-2 Variant of Concern (VOC) 202012/01 by Public Health England (PHE)*,* carries a large number of mutations along the whole genome [1]. Being more infectious than other circulating strains of SARS-CoV-2 [2], VOC 202021/01 quickly became the dominant circulating variant in the UK, leading to lockdowns and travel restrictions in this country. Two other lineages, B.1.351 and P.1, are under strict observation worldwide. B.1.351 has been identified initially in South Africa (December 2020) and to date a low

number of cases have been reported also in other countries in Africa, US, and in Europe [3]. P.1 emerged in Manaus (Brazil), in January 2021, and is currently being reported in many countries of South and North America, Europe and Japan [4]. The most significant variants identified in these strains are those located in the S gene, encoding for the Spike protein. Interestingly, one specific mutation, named N501Y, arose independently in all three variants. Since Spike protein is key to allow virion entry into host's cells, this mutation might improve the Spike protein's ability to bind to the cell receptor angiotensin-converting enzyme-2 (ACE-2), increasing the virus ability to infect hosts. Other mutations occurring within the S gene might affect the antigenic features of Spike, which is a key target of the host immune response. To this extent, since neutralizing antibodies are elicited by S protein epitopes, mostly S1 subunit RBD and NTD domains [5], S gene variants represent a serious threat also for the vaccinated or naturally immunized individuals.

B.1.351 and P.1, indeed, appeared to escape immunity in vaccinated individuals [6]. On the

contrary, several reports suggested that B.1.1.7 increased its dissemination capability while keeping its susceptibility to anti-S antibodies triggered by vaccination [7-8]. Polyclonal nature of humoral response warrants neutralization capability for both vaccinated and naturally immunized individuals. However, Spike epitopes binding by monoclonal therapeutic antibodies might be weakened in circulating strains carrying both N501Y and HV60/70 variants. Noteworthy, under the selective pressure of the steadily growing number of vaccinated individuals, B.1.1.7 might rapidly acquire further genetic variants that resist vaccine-elicited anti-S polyclonal antibodies. Hence, it is of paramount importance to tackle the dissemination of this B.1.1.7 variant with effective diagnostic strategies. Exploiting its S gene target failure (SGTF), a reverse transcription PCR (RT-PCR) assays (e.g., TaqPath COVID-19 RT-PCR assay, Thermo Fisher Scientific), has been successfully applied as a proxy to identify the B.1.1.7 lineage and to monitor its geotemporal dissemination. This approach, also known as Spike gene "drop out" enabled many laboratories worldwide to rapidly screen for the B.1.1.7 variant and to monitor its circulation. Interestingly, Borges *et al.* [9] also detected TaqPath S-gene positive B.1.1.7 strains having Cycle threshold (Ct) values for S gene > 5 units higher than the maximum Ct value obtained for the other two targets (N and ORF1ab). This profile, named

Figure 1. Retrospective geographical distribution of putative B.1.1.7 SARS-CoV-2 variants in north Sardinia.

"Spike gene target late amplification", was also consistently associated with the VOC 202012/01. However, the TaqPath methodology requires a two-step RT-PCR assay and might be applied in retrospective studies only when original samples or extracted RNA are still available. These observations prompted our research group to carefully evaluate the RT-PCR profile obtained with another assay targeting Spike gene, together with E, RdRP, and N genes (e.g., Allplex COVID-19 assay, Seegene). This assay was not reported to "fail" in detecting the VOC 202012/01. However, in addition to being unveiled by "target failure" or "target late amplification", mutations in the target sequence might moderately affect amplification efficiency, leading to minor changes of the amplification curve slope. In the past several months, our laboratory has analyzed a large amount of samples with the Allplex RT-PCR assay. Hence, we aimed to identify putative variants of the S gene by carefully evaluating the amplification plot of the S gene from the thermal cyclers digital records, followed by next-generation sequencing (NGS) of the corresponding archived genomic samples.

