



Article

Evidence-Based Evaluation of PCR Diagnostics for SARS-CoV-2 and the Omicron Variants by Gold Standard Sanger Sequencing

Sin Hang Lee

Abstract

Both SARS-CoV-2 and SARS-CoV-1 initially appeared in China and spread to other parts of the world. SARS-CoV-2 has generated a COVID-19 pandemic causing more than 6 million human deaths worldwide, while the SARS outbreak quickly ended in six months with a global total of 774 reported deaths. One of the factors contributing to this stunning difference in the outcome between these two outbreaks is the inaccuracy of the RT-qPCR tests for SARS-CoV-2, which generated a large number of false-negative and false-positive test results that have misled patient management and public health policymakers. This article presents Sanger sequencing evidence to show that the RT-PCR diagnostic protocol established in 2003 for SARS-CoV-1 can in fact detect SARS-CoV-2 accurately due to the well-known ability of the PCR to amplify similar, homeologous sequences. Using nested RT-PCR followed by Sanger sequencing to retest 50 patient samples collected in January 2022 and sold as RT-qPCR positive reference confirmed that 21 (42%) were false-positive. Routine sequencing of the RT-PCR amplicons of the receptor-binding domain (RBD) and N-terminal domain (NTD) of the Spike protein (S) gene is a tool to avoid false positives and to study the effects of amino acid mutations and multi-allelic single nucleotide polymorphisms (SNPs) in the circulating variants for investigation of their impacts on vaccine efficacies, therapeutics and diagnostics.

Copyright © The Authors — Published Under the Creative Commons License
Share/Alike (see <https://creativecommons.org/licenses/>)

Correspondence: Sin Hang Lee; email: shlee01@snet.net Tel. 203 878-1438

Milford Molecular Diagnostics Laboratory 2044 Bridgeport Avenue, Milford, CT 06460, USA

Keywords

SARS-CoV-2; SARS-CoV-1; RT-PCR; Sanger sequencing; RT-qPCR; receptor-binding domain (RBD); N-terminal domain (NTD); Omicron; multi-allelic SNPs; false-positive

Contents

1	Introduction	145
2	Materials and Methods	146
3	Results	149
4	Discussion	180
5	Conclusion	184
6	References	184
7	Author statements	188

1 Introduction

The SARS-CoV-2 virus that causes the COVID-19 pandemic is genetically closely related to the SARS-CoV-1 virus that caused the outbreak of severe acute respiratory syndrome (SARS) in late 2002. Both viruses have a genome of single-stranded positive-sense RNA of nearly 30,000 nucleotides that share a 79% similarity [1,2], and both use the angiotensin-converting enzyme 2 (ACE2) as their major receptor to enter the host cell [3].

As of 4 April 2022, there were more than 491 million cumulative human cases and more than 6 million deaths due to COVID-19 [4], which were reported worldwide with a case fatality rate of 1.22% since its outbreak in late 2019. By contrast, the SARS outbreak ceased in July 2003 with a global total of 8,098 reported cases and 774 deaths [5], a case fatality rate of 9.7%, which is 7.95-fold higher than that of the COVID-19 pandemic. Comparative studies suggested that a higher transmission rate of SARS-CoV-2 among human populations was responsible for the high death toll of COVID-19 [6,7].

However, there are also public health measure differences in managing these two outbreaks, which might have contributed to the higher global death toll of COVID-19. For example, the public record shows that during the 2002/2003 SARS outbreak in China, the laboratory diagnostics for SARS cases were based on conventional RT-PCR using a series of primers. After purification of the PCR products, cycling sequencing reactions were performed to

determine the nucleotide sequence for the definitive molecular diagnosis of SARS-CoV-1 infections [8]. According to one report, the US CDC-designed PCR primers were directed to the polymerase gene of all coronaviruses and amplified a 405 bp fragment from the newly emerging coronavirus. The amplicon was then sequenced and compared with the GenBank reference sequences for molecular diagnosis [9]. In another document, the CDC recommended using three specific primers to perform RT-PCR on patient samples and to sequence a 348-bp PCR amplicon “to verify the authenticity of the amplified product” [10]. With accurate diagnoses based on DNA sequencing, prompt isolation of patients and early treatment, the SARS outbreak ended in July [11]; the pandemic was stopped in 2003 by applying travel restrictions and isolating individuals infected by SARS-CoV-1 [12]. To reaffirm this gold-standard approach to diagnose RNA viruses, the Food and Drug Administration (FDA) also issued a guideline on January 2, 2009 that detection of enterovirus RNA requires generating RT-PCR amplicons from two different genomic regions of the virus and to perform bi-directional sequencing on one of the amplicons; and the sequence of the amplicon should match the reference or consensus sequence of the virus [13].

