

Single base mutations in the nucleocapsid gene of SARS-CoV-2 affects amplification efficiency of sequence variants and may lead to assay failure



Nathaniel Storey^{a,*}, Julianne R Brown^a, Rui P A Pereira^b, Denise M O'Sullivan^b, Jim F Huggett^{b,c}, Rachel Williams^d, Judith Breuer^e, Kathryn A Harris^a

^a Microbiology, Virology and Infection Prevention and Control, Great Ormond Street Hospital for Children NHS Foundation Trust, London, United Kingdom

^b National Measurement Laboratory at LGC, Teddington, United Kingdom

^c School of Biosciences and Medicine, Faculty of Health and Medical Sciences, University of Surrey, United Kingdom

^d Division of Infection and Immunity, University College London, London, United Kingdom

^e Department of Infection, Immunity and Inflammation, UCL Great Ormond Street Institute of Child Health, University College London, London, United Kingdom

ARTICLE INFO

Keywords:

SARS-CoV-2
Nucleocapsid
Mutations
PCR
Failure
Silico

ABSTRACT

Reverse transcriptase quantitative PCR (RT-qPCR) is the main diagnostic assay used to detect SARS-CoV-2 RNA in respiratory samples. RT-qPCR is performed by specifically targeting the viral genome using complementary oligonucleotides called primers and probes. This approach relies on prior knowledge of the genetic sequence of the target. Viral genetic variants with changes to the primer/probe binding region may reduce the performance of PCR assays and have the potential to cause assay failure. In this work we demonstrate how two single nucleotide variants (SNVs) altered the amplification curve of a diagnostic PCR targeting the Nucleocapsid (N) gene and illustrate how threshold setting can lead to false-negative results even where the variant sequence is amplified. We also describe how *in silico* analysis of SARS-CoV-2 genome sequences available in the COVID-19 Genomics UK Consortium (COG-UK) and GISAID databases was performed to predict the impact of sequence variation on the performance of 22 published PCR assays. The vast majority of published primer and probe sequences contain sequence mismatches with at least one SARS-CoV-2 lineage. We recommend that visual observation of amplification curves is included as part of laboratory quality procedures, even in high throughput settings where thresholds are set automatically and that *in silico* analysis is used to monitor the potential impact of new variants on established assays. Ideally comprehensive *in silico* analysis should be applied to guide selection of highly conserved genomic regions to target with future SARS-CoV-2 PCR assays.

1. Introduction

The SARS-CoV-2 pandemic required rapid implementation of diagnostic testing. The only method capable of meeting this need was reverse transcriptase quantitative PCR (RT-qPCR) which worked by detecting SARS-CoV-2 RNA. Reliable RT-qPCR assays played an important role in guiding patient management and limiting the onward transmission of SARS-CoV-2 [1]. Diagnostic laboratories across the world are currently using a wide variety of different commercial and locally developed RT-qPCR assays that target a range of SARS-CoV-2 genes.

SARS-CoV-2 genetic variants exist that have the potential to lead to phenotypic changes manifesting as differences in viral infectivity, burden, immunogenicity and tropism during the course of infection [2–8]. Genetic variants form a natural part of infection as SARS-CoV-2, a

zoonotic virus, evolves with its new human host and can become established in the population. Variants of concern (VOC) are variants of SARS-CoV-2 that may be more infectious, cause more severe disease or lead to vaccine or immunological escape [9]. Consequently accurate and rapid detection of VOCs is important to manage a pandemic. RT-qPCR is performed by using short DNA molecules (primers and probes) that complement and directly bind the SARS-CoV-2 genome, therefore genetic variants may affect assay performance. An emerging sequence variant could escape detection in molecular assays. This may be an infrequent event, but if it occurs the impact could be very significant. Individuals could be incorrectly told they are not infected, potentially increasing the spread of a new variant. This is even more challenging during the SARS-CoV-2 pandemic, as new sequence variants arise frequently [10,11]. Consequently it is important that diagnostic developers and those applying RT-qPCR for SARS-CoV-2 diagnosis are aware of the

* Corresponding author.

E-mail addresses: nathaniel.storey@gosh.nhs.uk (N. Storey), Julianne.Brown@gosh.nhs.uk (J.R. Brown), Rui.Pereira@LGCgroup.com (R.P.A. Pereira), denise.osullivan@lgcgroup.com (D.M. O'Sullivan), Jim.Huggett@lgcgroup.com (J.F. Huggett), rachel.williams@ucl.ac.uk (R. Williams), j.breuer@ucl.ac.uk (J. Breuer), Kathryn.Harris@gosh.nhs.uk (K.A. Harris).

<https://doi.org/10.1016/j.jcvp.2021.100037>

Received 16 April 2021; Received in revised form 11 August 2021; Accepted 13 August 2021

2667-0380/© 2021 The Authors. 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/>)

potential impact of VOCs on their assays both in terms of surveillance and RT-qPCR output.