Methodology
RT-PCR digital records analyses

SARS-CoV-2 positive tests were selected from the whole dataset of Allplex SARS-CoV-2 assays (Seegene, Seoul, Korea) archived in our laboratory, as part of the diagnostic laboratory routine. RT-PCR were performed with the CFX96™ Real-time PCR Detection System-IVD (Bio-Rad Laboratories Inc, CA, USA). As for most thermal cyclers, the instrument software applies fluorophore-specific channel filters to remove background noise, and smoothing algorithms to obtain clearer amplification sigmoidal curves. Hence, the relative fluorescence from each sample was calculated with Bio-Rad CFX Manager V3.1 (Bio-Rad Laboratories) with the baseline-subtracted curve-fit setting and selecting the fluorescence from FAM (gene E) and Cal Red 610 (RdRP/S gene). The shape of the amplification curves and, specifically, of the phase-2 (or logarithmic phase) was carefully scrutinized to assess the efficiency of the S gene amplification. Specifically, the following parameters were chosen to screen samples of interest: a weak phase-2 (low slope) and/or a sequence of at least 2 inflection points indicating a curve increasing in slope each cycle and
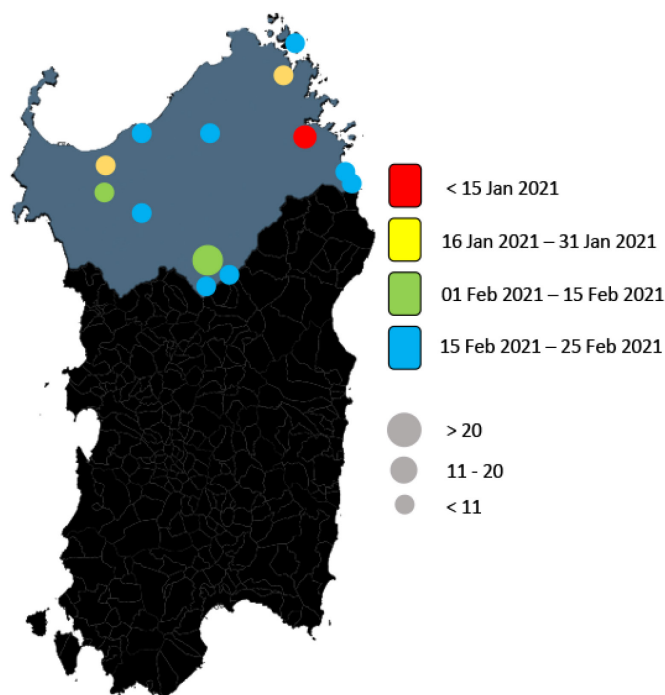
then decreasing in slope each cycle (Supplementary Figure 1).

Next generation sequencing libraries preparation

We implemented the CleanPlex SARS-CoV-2 Research and Surveillance Panels protocol (Paragon Genomics, Inc, Hayward, CA, USA) for target enrichment and library preparation. All the protocol steps, including RT with primers annealing, multiplex PCR (mPCR), digestion and second PCR, were performed according to the manufacturer's instructions. A group of four RNA viral samples was selected for whole genome sequencing. Specifically, mPCR was conducted using the 2-pool workflow, in order to obtain a higher coverage, preparing 2 mPCR reactions per sample as recommended. Unique i7 and i5 indexes from CleanPlex Dual-Indexed PCR Primers for Illumina, Set A (Paragon Genomics) were introduced via second PCR, adjusting to 24 cycles based on the number of amplicons (343) generated by mPCR. All the

purification steps were performed using CleanMag Magnetic Beads (Paragon Genomics). Quality control was performed with the 2100 Bioanalyzer instrument using Agilent High Sensitivity DNA Kit (Agilent Technologies, Santa Clara, CA, USA). Libraries showed an average peak at 272 bp. Libraries were then quantified using a Qubit Fluorometer with the dsDNA HS Assay Kit (Life Technologies, Carlsbad, CA, USA, now Thermo Fisher Scientific). After confirmation of the library quality, libraries were normalized to 10 nM and samples with unique index combinations were pooled in equimolar ratios to reach the recommended final concentration of 4 nM for sequencing. After a further quantification with Qubit dsDNA HS Assay Kit, pooled libraries were prepared following the Standard Normalization protocol on MiSeq System Denature and Dilute Libraries Guide (Illumina, San Diego, CA, USA), and finally denatured and diluted to 11 pM. Sequencing was carried-out on a MiSeq instrument (Illumina) with Reagent Kit v3, using 20 pM PhiX

Table 1. Substitutions and indels identified in four SARS-CoV-2 positive samples showing a "low-slope" amplification curve for the S gene. Bold text indicates variants defining the B.1.1.7 lineage. The genomic position affected by the mutation is indicated in round brackets.