Contrary to the previously established protocol and guideline set by the CDC and the FDA for the diagnosis of SARS-CoV-1 and for RNA viruses, the SARS-CoV-2 commercial RT-qPCR assay kits are generating a Ct number, an unproven surrogate for nucleotide sequence, for “the presumptive qualitative detection of nucleic acid from the 2019-nCoV” under emergency use authorization [14]. Using conventional RT-PCR and Sanger sequencing, as recommended by the CDC for SARS-CoV-1 in 2003, to retest two sets of patient samples showed that the current commercial RT-qPCR test kits for SARS-CoV-2 assays generated at least 20% false-negative and 30% false-positive results on nasopharyngeal swab samples collected from

patients with respiratory infection in early 2020 [15] and 47% false positives in the nasopharyngeal swab samples collected from patients with respiratory infection in the month of October, 2020 in the United States [16], before any variants of concern emerged.

Accurate viral detection is a starting point to contain the COVID-19 pandemic [17,18]. Early accurate diagnosis with early isolation and early treatment of the patients can significantly reduce the number of deaths. A comparative study of case infection rate (CIR) and case fatality rate (CFR) between healthcare workers (HCW) and non-healthcare workers (non-HCW) in Wuhan during the SARS-CoV-2 outbreak showed that while the CIR of HCWs (2.10%) was dramatically higher than that of non-HCWs (0.43%), the CFR of HCWs (0.69%) was significantly lower than that of non-HCWs (5.30%) [19]. Improving test sensitivity and specificity remains an urgent need [17-18, 20].

The purpose of this study was to introduce a generic amplicon sequencing protocol implementable in diagnostic laboratories, as recommended by the CDC [10] and the FDA [13], to verify the definitive detection of SARS-CoV-2 in patient samples, including determination of its variants by partial S gene sequencing.

Accurate determination of the mutations in the RBD and NTD of the S gene of the SARS-CoV-2 Omicron variants is needed in selecting therapeutics for COVID-19 patients. The current standard care in antiviral treatment for moderate to severe COVID-19 includes the use of the monoclonal antibody combination REGN10933 (casivirimab) and REGN10897 (imdevimab) [21]. However, the K417N, E484A, S477N, and Q493R mutations in the RBD would lead to loss of electrostatic interactions with REGN10933, whereas a mutation of G446S would lead to steric clashes with REGN10897 [22], causing neutralization escapes [23]. The Q493R and Q498R mutations are known

to introduce additional electrostatic interactions with ACE2 residues Glu35 and Asp38, respectively, whereas S477N enables hydrogen-bonding with ACE2 Ser19. Collectively, these latter mutations strengthen ACE2 binding and could be a factor in the enhanced transmissibility of Omicron relative to previous variants [21]. In addition, the deletions of NTD amino acid sequences, such as $\Delta 69-70$, $\Delta 141-144$ and $\Delta 146$ are known to be associated with immune escape in certain patients because these deletions may hinder NTD recognition by neutralizing antibodies from convalescent plasma [24].

2 Materials and Methods

2.1. RT-qPCR positive reference samples for evaluation

A total of 50 nasopharyngeal swab specimens from patients with clinical respiratory infection, which were collected in the month of January 2022 and tested positive for SARS-CoV-2 by an RT-qPCR assay, were re-tested in this study by Sanger sequencing for the presence of the Omicron variant. Another 16 nasopharyngeal swab samples from patients with clinical respiratory infection, which were collected in October, 2020 and verified to be true-positive for SARS-CoV-2 by bidirectional partial Sanger sequencing of the N gene and S gene RBD [15], were used to evaluate the effectiveness of the SARS-CoV-1 specific PCR primers [10] in detecting SARS-CoV-2 genomic RNA.

These RT-qPCR positive reference specimens without patient identifications were purchased from Boca Biolistics Reference Laboratory, Pompano Beach, FL, a commercial reference material laboratory endorsed by the FDA as a supplier of clinical samples positive for SARS-CoV-2 by RT-qPCR assays. According to the commercial supplier, the swabs were immersed in VTM or saline after collection and stored in freezer at -80°C temperature following the initial testing.