In this report we investigate how sequencing can be used locally to determine the sequence lineages of 332 consecutive samples that were SARS-CoV-2 PCR positive in our laboratory between 26th March 2020 and 17th December 2020. We also explored the effect two single-nucleotide variants (SNVs) have on the performance of an RT-qPCR diagnostic PCR assay and if different analysis parameters (such as threshold setting or applying Cq cut offs) can lead to different results. We then use this sequence data to perform *in silico* analysis of primer and probe sequences from 22 published SARS-CoV-2 PCR assays to predict potential impact of VOCs on these approaches and make recommendations for a proactive and standardised approach to assay design and monitoring.

2. Materials and methods

2.1. Samples

Nasopharyngeal swabs and aspirates, collected from patients and healthcare workers tested for SARS-CoV-2 by reverse-transcriptase quantitative PCR (RT-qPCR) between 26th March 2020 and 17th December 2020. These samples were tested as part of the routine diagnostic service at Great Ormond Street Hospital NHS Foundation Trust (GOSH) and included both symptomatic and asymptomatic individuals. Samples were selected for sequencing if they were the first positive sample from an individual and had a sufficient viral load for successful sequencing determined by a quantification cycle (Cq, also described as cycle threshold, Ct) lower than 34. A total of 332 samples were sequenced.

2.2. RT-qPCR and sequencing

Dry, flocked swabs were re-suspended with 600 or 1,200 µl nuclease-free water (for single or double swabs, respectively). Total nucleic acid was purified from 250 µl swab suspension fluid using the Hamilton STAR robotic liquid handler and Mag-Bind® Viral DNA/RNA kit (Omega Biotek). RNA was eluted in 100 µl elution buffer. Each 250 µl specimen was spiked with 1.1 µl Phocine Distemper Virus (PDV) cell culture isolate (cultured in vero cells) before the nucleic acid extraction to act as an internal positive control. PDV is an established qualitative control used in all clinical RT-PCR assays at GOSH to control for PCR inhibition or nucleic acid extraction failure. Negative extraction controls were included alongside the specimen extractions that contained water in place of swab suspension fluid; these were also spiked with PDV [12].

RT-qPCR reactions were performed in 25 µl reaction volumes with 7.5 µl of RNA, 1X One Step PrimeScript III RT-PCR mastermix (Takara); 0.4 µM forward primer (N-gene Taq1); 0.6 µM reverse primer (N-gene Taq 2); 0.3 µM probe (N-gene Taq probe); 0.125 µM forward primer (PDV-F); 0.125 µM reverse primer (PDV-R); 0.125 µM probe (PDV-probe) per reaction, primer and probe sequences previously published and provided in the appendix [12,13]. Thermal cycling was performed on a QuantStudio5 thermocycler (ThermoFisher) using manufacturer (Takara) recommended fast cycling conditions and 45 cycles. A negative control (water) and a positive control were included in each run. This positive control was an RNA transcript of the SARS-CoV-2 nucleocapsid (N) gene synthesised by *in vitro* transcription of a linearised plasmid containing the entire N gene. Full length transcripts were confirmed using the 2100 Bioanalyzer system (Agilent) and the RNA concentration was estimated using the Qubit Fluorometer (ThermoFisher) and used at a concentration of 100 copies / µl [12]. Sequencing was performed in the UCL Pathogen Genomics Unit (PGU) using either Oxford Nanopore Technologies (ONT) with 2 kb primers [14], ONT with V3 primers [15] or ARTIC Illumina sequencing protocol V.5 [16] as part of the COVID-19 Genomics UK Consortium (COG-UK) [17] and sequences were assigned relevant COG-UK identifiers.

The cycling programmes were run and the data collected using the QuantStudio 5 Dx Software v1.0.2. Cq values were obtained using a

Table 1

List of the PCR assays with published primer and probe sequences used for *in silico* analysis. Sequences are provided in appendix and are also available at: <https://www.who.int/publications/m/item/molecular-assays-to-diagnose-covid-19-summary-table-of-available-protocol>.

Source	Target gene (s)
China CDC (Wang et al. [23])	Orf1ab; N
Charite (Corman et al. [24])	RdRp; E; N
Hong Kong (Chu et al. [25])	Orf1b-nsP14; N
Thailand	N
Japan	Orf1a; S; N
USA CDC	N (N1,N2, N3)
Institut Pasteur	RdRp (IP2,IP4); E
To et al. [26]	RdRp
Chan et al. [27]	RdRp; S
Lu et al. [28]	Orf1a
Grant et al. [13]	N

manual threshold method (threshold set mid-point through the exponential phase of amplification, as per manufacturer guidance and standard practice in clinical PCR assays at GOSH). A positive result was defined as amplification detected above the threshold within 45 cycles. Images of the amplification curves were taken directly from the analysis software.