PC1	#2	#4	#5	#32
(Val26=)	(Ser216=)	(Ser216=)	(Ser216=)	(Ser216=)
(Val60=)	(Phe924=)	(Phe924=)	(Phe924=)	(Phe924=)
(Phe924=)	**(Thr1001Ile) (ORF1ab)**	**(Thr1001Ile) (ORF1ab)**	**(Thr1001Ile) (ORF1ab)**	**(Thr1001Ile) (ORF1ab)**
(Ser2625=)	**(Ala1708Asp) (ORF1ab)**	(Asn1662Ser)	(Asn1662Ser)	**(Ala1708Asp) (ORF1ab)**
(Ala3623Ser)	(Phe1907=)	**(Ala1708Asp) (ORF1ab)**	**(Ala1708Asp) (ORF1ab)**	(Phe1907=)
(Pro4715Leu)	**(Ile2230Thr) (ORF1ab)**	(Phe1907=)	(Phe1907=)	**(Ile2230Thr) (ORF1ab)**
(Leu5Phe)	**(Ser3675_Phe3677del) (ORF1ab)**	(Asn2147=)	(Asn2147=)	**(Ser3675_Phe3677del) (ORF1ab)**
(Ala222Val)	(Pro4715Leu)	**(Ile2230Thr) (ORF1ab)**	**(Ile2230Thr) (ORF1ab)**	(Leu3829Phe)
(Asp614Gly)	(Pro4804=)	(Lys3353Arg)	(Lys3353Arg)	(Pro4715Leu)
(Gln675His)	(His5005=)	**(Ser3675_Phe3677del) (ORF1ab)**	**(Ser3675_Phe3677del) (ORF1ab)**	(Pro4804=)
(Arg682=)	(Thr5304=)	(Pro4715Leu)	(Pro4715Leu)	(His5005=)
(Ala220Val)	**(His69_Val70del) (S)**	(Pro4804=)	(Pro4804=)	(Thr5304=)
(Ser235Phe)	(Tyr145del)	(His5005=)	(His5005=)	(Asp6824=)
	(Asn501Tyr) (S)	(Thr5304=)	(Thr5304=)	**(His69_Val70del) (S)**
	(Ala570Asp) (S)	(Leu6703Met)	(Leu6703Met)	(Tyr145del)
	(Asp614Gly)	**(His69_Val70del) (S)**	**(His69_Val70del) (S)**	**(Asn501Tyr) (S)**
	(Pro631Ser)	(Tyr145del)	(Tyr145del)	**(Ala570Asp) (S)**
	(Pro681His) (S)	**(Asn501Tyr) (S)**	**(Asn501Tyr) (S)**	(Asp614Gly)
	(Thr716Ile) (S)	**(Ala570Asp) (S)**	**(Ala570Asp) (S)**	(Pro631Ser)
	(Ser982Ala) (S)	(Asp614Gly)	(Asp614Gly)	**(Pro681His) (S)**
	(Asp1118His) (S)	(Arg634His)	**(Pro681His) (S)**	**(Thr716Ile) (S)**
	(Gln27*) (ORF8)	**(Pro681His) (S)**	**(Thr716Ile) (S)**	**(Ser982Ala) (S)**
	(Arg52Ile) (ORF8)	**(Thr716Ile) (S)**	**(Ser982Ala) (S)**	**(Asp1118His) (S)**
	(Tyr73Cys) (ORF8)	**(Ser982Ala) (S)**	**(Asp1118His) (S)**	(Ile47=)
	(Met1?)	**(Asp1118His) (S)**	**(Gln27*) (ORF8)**	(Tyr31=)
	(Asp3His)	**(Gln27*) (ORF8)**	**(Arg52Ile) (ORF8)**	**(Gln27*) (ORF8)**
	(Asp3Val)	**(Arg52Ile) (ORF8)**	**(Tyr73Cys) (ORF8)**	**(Arg52Ile) (ORF8)**
	(Asp3Glu)	**(Tyr73Cys) (ORF8)**	(Met1?)	(Lys68*)
	(Arg203Lys)	(Met1?)	(Asp3His)	**(Tyr73Cys) (ORF8)**
	(Arg203=)	(Asp3His)	(Asp3Val)	(Met1?)
	(Gly204Arg)	(Asp3Val)	(Asp3Glu)	(Asp3His)
	(Ser235Phe)	(Asp3Glu)	(Arg14Cys)	(Asp3Val)
		(Arg203Lys)	(Arg203Lys)	(Asp3Glu)
		(Arg203=)	(Arg203=)	(Arg203Lys)
		(Gly204Arg)	(Gly204Arg)	(Arg203=)
		(Ser253Phe)	(Ser235Phe)	(Gly204Arg)
				(Ser235Phe)

control spike-in of 5% for low-diversity libraries, setting 2 x 150 cycles, and generating paired-end reads.

Genomic analyses and variants assessments

NGS raw data (FASTQ files) were generated from MiSeq Local Run Manager (Illumina) and uploaded on the SOPHiA DDM platform (SOPHiA Genetics, Lausanne, Switzerland) for external quality check, variant call review, and determination of the consensus genome. Finally, SARS-CoV-2 lineages were determined using Pangolin (https://github.com/cov-lineages/pangolin).