2.2. Extracting viral RNA from infected cells

As previously reported, the test was designed to detect the viral RNA in the infected cells as well as in cell-free fluid [15, 16, 25]. To this end, about 1 mL of the nasopharyngeal swab rinse was transferred to a graduated 1.5 mL microcentrifuge tube and centrifuged at $\sim 16,000\times g$ for 5 min to pellet all cells and cellular debris. The supernatant was discarded except the last 0.2 mL, which was left in the test tube with the pellet. To each test tube containing the pellet and the residual fluid, 200 μL of digestion buffer containing 1% sodium dodecyl sulfate, 20mM Tris-HCl (pH 7.6), 0.2M NaCl and 700 $\mu\text{g}/\text{mL}$ proteinase K, was added. The mixture was digested at 47°C for 1 hr in a shaker. An equal volume (400 μL) of acidified 125:24:1 phenol:chloroform:isoamyl alcohol mixture (Thermo Fisher Scientific Inc.) was added to each tube. After vortexing for extraction and centrifugation at $\sim 16,000\times g$ for 5 min to separate the phases, the phenol extract was aspirated out and discarded. Another volume of 300 μL of acidified 125:24:1 phenol:chloroform:isoamyl alcohol mixture was added to the aqueous solution for a second extraction. After centrifugation at $\sim 16,000\times g$ for 5 min to separate the phases, 200 μL of the aqueous supernatant without any material at the interface was transferred to a new 1.5 mL microcentrifuge tube. To the 200 μL aqueous sample, 20 μL of 3M sodium acetate (pH5.2) and 570 μL of 95% ethanol were added. The mixture was placed in a cold metal block in a freezer set at -15 to -20°C for 20 min, and then centrifuged at $\sim 16,000\times g$ for 5 min. The precipitated nucleic acid was washed with 700 μL of cold 70% ethanol. After a final centrifugation at $\sim 16,000\times g$, the 70% ethanol was completely removed with a fine-tip pipette, and the microcentrifuge tube with opened cap was put into a vacuum chamber for 10 minutes to evaporate the residual ethanol. The nucleic acids in each tube were dissolved in 50 μL of diethylpyrocarbonate-treated water (ThermoFisher), and contained

residual human genomic DNA serving as indicator of sample adequacy. All nucleic acid extracts were tested immediately or stored at -80°C until testing.

2.3. PCR conditions

To initiate the primary RT-PCR, a total volume of 25 μL mixture was made in a PCR tube containing 20 μL of ready-to-use LoTemp® PCR mix with denaturing chemicals (HiFi DNA Tech, LLC, Trumbull, CT, USA), 1 μL (200 units) of Invitrogen SuperScript III Reverse Transcriptase, 1 μL (40 units) of Ambion™ RNase Inhibitor, 0.1 μL of Invitrogen 1 M DTT (dithiothreitol), 1 μL of 10 μmolar forward primer in TE buffer, 1 μL of 10 μmolar reverse primer in TE buffer and 1 μL of sample nucleic acid extract.

The ramp rate of the thermal cycler was set to 0.9 °C/s. The program for the temperature steps was set as: 47°C for 30 min to generate the cDNA, 85°C 1 cycle for 10 min, followed by 30 cycles of 85°C 30 sec for denaturing, 50°C 30 sec for annealing, 65°C 1 min for primer extension, and final extension 65°C for 10 minutes.

The nested PCR was conducted in a 25 μL volume of complete PCR mixture containing 20 μL of ready-to-use LoTemp® mix, 1 μL of 10 μmolar forward primer, 1 μL of 10 μmolar reverse primer and 3 μL of molecular grade water.

To initiate the nested PCR, a trace (about 0.2 μL) of primary PCR products was transferred by a micro-glass rod to the complete nested PCR mixture. The thermocycling steps were programmed to 85°C 1 cycle for 10 min, followed by 30 cycles of 85°C 30 sec for denaturing, 50°C 30 sec for annealing, 65°C 1 min for primer extension, and final extension 65°C for 10 minutes.

Transferring of PCR products was carried out by micro-glass rods in a PCR station, not by micropipetting, to avoid aerosol contamination.