2.3. In silico primer probe analysis

Primer and probe sequences from 22 published assays and 11 different sources were analysed (Table 1), sequences are in the appendix. Twenty one of these assays are published on the WHO website [1] and the remaining one is a locally designed assay that has been published [12,13]. SARS-CoV-2 sequence regions on the leading strand, corresponding to primer, probe or gene targets were identified in a SARS-CoV-2 reference sequence Wuhan-Hu-1 (Accession: NC_045512.2). Sequences corresponding to these regions with an additional 75 bases and 15 bases from the 5' and 3' ends respectively, were extracted *in silico* from SARS-CoV-2 consensus sequence data, sequenced in the UCL PGU using the ARTIC protocol with V3 primers [15] and assembled in to consensus sequences following the ARTIC-nCoV-bioinformatics SOP-v1.1.0 protocol [18]. These COG-UK sequences correspond to samples processed by GOSH between March and December 2020, alongside 8 selected variant of concern sequences provided by GISAID [19] (COG-UK, GISAID ID's and lineage information available in Supplementary data, Table S1 and S2). Sequence extraction was performed with Samtools faidx version 1.9 [20] and multiple sequence alignment performed using Clustal Omega version 1.2.4 [21] for identification of SNVs, insertions and deletions. Lineage of SARS-CoV-2 sequences were determined with Pangolin 2.1.7 [22] and lineages version 2020-05-19.

3. Results

3.1. Impact of SNV on qPCR assay

329 clinical samples that were positive for SARS-CoV-2 at our centre between 26th March 2020 and 17th December 2020 were sequenced and assigned to a lineage; three samples were not assigned a lineage due to high number of ambiguous bases (>50% N-content). Lineages assigned to more than one sample are shown in Table 2.

Thirty samples were the B.1.1.7 lineage, a UK variant of concern (VOC 202,012/01) and 11 samples were the B.1.177.19 lineage. Both were shown by *in silico* analysis to have sequence variation in the primer and probe binding sites of the N-gene assay used in our laboratory [12,13]. In both cases this was a single nucleotide variant (SNV). Fig. 1b shows the amplification plots for six samples, two that were B.1.1.7 lineage, two that were B.1.177.19 lineage and one each of the wild-

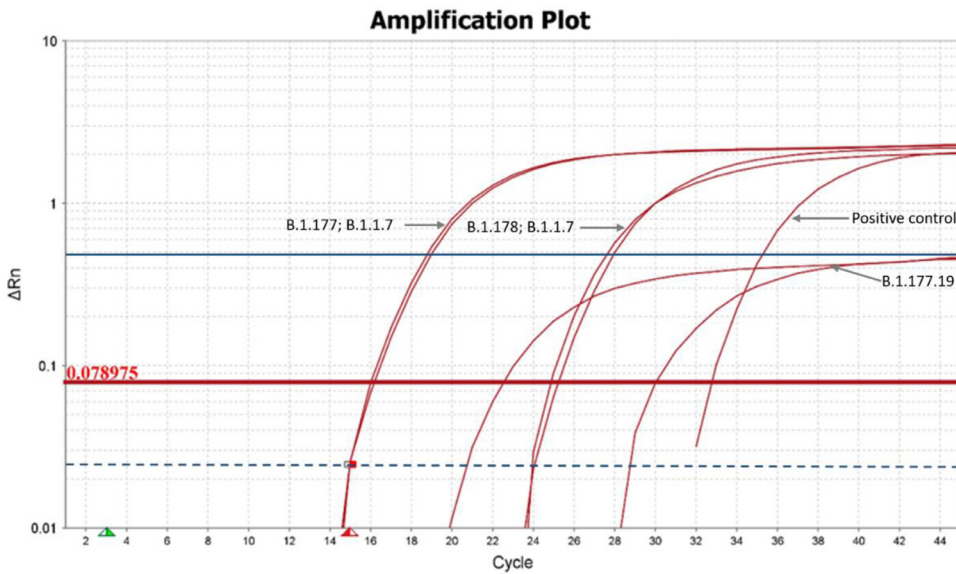


Fig. 1. a. Alignment of the DNA sequence of 3 samples (LOND-12E8582: B.1.1.7, LOND-12E8670: B.1.177.19, LOND-12E85EC: B.1.177.19) alongside primer and probe sequences of the N-gene assay in operation at GOSH. b: Amplification plot for N-gene assay (Grant et al.)^{2,3} performed on the Quantstudio 5 (ThermoFisher) using One Step PrimeScript III RT-PCR mastermix (Takara). Plot shows positive control (N-gene RNA transcript 100 copies/ μ l) and 6 clinical samples (1 lineage B.1.177, 1 lineage B.1.178, 2 lineage B.1.177 and 2 lineage B.1.177.19). Fluorescence threshold set manually at the mid-point of the exponential phase of amplification for the positive control (solid red line). The dashed blue line is an alternative threshold set manually at the mid-point of the exponential phase of amplification for the B.1.177.19 samples. The solid blue line is an alternative threshold set at a higher fluorescence value. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

■ SARS-CoV-2

Table 2
Table of SARS-CoV-2 lineages identified in more than one sample with and number of sequences assigned to each lineage.