Results and Discussion

The Microbiology and Virology Division, University Hospital of Sassari, Italy, serves as a regional reference center and has tested approximately 178,151 samples since March 2020. Between January 1st 2021 and February 25th 2021, 34,305 samples were analyzed by the Allplex COVID-19 assay, Seegene; 5,555 of these determinations were resolved as positive. Since most swab samples collected in the area of North Sardinia were analysed in our diagnostic laboratory, we consider our digitally recorded dataset of Allplex COVID-19 assay as large and representative for the whole sampling area. Given the spread of VOC 202012/01 (B.1.1.7 lineage), our diagnostic laboratory was prompted to actively test for this variant on all SARS-CoV-2 positive swab samples. To this end, while waiting for development RT-PCR assays enabling the direct detection of specific variants, according to new and specific probe/primers design, we carefully analyzed the shape of S gene amplification curve as a potential proxy for S gene variants, as described in the methods section. Quite fortuitously, we observed a cluster of 4 samples with a weak phase-2 (low slope) and inflection in a point midway up the phase-2 of the curve (Supplementary Figure 1). To assess whether the amplification pattern was due to S gene mutations affecting the efficiency of the RT-PCR test, the 4 samples were subjected to RNA extraction and whole genome sequencing, as described in the methods section. According to the Pangolin application, the analyses of NGS data enabled us to assign these viral genomes to SARS-CoV-2 lineages, confirming the presence of the B.1.1.7 variant. In further analysis of the whole viral genomes, the lineage has also been confirmed by the identification of 15 of 17 variant defining mutations (Table 1). In addition to the modifications known to be B.1.1.7 lineage-specific, we reported the substitution and indels signatures (Table 1) characterizing each sample. We then analyzed,

retrospectively, all digital records obtained from the thermal cyclers where the Allplex COVID-19 assay was performed. Among the positive tests recorded since January 1st, 64 were detected as showing curve profiles similar to those described above and in Supplementary Figure 1. We finally investigated the geotemporal distribution of the samples identified as B.1.1.7, according to the site of sampling and/or recorded permanent place of residence. The data obtained, while limited and without further epidemiological characterization, enable a preliminary picture of B.1.1.7 spread in the area of North Sardinia, Italy. Noteworthy, on February 25th 2021, while we were writing this paper, Seegene Inc. published a technical note assessing that "*investigations with full-genome sequenced around 400 known samples (300 with normal curve and 100 with abnormal curve), tested by Allplex SARS-CoV-2 Assay, showed that in the presence of B.1.1.7 lineage, the shape of the raw curve of the CalRed610 channel (corresponding to RdRp/S genes) differs from the expected shape of the curve, allowing to a pre-screening of this lineage...*".

In conclusion, given the wide availability of the Allpex SARS-CoV-2 assay in many countries, the retrospective analyses of digitally recorded dataset and the detection of the S gene amplification curve profiles, as reported in this paper, might contribute in a cost-effective manner to enable large retrospective surveys of B.1.1.7 geotemporal dissemination in most countries worldwide.

Authors' contributions

GI contributed to the study design and to genomes sequencing, conducted RT-PCR test, supervised data analyses, analyzed digital records, and wrote the manuscript. RS and MA conducted viral genomes sequencing and wrote the manuscript. FA contributed to the study design and to genomes sequencing, conducted RT-PCR test, analyzed digital records, and wrote the manuscript. SR contributed to the study design. SU conceived the study, supervised the study progress and data analyses, and wrote the manuscript. All authors approved the submitted version.

References

1. Public Health England (PHE) (2020) Investigation of novel SARS-COV-2 variant: Variant of Concern 202012/01: Technical briefing document on novel SARS-CoV-2 variant.

Available: https://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201. Accessed: 20 February 2021.

2. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, Hinsley WR, Laydon DJ, Dabrera G, O'Toole A, Amato R, Ragonnet-Cronin M, Jorgensen D, Myers R, Hill V, Boyd O, Loman NJ, McCrone JT, Goncalves S, Kwiatkowski DP, The COVID-19 Genomic UK (COG-UK) consortium, Flaxman S, Ratmann O, Bhatt S, Hopkins S, Gandy A, Rambaut A, Ferguson NM (2020) Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data - Preprint 20201230.