2.4. DNA sequencing

The crude nested PCR products showing an expected amplicon at agarose gel electrophoresis were subjected to automated Sanger sequencing without further purification. To initiate a Sanger reaction, a trace (about 0.2 µL) of nested PCR products was transferred by a micro-glass rod into a thin-walled PCR tube containing 1 µL of 10 µmolar sequencing primer, 1 µL of BigDye® Terminator (v 1.1/Sequencing Standard Kit), 3.5 µL 5× buffer, and 14.5 µL water in a total volume of 20 µL. Twenty (20) enzymatic primer extension/termination reaction cycles were run according to the protocol supplied by the manufacturer (Applied Biosystems, Foster City, CA, USA).

After a dye-terminator cleanup, the Sanger reaction mixture was loaded in an Applied Biosystems SeqStudio Genetic Analyzer for sequence analysis. Sequence alignments were performed against the standard sequences stored in the GenBank database by online BLAST. The sequences were also visually analyzed for nucleotide mutations and indels.

2.5. PCR primers

The sequences of the 3 hemi-nested RT-PCR primers used to generate a 348-bp amplicon of the SARS-CoV-2 ORF1ab gene were listed in a CDC document [10]. Their sequences and the sequences of the nested RT-PCR primers used in this study for amplification of the N gene, the RBD and the S gene NTD [15, 16, 25] are summarized in Table 1.

Table 1. Sequences of PCR primers used to generate nested RT-PCR amplicons for Sanger sequencing

PCR Amplicon	Primer Type	Start	End	Sequence 5'-3'	Size (bp)
SARS-CoV ORF1ab gene Heminested CDC 2003	Cor-p-F2 (+) F.			CTAACATGCTTAGGATAATGG	368
	Cor-p-R1 (-) R.			CAGGTAAGCGTAAAACATCATC	
	Cor-p-F3 (+) F.			GCCTCTCTTGTTCTTGCTCGC	<u>348</u>
	Cor-p-R1 (-) R.			CAGGTAAGCGTAAAACATCATC	
SARS-CoV-2 N gene Co4/Co3 Nested	Co1 primary F.	28707	28727	ACATTGGCACCCGCAATCCTG	416
	Co8 primary R.	29102	29122	TTGGGTTTGTCTGGACCACG	
	Co4 nested F.	28720	28740	CAATCCTGCTAACAATGCTGC	<u>398</u>
	Co3 nested R.	29097	29117	TTTGTCTGGACCACGTCTGC	
SARS-CoV-2 S gene RBD S9/S10 Nested	SS1 primary F.	22643	22663	TGTGTTGCTGATTATTCCTGTC	460
	SS2 primary R.	23082	23102	AAAGTACTACTACTCTGTATG	
	S9 nested F.	22652	22672	GATTATTCCTGCTATATAAT	<u>445</u>
	S10 nested R.	23076	23096	CTACTACTCTGTATGGTTGGT	
SARS-CoV-2 S gene NTD SB7/SB8 Nested	SB5 primary F.	21619	21639	AACCAGAACTCAATTACCCCC	505
	SB6 primary R.	22103	22123	TTTGAAATTACCCTGTTTTCC	
	SB7 nested F.	21628	22648	TCAATTACCCCTGCATACAC	<u>490</u>
	SB8 nested R.	22097	22117	ATTACCCTGTTTTCTTCAAG	

Table 1 summarizes the 4 sets of PCR primers used in this study. The intended nested PCR amplicon size is underlined. Although not used in this study, the general primer set for amplification of the S gene NTD has been further modified to bypass the Δ24-26 and A27S mutations of the Omicron BA.2, BA.4 and BA.5 subvariants. The sequences of the new general primer set for the S gene NTD amplification are:

- SB11 5'-TCTCTAGTCAGTGTGTTAATC-3' Primary Forward
- SB6 5'-TTTGAAATTACCCTGTTTTCC-3' Primary Reverse
- SB12 5'-TTAATCTTACAACCAGAACTC-3' Nested Forward
- SB8 5'-ATTACCCTGTTTTCTTCAAG-3' Nested Reverse

2.6. Determination of variants of concern and interest was based on the amino acid mutations found in partial sequencing of the S gene and N gene listed in Table 2

Table 2. Key amino acid mutations in the S gene RBD, S gene NTD and the N gene used for variant determination [26–28]