Lineage	Samples assigned
B	8
B.1	63
B.1.1	100
B.1.1.1	8
B.1.1.10	4
B.1.1.161	3
B.1.1.164	3
B.1.1.37	7
B.1.1.4	3
B.1.1.7	30
B.1.1.89	8
B.1.177*	10
B.1.177.19	11
B.1.258	3
B.1.36.17	2
B.1.367	4
B.1.5	29
B.1.98	3
B.3	2
B.40	12
None	3

*VOC 202,012/01, 20I/501Y.V1.

type lineages B.1.177 and B.1.178. The first SNV (C28932T), in lineage B.1.1.7 (LOND-12E8582, Fig. 1a), was at the 5' end of the forward primer and had no effect on the assay performance compared to the wild-type lineages. The second SNV (G29027T), in lineage B.1.177.19 (LOND-12E8670 and LOND-12E85EC, Fig. 1a), was towards the 3' end of the probe sequence, impacting assay performance by reducing the amplitude of the amplification curve. All 11 samples of lineage B.1.177.19 exhibited this reduced amplitude. The fluorescence threshold was set manually using the standard laboratory procedure, which is mid-point through the exponential phase of amplification for the positive control which contains the correct sequence (solid red line in Fig. 1). However, using the same Cq threshold for the variant samples measures a different part of the amplification curve, giving an artificially high Cq value. The dashed blue line in Fig. 1b is a more appropriate threshold for these variant samples. If the threshold was set higher (solid blue line in Fig. 1b), the variant was measured as negative even where viral burden was high.

3.2. In silico analysis reveals multiple SNVs in published primer probe sequences

Sequence data from 332 samples that tested positive at GOSH during the period 26th March 2020 and 17th December 2020 was analysed for sequence variation in primer and probe binding regions of the 22 published PCR assays listed in Table 1. Eighteen of the 22 assays had at least one SNV in one or more of the primer or probe sequences in $\geq 1\%$ of the SARS-CoV-2 sequences analysed (Fig. 2). Twenty-six individual primer/probe sequences had at least one SNV with sequences from one or more of the lineages included in the analysis (Table 3).

Six primer/probe sequences had a least one SNV in every sequence analysed: To et al. [26] (RdRp gen, R primer), Japan (N gene, R2 primer), Charite (Corman et al. [24]) (RdRp gene, P1 probe and R1 primer), Chan et al. [27] (RdRp gene, F primer and R primer). One primer sequence (China CDC (Wang et al. [23]) (N gene, F primer)) had at least one SNV in 57% of sequences analysed.

4. Discussion

Viral variant genetic changes can result in primer or probe sequences that no longer perfectly complement the genetic region they target. Sequence changes to the primer and probe binding regions can have no, marginal or a catastrophic effect on PCR performance. Sequence mismatches in the five prime region of a primer or probe rarely affect assay performance whereas those in the three prime region can have a greater impact. We demonstrated how two different SNVs can impact on the performance of a real-time PCR assay targeting the N gene of SARS-CoV-2 (Fig. 1). These outcomes can be predicted from *in silico* analysis as the SNV in the probe may impact on binding or hydrolysis (Fig. 1a: aligned assay to genome), as has been previously documented for SARS-CoV-2 [29-31]. The result was a different shaped amplification curve which also had reduced amplitude when the B.1.177.19 samples are analysed.

Observation of the new curve allows an experienced user to set an alternative threshold for a better assessment of the Cq value (dotted blue line Fig. 1b), crucially the optimal threshold differs and the two curves should not be compared using the same threshold. However, automatic threshold setting, which is commonly applied in commercial assays, especially in high-throughput settings, would compare all curves at the same threshold. If this threshold was set based on the assay targeting the original sequence it would lead to the Cq of the assay targeting B.1.177.19 being estimated high, closer to the curve plateau. If

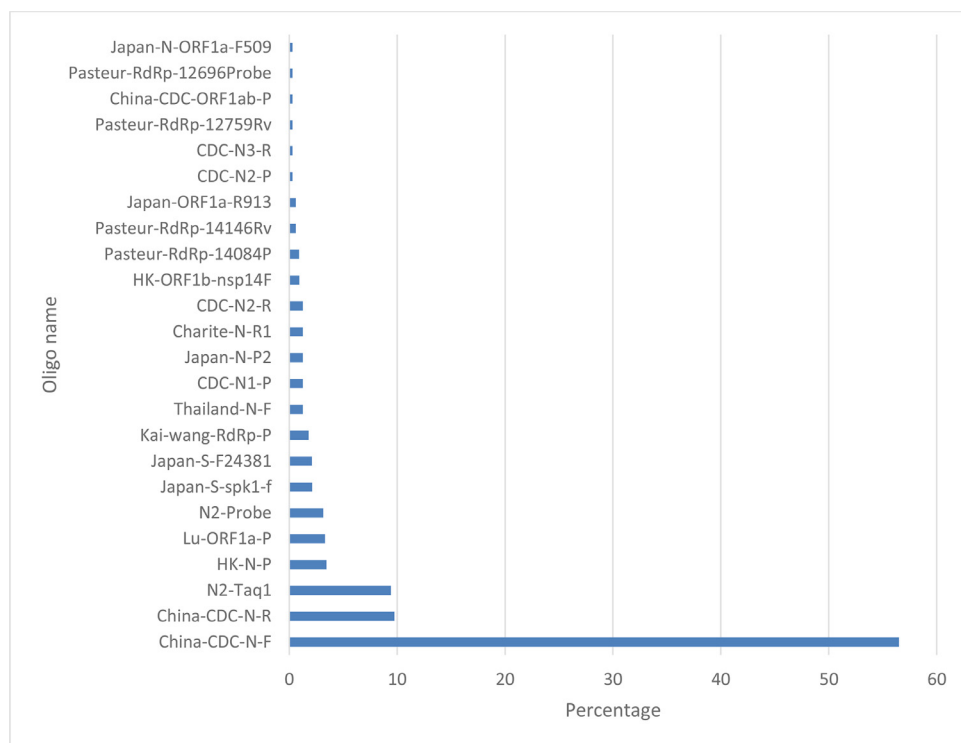


Fig. 2. Percentage of SARS-CoV-2 sequences ($n = 340$) containing 1 or more SNVs compared to oligo sequence; assay sequences with 100% of sequences containing SNVs are not displayed.