3. Tegally H, Wilkinson E, Giovanetti M, Iranzadeh A, Fonseca V, Giandhari J, Doolabh D, Pillay S, San EJ, Msomi N, Mlisana K, von Gottberg A, Walaza S, Allam M, Ismail A, Mohale T, Glass AJ, Engelbrecht S, Zyl GV, Preiser W, Petruccione F, Sigal A, Hardie D, Marais G, Hsiao M, Korsman S, Davies MA, Tyers L, Mudau I, York D, Maslo C, Goedhals D, Abrahams S, Laguda-Akingba O, Alcantara LCJ, Kosakovsky Pond SL, Weaver S, Martin D, Lessells RJ, Bhiman JN, Williamson C, de Oliveira T (2020) Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa - Preprint 20201221.

4. Faria NR, Morales Claro I, Candido D, Moyses Franco LA, Andrade PS, Coletti TM, Silva CAM, Sales FC, Manuli ER, Aguiar RS, Gaburo N, da C. Camilo C, Fraiji NA, Crispim MAE, do Perpetuo S.S. Carvalho M, Rambaut A, Loman N, Pybus OG, Sabino EC, on behalf of CADDE Genomic Network (2021) Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. Virological. Available: https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586. Accessed: 20 February 2021.

5. Liu L, Wang P, Nair MS, Yu J, Rapp M, Wang Q, Luo Y, Chan JFW, Sahi V, Figueroa A, Guo XV, Cerutti G, Bimela J, Gorman J, Zhou T, Chen Z, Yuen KY, Kwong PD, Sodroski JG, Yin MT, Sheng Z, Huang Y, Shapiro L, Ho DD (2020) Potent Neutralizing Antibodies against Multiple Epitopes on SARS-CoV-2 Spike. Nature 584: 450-456.

6. Garcia-Beltran WF, Lam EC, Denis KS, Nitido AD, Garcia ZH, Hauser BM, Feldman J, Pavlovic MN, Gregory DJ, Poznansky MC, Sigal A, Schmidt AG, Iafrate AJ, Naranbhai V, Balazs AB (2021) Circulating SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. Preprints 20210214.

7. Collier DA, De Marco A, Ferreira I, Meng B, Datir R, Walls AC, Kemp SA, Bassi J, Pinto D, Silacci Fregni C, Bianchi S, Tortorici MA, Bowen J, Culap K, Jaconi S, Cameroni E, Snell G, Pizzuto MS, Franzetti Pellanda A, Garzoni C, Riva A, The CITIID-NIHR BioResource COVID-19 Collaboration, Elemer A, Kingston N, Graves B, McCoy LE, Smith KGC, Bradley JR, Temperton N, Ceron-Gutierrez L, Barcenas Morales G, The COVID-19 Genomics UK (COG-UK) consortium, Harvey W, Virgin HW, Lanzavecchia A, Piccoli L, Doffinger R, Wills M, Veesler D, Corti D, Gupta RK (2021) SARS-CoV-2 B.1.1.7 sensitivity to mRNA vaccine-elicited, convalescent and monoclonal antibodies. Preprint 20210119.

8. Xie X, Yang L, Liu J, Zhang X, Zou J, Fontes-Garfias CR, Xia H, Swanson KA, Cutler M, Cooper D, Menachery VD, Weaver S, Dormitzer PR, Shi PY (2021) Neutralization of SARS-CoV-2 Spike 69/70 Deletion, E484K and N501Y Variants by BNT162b2 Vaccine-Elicited Sera. Preprints 20210127.

9. Borges V, Sousa C, Menezes, L, Gonçalves AM, Picão M, Almeida JP, Vieita M, Santos R, Silva AR, Costa M, Carneiro L, Isidro J, Duarte S, Vieira L, Guiomar R, Silva S, Nunes B, Gomes JP (2021) Tracking SARS-CoV-2 VOC 202012/01 (lineage B.1.1.7) dissemination in Portugal: insights from nationwide RT-PCR Spike gene drop out data. Virological. Available: https://virological.org/t/tracking-sars-cov-2-voc-202012-01-lineage-b-1-1-7-dissemination-in-portugal-insights-from-nationwide-rt-pcr-spike-gene-drop-out-data/600. Accessed: 20 February 2021.

Corresponding author
Professor Sergio Uzzau, MD. PhD
Department of Biomedical Sciences
University of Sassari
Viale San Pietro 43b
07100 Sassari
Email: uzzau@uniss.it

Conflict of interests: No conflict of interests is declared.

Annex – Supplementary Items

Supplementary Figure 1. Comparative example of linear and "low-slope" amplification curves in two positive SARS-CoV-2 samples. E-Gene (blue line) and RdRp/S-gene (red line) amplification curves describe the trend differences between a native SARS-CoV-2 genome (left) and putative B.1.1.7 lineage (right).