WHO Name Variant	Pango lineage	ACE2 RBD mutations	NTD mutations	N-gene mutations
Alpha	B.1.1.7	N501Y	69del, 70del, 144del	N. A.
Beta	B.1.351	K417N, E484K, N501Y	D80A	N. A.
Gamma	P.1	K417T, E484K, N501Y	D138Y	N. A.
Delta	B.1.617.2	L452R, T478K	T95I, G142D, E156del, F157del, R158G	N. A.
Omicron BA.1	B.1.1.529.1	S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, 498R, N501Y, Y505H	A67V, Δ69-70, T95I, G142D, Δ143-145	R203K, G204R
Omicron BA.2	B.1.1.529.2	S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, S477N, T478K, E484A, Q493R, Q498R, N501Y, Y505H	Δ24-26, A27S, G142D	R203K, G204R
Epsilon	B.1.427	L452R		N. A.
Epsilon	B.1.429	L452R	W152C	N. A.
Eta	B.1.525	E484K	A67V, 69del, 70del, 144del	N. A.
Iota	B.1.526	E484K	T95I	N. A.
Kappa	B.1.617.1	L452R, E484Q	G142D, E154K	N. A.
Kappa	B.1.617.3	L452R, E484Q	G142D	N. A.
Lambda	C.37	L452Q, F490S	G75V, T76I	N. A.

Table 2 shows that sequencing the 445-bp ACE2 RBD nested PCR amplicon can detect the key amino acid mutations from S371 to Y505. The combination patterns of these RBD mutations with additional information from the NTD sequencing can reliably diagnose all major variants of concern and variants of interest.

3 Results

Since Sanger sequencing is used to provide physical evidence, based on which the diagnostic technology and data are evaluated, a higher-than-usual number of electropherograms are presented in the Results.

3.1. Using SARS-CoV-1 specific RT-PCR primers to detect SARS-CoV-2

Sixteen (16) SARS-CoV-2 positive samples collected in October 2020 were selected for hemi-nested RT-PCR amplification with the 3 PCR primers, which the CDC designed and recommended for SARS-CoV-1 specific RT-PCR diagnosis in 2003 [10]. They all generated a 348-bp amplicon with an identical 306-base interprimer sequence.

One of the 16 pairs of bidirectional sequencing electropherograms is presented in Figures 1A and 1B for illustration (overleaf). The 5’-3’ composite sequence derived from the two electropherograms presented in Figures 1A and 1B is as follows:

```
GCCTCTCTTGTTCTTGCTCGCAAACATACAACG
TGTTGTAGCTTGTACACCGTTTCTATAGATTA
GCTAATGAGTGTGCTCAAGTATTGAGTGAAAT
GGTCATGTGTGGCGGTTCACTATATGTTAAACC
AGGTGGAACCTCATCAGGAGATGCCACAACCTG
CTTATGCTAATAGTGTTTTTAAACATTTGTCAAG
CTGTCACGGCCAATGTTAATGCACTTTTATCTA
CTGATGGTAACAAAATTGCCGATAAGTATGTC
CGCAATTTACAACACAGACTTTATGAGTGTCTC
TATAGAAATAGAGATGTTGACACAGACTTTGT
GGATGAGTTTTACGCTTACCTG
```