Table 3

Lineages with SNVs in the 22 published primers and probes (where lineage is defined). Excluding primers where >50% sequences have SNVs. Average SNV count in target region shown for sequences containing >0 SNVs.

Assay	Primer/probe	Target gene	Average SNV count	Lineages with SNVs
China CDC	Forward primer	N	2.94	B.1.1, B.1.1.7, B.1.1.4, B.1.1.1, B.1.1.89, B.1.1.296, B.1.1.161, B.1.1.37, B.1.1.164, B.1.1.10, B.1, B.1.1.196, B.1.1.220, B.1.1.203, B.1.1.315, B.1.351, B.1.1.207
China CDC	Reverse primer	N	1	B.1.1.7, B.1
China CDC	Probe (5'-FAM/BHQ1-3')	ORF1ab	1	B.39
CDC	N1-Probe	N	1	B.1.1
CDC	N2-Probe	N	1	B.1.1
CDC	N2-R	N	1	B.1.1.4, B.1.1.315
CDC	N3-R	N	1	B.1
Charite	N_Sarbeco_R1	N	1	B.1.367
Charite	RdRP_SARSr-P2 (5'-FAM/BBQ-3')	RdRp	1	B.1.177
Hong Kong	HKU-NP (5'-FAM/TAMRA-3')	N	1	B.1.177.19
Hong Kong	HKU-ORF1b-nsp14F	ORF1b-nsp14	1	B.1.1.4
Japan	NIID_WH-1_F509	ORF1a	1	B.1.1.7
Japan	NIID_2019-nCoV_N_P2 (FAM/BHQ)	N	1	B.1.1.4, B.1.1.315
Japan	NIID_2019-nCoV_N_R2	N	1	B.1.1.207
Japan	NIID_WH-1_R913	ORF1a	1	B
Japan	NIID_WH-1_F24381	S	1	B.1
Japan	WuhanCoV-spk1-f	S	1	B.1
Kai-Wang To	RdRp/Helicase Probe	RdRp	1	B.1, B.1.154
Lu	ORF1a probe	ORF1a	1	B.1, B.1.1, B.1.1.37
Grant	N2-Probe	N	1	B.1.177.19
Grant	N2-Taql	N	1	B.1.1.7
Institut Pasteur	nCoV_IP2-12759Rv	RdRp	1	B.1.1.7
Institut Pasteur	nCoV_IP4-14084Probe (FAM/BHQ-1)	RdRp	1	B.1.1.7
Institut Pasteur	nCoV_IP4-14146Rv	RdRp	1	B.1.1.1
Thailand	WH-NIC N-F	N	1	B.1.1

the threshold was set at a higher fluorescence, or amplitude was further reduced as a result of the SNV, then the B.1.177.19 could be considered negative, even when the viral quantity was high (solid blue line Fig. 1b). Consequently in the absence of visual inspection of the amplification plots, automatic threshold settings could lead to false-negative results when assay performance is impacted in this way. Patients infected with

variants of concern, with high viral burden, could potentially be told they do not have COVID-19.

PCR is vulnerable to sequence variation in primer and probe binding sites, even minor sequence changes can lead to assay failure and false-negative results [31]. Multiplexed PCR assays targeting two or more different genes within a single pathogen improve the reliability of the

assay for detecting variants of concern, as the chance of mutations affecting all targets is smaller. Many commercial assay providers favour this approach. Ultimately, it would be prudent to redesign the effected assay of a multiplexed solution when sequence variation is detected due to the fact that variants may accumulate and eventually affect all assays. One downside of multiple target assays is that they are difficult to incorporate into a routine clinical service where multiplexing of targets is used to include a number of different pathogens that cause the same clinical syndrome. Another consideration when applying a multiplex approach is that there may be a reduction in the performance of the assays due to optimisation challenges.

A simple way to determine whether PCR assays may be affected by sequence variation is to perform *in silico* analysis of primer and probe binding regions of all circulating variants of SARS-CoV-2. The data we have presented represents the SARS-CoV-2 in our centre during this time period and out of the 22 published assays included in the analysis, only three assays had perfectly complementary primer and probe sequences for all SARS-CoV-2 sequences included in the analysis. Similar findings were observed in another study performing *in silico* analysis [32].

Many of the sequence changes were a single SNV in one primer or the probe and may not actually affect wet-laboratory performance of the assay: however this cannot easily be determined based on *in silico* approaches alone as other factors (reagent choice, assay optimum, etc.) will also affect how the PCR will perform. Furthermore, the effect of mutations can be cumulative, therefore it would seem prudent to modify the assays as soon as possible to avoid impact of further genetic changes over time. However, this may be challenging if SNVs are occurring frequently, especially as it will not be clear which will become VOCs. Therefore, sequence data should be proactively monitored for additional SNVs and lineages known to have SNVs in regions targeted by assays should be monitored for sudden increase in frequency. Ultimately, *in silico* analysis can guide the re-design of PCR assays to target the most conserved genomic regions of SARS-CoV-2.