Figure 1A

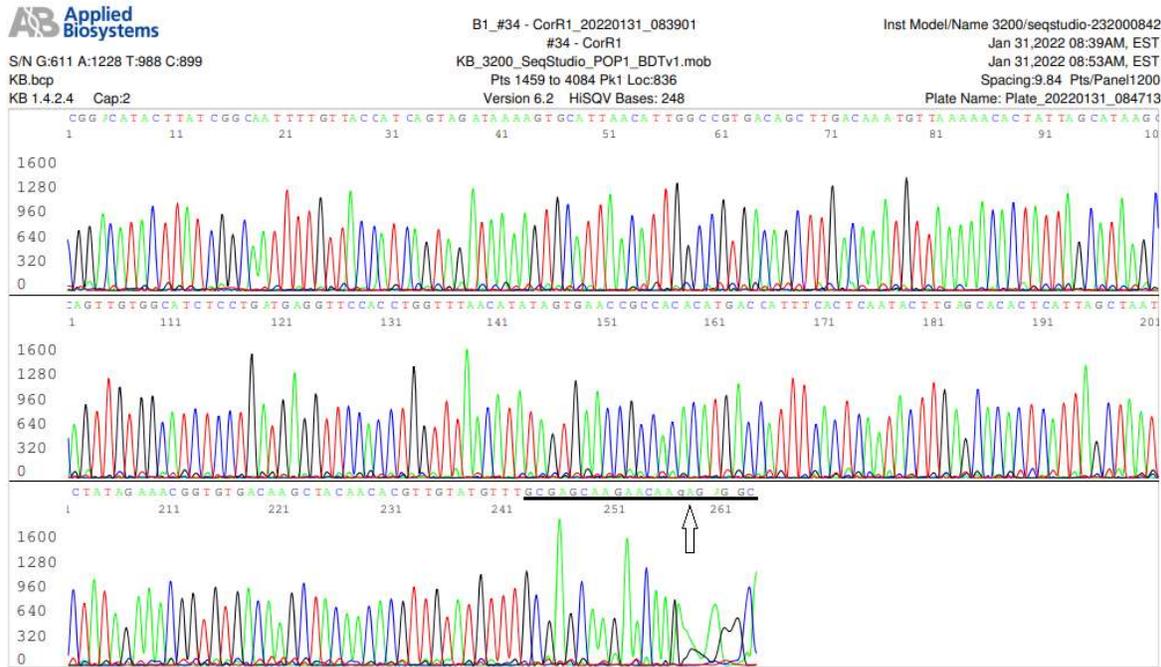
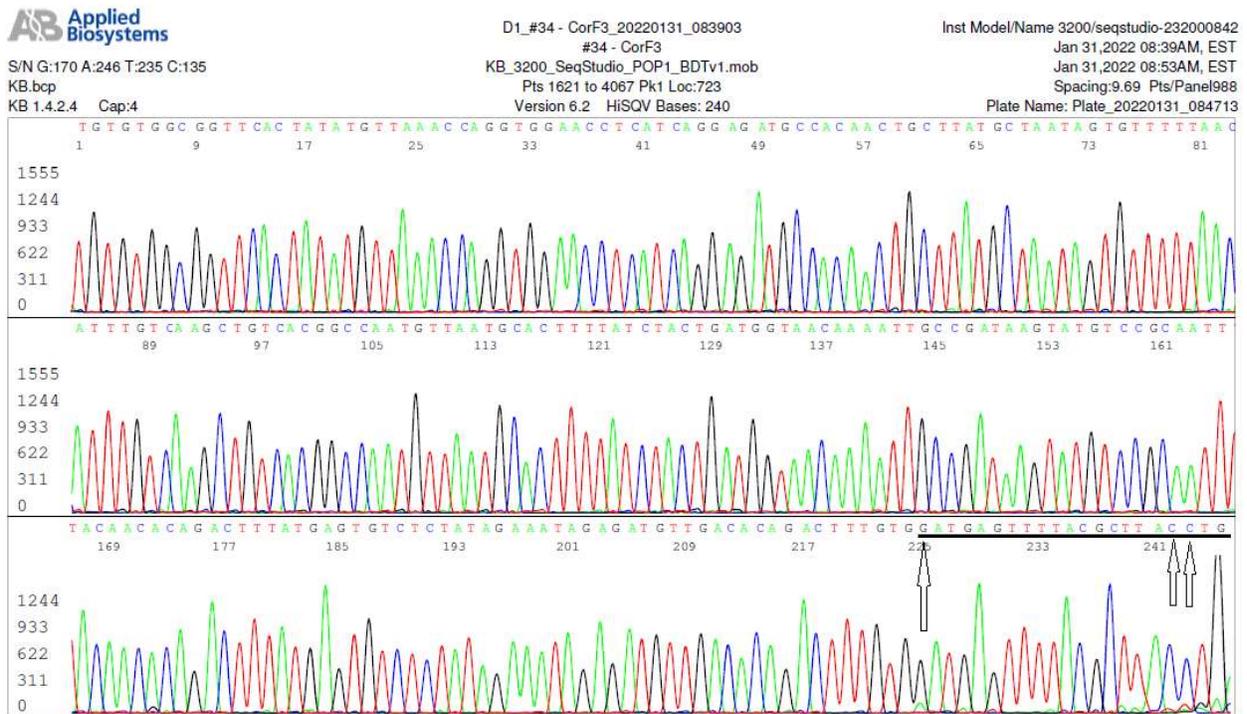


Figure 1B



The two computer-generated bidirectional sequencing electropherograms presented in Figures 1A and 1B show the 3’-5’ sequence of a SARS-CoV-2 gene RT-PCR amplicon, using the CDC-recommended SARS-CoV-1 Cor-p-R1 (-) reverse PCR primer 5’-CAGGTAAGCGTAAACTCATC -3’ as the sequencing primer (Figure 1A), and the 5’-3’ sequence of the same amplicon, using the CDC-recommended forward PCR primer Cor-p-F3 (+) 5’-

GCCTCTCTTGTCTTCTTGCTCGC-3’ as the sequencing primer (Figure 1B), respectively. The RT-PCR amplification was successful in spite of 4 mismatched nucleotides pointed by 4 arrows in the two underlined primer sequences. One mismatch is in the forward primer (Figure 1A) and 3 mismatches are in the reverse primer (Figure 1B). One of the mismatched nucleotides, a base G, is located in the 3’ end of the reverse primer (Figure 1B).