The majority of clinical diagnostic laboratories in the UK are currently using SARS-CoV-2 RT-qPCR assays supplied by commercial providers and rely on assay manufacturers to perform validation and on-going quality assurance to ensure assays are fit for purpose. There is an expectation that this includes the impact of sequence variation on assay performance, however there is no standard guidance for commercial assay providers (or diagnostic laboratories) on how to monitor the impact of sequence variation on PCR performance during a pandemic.

Such guidance should include the frequency with which the *in silico* analysis should be performed, the reference database of variant sequences used, recommended bioinformatics tools and how to assess the impact of sequence variation on assay performance.

The *in silico* analysis we present here could be broadened to include all lineages circulating locally and also any other global lineages of interest in GISAID [19]. The availability of SARS-CoV-2 sequence data in the UK provides an excellent opportunity to perform surveillance of sequence data and to pre-empt the impact of any variation on diagnostic tests. We propose that a standardised approach should be adopted to continuously monitor the impact of sequence variants on diagnostic tests before they become prevalent in the community and guide the selection of highly conserved regions of the SARS-CoV-2 genome to use as PCR targets to improve diagnostic tests in the future.

5. Funding statement

This work was supported by the UK National Measurement System and the European Metrology Programme for Innovation and Research (EMPIR) joint research project [18HLT03] “SEPTIMET” which has received funding from the EMPIR programme co-financed by the Participating States and the European Union’s Horizon 2020 research and innovation programme.

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.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.jcvp.2021.100037](https://doi.org/10.1016/j.jcvp.2021.100037).

Appendix: Full list of primer and probe sequenced from published assays used in this analysis. Available from WHO website: <https://www.who.int/publications/m/item/molecular-assays-to-diagnose-covid-19-summary-table-of-available-protocols>

China CDC	ORF1ab	
(Wang et al. [23])	Forward primer	CCC TGT GGG TTT TAC ACT TAA
	Reverse primer	ACG ATT GTG CAT CAG CTG A
	Probe (5'-FAM/BHQ1-3')	CCG TCT GCG GTA TGT GGA AAG GTT ATG G
	N	
	Forward primer	GGG GAA CTT CTC CTG CTA GAA T
	Reverse primer	CAG ACA TTT TGC TCT CAA GCT G
	Probe (5'-FAM/TAMRA-3')	TTG CTG CTT GAC AGA TT
Charite (Corman et al. [24])	RdRP	
	RdRP_SARsF-F2	GTG ARA TGG TCA TGT GTG GCG G
	RdRP_SARsF-R1	CAR ATG TTA AAS ACA CTA TTA GCA TA
	RdRP_SARsF-P2 (5'-FAM/BBQ-3') (Specific for 2019-nCoV)	CAG GTG GAA CCT CAT CAG GAG ATG C
	RdRP_SARsF-P1 (5'-FAM/BBQ-3') (Pan Sarbeco-Probe)	CCA GGT GGW ACR TCA TCM GGT GAT GC
	E	
	E_Sarbeco_F1	ACA GGT ACG TTA ATA GTT AAT AGC GT
	E_Sarbeco_R2	ATA TTG CAG TAC GCA CAC A
	E_Sarbeco_P1 (5'-FAM/BBQ-3')	ACA CTA GCC ATC CTT ACT GCG CTT CG
	N	
	N gene_N_Sarbeco_F1	CAC ATT GGC ACC CGC AAT C
	N_Sarbeco_R1	GAG GAA CGA GAA GAG GCT TG
	N_Sarbeco_P1 (5'-FAM/BBQ-3')	ACT TCC TCA AGG AAC AIT GCC A
Hong Kong (Chu et al. [25])	ORF1b-nsp14	
	HKU-ORF1b-nsp14F Sarbeco	TGG GGY TTT ACR GGT AAC CT
	HKU-ORF1b-nsp14R Sarbeco	AAC RCG CTT AAC AAA GCA CTC
	HKU-ORF1b-nsp141P (5'-FAM/TAMRA-3') Sarbeco	TAG TTG TGA TGC WAT CAT GAC TAG
	N	

(Continued)