Submission of this 348-base sequence for BLAST alignment analysis showed that the 306-base interprimer sequence has a 100% match with more than 1,000 SARS-CoV-2 ORF1ab gene sequences recently deposited in the GenBank and the corresponding segment of the SARS-CoV-2 Wuhan Hu-1 prototype sequence. One of the >1,000 matches is presented in Figure 2A, a segment of SARS-CoV-2 ORF1ab gene sequence derived from a sample collected in Minnesota, USA

on January 30, 2022 with GenBank sequence ID# OM775626. This reference sequence was copied from the GenBank database and pasted in Figure 2B for comparison with a corresponding SARS-CoV-2 Wuhan Hu-1 prototype sequence (GenBank Sequence ID# NC_045512.2), presented in Figure 2C, to show that there is only one-base difference between the OM775626 and the Wuhan Hu-1 prototype sequence in this 348-base segment in the reverse primer-binding site.

Figure 2A

Severe acute respiratory syndrome coronavirus 2 isolate
 SARS-CoV-2/human/USA/MN-CDC-QDX33502310/2022, complete genome
 Sequence ID: [OM775626.1](#) Length: 29747 Number of Matches: 1
 Range 1: 15278 to 15621 [GenBankGraphics](#) [Next Match](#) [Previous Match](#)

Alignment statistics for match #1			
Score	Expect	Identities	Gaps
625 bits(338)	1e-174	342/344 (99%)	0/344 (0%)
<u>Query 1</u>	GCCTCTCTTGTCTTCTTGCTCGCAACATACACGTTGTGTAGCTTGTGCACCCGTTTCTAT	60	
<u>Sbjct 15278</u>	GCCTCACTTGTCTTCTTGCTCGCAACATACACGTTGTGTAGCTTGTGCACCCGTTTCTAT	15337	
<u>Query 61</u>	AGATTAGCTAATGAGTGTGCTCAAGTATTGAGTGAATGGTCATGTGTGGCGGTTCACTA	120	
<u>Sbjct 15338</u>	AGATTAGCTAATGAGTGTGCTCAAGTATTGAGTGAATGGTCATGTGTGGCGGTTCACTA	15397	
<u>Query 121</u>	TATGTTAAACCAGGTGGAACTCATCAGGAGATGCCACAACCTGCTTATGCTAATAGTGT	180	
<u>Sbjct 15398</u>	TATGTTAAACCAGGTGGAACTCATCAGGAGATGCCACAACCTGCTTATGCTAATAGTGT	15457	
<u>Query 181</u>	TTTAACATTTGTCAAGCTGTACGGCCAATGTTAATGCACCTTTATCTACTGATGGTAAC	240	
<u>Sbjct 15458</u>	TTTAACATTTGTCAAGCTGTACGGCCAATGTTAATGCACCTTTATCTACTGATGGTAAC	15517	
<u>Query 241</u>	AAAATTGCCGATAAGTATGTCGCAATTTACACACAGACTTTATGAGTGCTCTATAGA	300	
<u>Sbjct 15518</u>	AAAATTGCCGATAAGTATGTCGCAATTTACACACAGACTTTATGAGTGCTCTATAGA	15577	
<u>Query 301</u>	AATAGAGATGTTGACACAGACTTTGTG ATGAGTTTTACGCTTA	344	
<u>Sbjct 15578</u>	AATAGAGATGTTGACACAGACTTTGTG AATGAGTTTTACGCTTA	15621	

Figure 2A is copy of a BLAST report from the GenBank showing a 348-base segment of SARS-CoV-2 genome sequence generated by a pair of PCR primers specifically designed by the CDC for SARS-CoV-1 RT-PCR diagnostics. This BLAST report only listed 344 of the 348 bases submitted for alignment because the reverse primer has 2 adjacent unmatched GG/TT bases near its 5’ end. One T/A mismatch in the forward primer and 1 G/A mismatch in the reverse primer are typed in red. The G/A mismatch in the 3’ end of the reverse primer did not prevent a successful PCR amplification.