China CDC	ORF1ab	
	HKU-NF Sarbeco	TAA TCA GAC AAG GAA CTG ATT A
	HKU-NR Sarbeco	CGA AGG TGT GAC TTC CAT G
	HKU-NP (5'-FAM/TAMRA-3') Sarbeco	GCA AAT TGT GCA ATT TGC GG
Thailand	N	
	WH-NIC N-F	CGT TTG GTG GAC CCT CAG AT
	WH-NIC N-R	CCC CAC TGC GTT CTC CAT T
	WH-NIC N-P (FAM/BQH1)	CAA CTG GCA GTA ACC A
Japan	ORF1a	
	NIID_WH-1_F501	TTC GGA TGC TCG AAC TGC ACC
	NIID_WH-1_R913	CTT TAC CAG CAC GTG CTA GAA GG
	NIID_WH-1_F509	CTC GAA CTG CAC CTC ATG G
	NIID_WH-1_R854	CAG AAG TTG TTA TCG ACA TAG C
	S	
	WuhanCoV-spk1-f	TTG GCA AAA TTC AAG ACT CAC TTT
	WuhanCoV-spk2-r	TGT GGT TCA TAA AAA TTC CTT TGT G
	NIID_WH-1_F24381	TCA AGA CTC ACT TTC CAC
	NIID_WH-1_R24873	ATT TGA AAC AAA GAC ACC TTC AC
	N	
	NIID_2019-nCoV_N_F2	AAA TTT TGG GGA CCA GGA AC
	NIID_2019-nCoV_N_R2	TGG CAG CTG TGT AGG TCA AC
	NIID_2019-nCoV_N_P2 (FAM/BHQ)	ATG TCG CGC ATT GGC ATG GA
CDC	N	
	2019-nCoV_N1 F	GAC CCC AAA ATC AGC GAA AT
	2019-nCoV_N1-R	TCT GGT TAC TGC CAG TTG AAT CTG
	2019-nCoV_N1-P (FAM/BHQ)	ACC CGG CAT TAC GTT TGG ACC
	2019-nCoV_N2-F	TTA CAA ACA TTG GCC GCA AA
	2019-nCoV_N2-R	GCG CGA CAT TCC GAA
	2019-nCoV_N2-P (FAM/BHQ)	ACA ATT TGC CCC CAG CGC TTC AG
	2019-nCoV_N3-F	GGG AGC CTT GAA TAC ACC AAA A
	2019-nCoV_N3-R	TGT AGC ACG ATT GCA TTG
	2019-nCoV_N3-P (FAM/BHQ)	AYC ACA TTG GCA CCC GCA ATC CTG
Institut Pasteur	RdRp gene / nCoV_IP2	
	nCoV_IP2-12669Fw	ATG AGC TTA GTC CTG TTG
	nCoV_IP2-12759Rv	CTC CCT TTG TGT
	nCoV_IP2-12696bProbe (HEX/BHQ-1)	AGA TGT CTT GTG CTG CCG GTA
	RdRp gene / nCoV_IP4	
	nCoV_IP4 nCoV_IP4-14059Fw	GGT AAC TGG TAT GAT TTC G
	nCoV_IP4-14146Rv	CTG GTC AAG GTT AAT ATA GG
	nCoV_IP4-14084Probe (FAM/BHQ-1)	TCA TAC AAA CCA CGC CAG G
	E_Sarbeco (CoVE)	
	E_Sarbeco_F1	ACA GGT ACG TTA ATA GTT AAT AGC GT
	E_Sarbeco_R2	ATA TTG CAG TAC GCA CAC A
	E_Sarbeco_P1 (FAM/BHQ)	ACA CTA GCC ATC CTT ACT GCG CTT CG
To et al. [26]	RdRp	
	RdRp/Helicase Forward	CGC ATA CAG TCT TRC AGG CT
	RdRp/Helicase Reverse	GTG TGA TGT TGA WAT GAC ATG GTC
	RdRp/Helicase Probe	TTA AGA TGT GGT GCT TGC ATA CGT AGA C
Chan et al. [27]	RdRp	
	RdRp Forward Sarbeco	CAA GTG GGG TAA GGC TAG ACT TT
	RdRp Reverse Sarbeco	ACT TAG GAT AAT CCC AAC CCA T
	S	
	S Forward	CCT ACT AAA TTA AAT GAT CTC TGC TTT ACT
	S Reverse	CAA GCT ATA ACG CAG CCT GTA
Lu et al. [28]	ORF1a	
	ORF1a forward	AGA TTG GTT AGA TGA TAG T
	ORF1a reverse	TTC CAT CTC TAA TTG AGG TTG AAC
	ORF1a probe	TGG TCA ACA AGA CGG CAG TGA GGA
Grant et al. [13]	N	
	N gene Taq1	TCTGGTAAAGGCCAACACAA
	N gene Taq2	TGTATGCTTTAGTGGCAGTACG
	N gene Probe	CTGTCATAAGAAATCTGCTGTGAGGC
PDV Internal Control		
	PDV-F	GCG GGT GCC TTT TAC AAG AAC
	PDV-R	CAG AAT AAG CAA AAT TGA TAG GAA CCA T
	PDV-Pr	TCT TTC CTC AAC CTC GTC CGT CAC AAG T

References

- [1] Emergencies Preparedness, W.H.H., in: Diagnostic Testing For SARS-CoV-2, World Health Organisation, 2020, p. 20.
- [2] E. de Sousa, et al., Mortality in COVID-19 disease patients: correlating the association of major histocompatibility complex (MHC) with severe acute respiratory syndrome 2 (SARS-CoV-2) variants, *Int. J. Infect. Dis.* 98 (2020) 454–459.
- [3] A. Erol, Are the emerging SARS-CoV-2 mutations friend or foe? *Immunol. Lett.* 230 (2021) 63–64.
- [4] K. Pancer, et al., The SARS-CoV-2 ORF10 is not essential in vitro or in vivo in humans, *PLoS Pathog.* 16 (12) (2020) e1008959.
- [5] X. Xie, et al., Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera, *Nat. Med.* 27 (2021) 620–621.
- [6] B.O. Villoutreix, et al., In silico investigation of the new UK (B.1.1.7) and South African (501Y.V2) SARS-CoV-2 variants with a focus at the ACE2-Spike RBD interface, *Int. J. Mol. Sci.* 22 (4) (2021).
- [7] B.E. Young, et al., Effects of a major deletion in the SARS-CoV-2 genome on the severity of infection and the inflammatory response: an observational cohort study, *Lancet N. Am. Ed.* 396 (10251) (2020) 603–611.
- [8] L. Zhang, et al., SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity, *Nat. Commun.* 11 (1) (2020) 6013.
- [9] SARS-CoV-2 Variants of Concern and Variants Under Investigation in England - Technical Briefing 7, Public Health England, 2021.
- [10] E.e. team, Updated rapid risk assessment from ECDC on the risk related to the spread of new SARS-CoV-2 variants of concern in the EU/EEA – first update, *Eurosurveillance* 26 (3) (2021) 2101211.
- [11] P. Simmonds, Rampant C→U hypermutation in the genomes of SARS-CoV-2 and other coronaviruses: causes and consequences for their short- and long-term evolutionary trajectories, *mSphere* 5 (3) (2020).
- [12] J.R. Brown, et al., Comparison of SARS-CoV2 N gene real-time RT-PCR targets and commercially available mastermixes, *bioRxiv* (2020) 2020.04.17.047118.
- [13] P.R. Grant, et al., Extraction-free COVID-19 (SARS-CoV-2) diagnosis by RT-PCR to increase capacity for national testing programmes during a pandemic, *bioRxiv* (2020) 2020.04.06.028316.
- [14] Resende, P.C., et al., SARS-CoV-2 genomes recovered by long amplicon tiling multiplex approach using nanopore sequencing and applicable to other sequencing platforms. 2020.
- [15] J.R. Tyson, et al., Improvements to the ARTIC multiplex PCR method for SARS-CoV-2 genome sequencing using nanopore, *bioRxiv* (2020) 2020.09.04.283077.
- [16] D.P. R&D, et al., COVID-19 ARTIC v3 Illumina Library Construction and Sequencing Protocol V.5, 2020 Available from <https://www.protocols.io/view/covid-19-artic-v3-illumina-library-construction-an-bibt kann>.
- [17] COVID-19 Genomics UK Consortium. Available from: <https://www.cogconsortium.uk/>.
- [18] N. Loman, W. Rowe, A. Rambaut, ARTIC-nCoV-BioinformaticsSOP-v1.1.0, 2020 Available from: <https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html>.
- [19] S. Elbe, G. Buckland-Merrett, Data, disease and diplomacy: gISAID's innovative contribution to global health, *Glob. Chall.* 1 (1) (2017) 33–46.
- [20] H. Li, et al., The sequence alignment/map format and SAMtools, *Bioinformatics* 25 (16) (2009) 2078–2079.
- [21] F. Sievers, et al., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol. Syst. Biol.* 7 (2011) 539.
- [22] O'Toole, Á., J. McCrone, and E. Scher. *Phylogenetic Assignment of Named Global Outbreak Lineages (Pangolin)*. Available from: <https://github.com/cov-lineages/pangolin>.
- [23] D. Wang, et al., Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in Wuhan, China, *JAMA* 323 (11) (2020) 1061–1069.
- [24] V.M. Corman, et al., Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR, *Eur. Surveill. Bull. Eur. sur les Mal. Transm. Eur. Commun. Dis. Bull.* 25 (3) (2020) 2000045.
- [25] D.K.W. Chu, et al., Molecular Diagnosis of a Novel Coronavirus (2019-nCoV) Causing an Outbreak of Pneumonia, *Clin. Chem.* 66 (4) (2020) 549–555.
- [26] K.K.W. To, et al., Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study, *Lancet Infect. Dis.* 20 (5) (2020) 565–574.
- [27] J.F.-W. Chan, et al., A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster, *Lancet N. Am. Ed.* 395 (10223) (2020) 514–523.
- [28] R. Lu, et al., Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding, *Lancet N. Am. Ed.* 395 (s) (2020) 565–574.
- [29] M. Artesi, et al., A recurrent mutation at position 26340 of SARS-CoV-2 is associated with failure of the E gene quantitative reverse transcription-PCR utilized in a commercial dual-target diagnostic assay, *J. Clin. Microbiol.* 58 (10) (2020) e01598-20.
- [30] M.R. Hasan, et al., A novel point mutation in the N gene of SARS-CoV-2 may affect the detection of the virus by RT-qPCR, *J. Clin. Microbiol.* (2021) JCM.03278-20.
- [31] D. Li, J. Zhang, J. Li, Primer design for quantitative real-time PCR for the emerging Coronavirus SARS-CoV-2, *Theranostics* 10 (16) (2020) 7150–7162.
- [32] M. Gand, et al., Use of Whole Genome Sequencing Data for a First in Silico Specificity Evaluation of the RT-qPCR Assays Used for SARS-CoV-2 Detection, *Int. J. Mol. Sci.* 21 (15) (2020).