Figure 2B

```

15241 atgtgataga gccatgccta acatgcttag aatgatggcc tctttgttc ttgctcgc aa
15301 acatacaacg tgttttagct tgtcacaccg tttctataga ttagctaata agtgtgctca
15361 agtattgagt gaaatgggca tgtgtggcgg ttcactatat gttaaaccag gtggaacctc
15421 atcaggagat gccacaactg cttatgctaa tagtgttttt aacatttgc aagctgtcac
15481 ggccaatggt aatgcacttt tatctactga tggtaaacaa attgccgata agtatgtccg
15541 caatttaca cacagacttt atgagtgtct ctatagaaat agagatgttg acacagactt
15601 tgtgatgatg ttttacgctt atgtgcgtaa acattttctca atgatgatac tctctgacga
    
```

Figure 2B is part of a SARS-CoV-2 ORF1ab gene sequence retrieved from the GenBank database, Sequence ID: OM775626 (submitted in February 2022). It contains a 306-base sequence fully matching the interprimer sequence presented in Figures 1A and 1B. The 3 CDC-recommended SARS-CoV-1 specific RT-PCR primer sequence sites are shaded gray or typed in red color. The mismatched nucleotides between the SARS-CoV-1 primers and the SARS-CoV-2 template are green-highlighted. It shows 2 nucleotide mismatches in the Cor-p-F2 (+) forward primary PCR primer position (shaded gray), 1 mismatch in the Cor-p-F3 (+) heminested forward PCR primer position (typed in red immediately downstream of the Cor-p-F2 (+) primer), and 3 mismatches in the Cor-p-R1 (-) heminested reverse PCR primer position (typed in red).

Figure 2C

```

15301 aaatgtgata gagccatgcc taacatgctt agaattatgg cctcacttgt tcttgcgcgc
15361 aacatacaaa cgtgtgttag cttgtcacac cgtttctata gattagctaa tgagtgtgct
15421 caagtattga gtgaaatggt catgtgtggc ggttcactat atgttaaacc aggtggaacc
15481 tcatcaggag atgccacaac tgcttatgct aatagtgttt ttaacatttg tcaagctgtc
15541 acggccaatg ttaatgcact tttatctact gatggtaaca aaattgccga taagtatgtc
15601 cgcaatttac aacacagact ttatgagtgt ctctatagaa atagagatgt tgacacagac
15661 tttgtgatg atg agttttacgc taaatgctg aaacatttct caatgatgat actctctgac
    
```

Figure 2C is part of a SARS-CoV-2 ORF1ab gene sequence retrieved from the GenBank Wuhan Hu-1 prototype Sequence ID: NC_045512.2. Compared to Sequence ID: OM775626, this Wuhan Hu-1 prototype sequence has one additional A/A mismatch against the Cor-p-R1 (-) heminested reverse PCR primer 14 bases away from the 3' end of the primer.

Based on the findings presented in Figures 1 and 2, the 3 SARS-CoV-1 Specific RT-PCR Primers recommended by the CDC in 2003 could easily have been used to detect the SARS-CoV-2 Wuhan Hu-1 prototype at the time of the outbreak for accurate RT-PCR/Sanger sequencing diagnosis of the COVID-19 cases to prevent or to curtail the subsequent pandemic.

3.2. SARS-CoV-2 was detected by RT-PCR and Sanger sequencing in only 29 of 50 RT-qPCR positive reference specimens

The results of nested RT-PCR amplification of the N gene and the S gene RBD of the 50 RT-qPCR positive samples are presented in Figure 3, panels A-E. Since the serial numbers M22-19 to M22-68 are for permanent Sanger sequencing identifications, these numbers will be referred to in the Results and Discussion sections of this paper for data correlation. The long numbers on the agarose gel images starting with S000 are ID numbers assigned

by the sample supplier for tracking their sources because these samples were sold as reference specimens, which may be used as the standard comparator to support medical device manufacturers' applications for FDA approval of new test kits.

Compared to the N gene PCR product bands, which were similar to that of the control P in fluorescence intensity on each run, the fluorescence intensity of the RBD PCR product bands varied greatly, although all the samples illustrated on each panel were processed in the same testing run, using the same nucleic acid extract to initiate the N gene

Figure 3

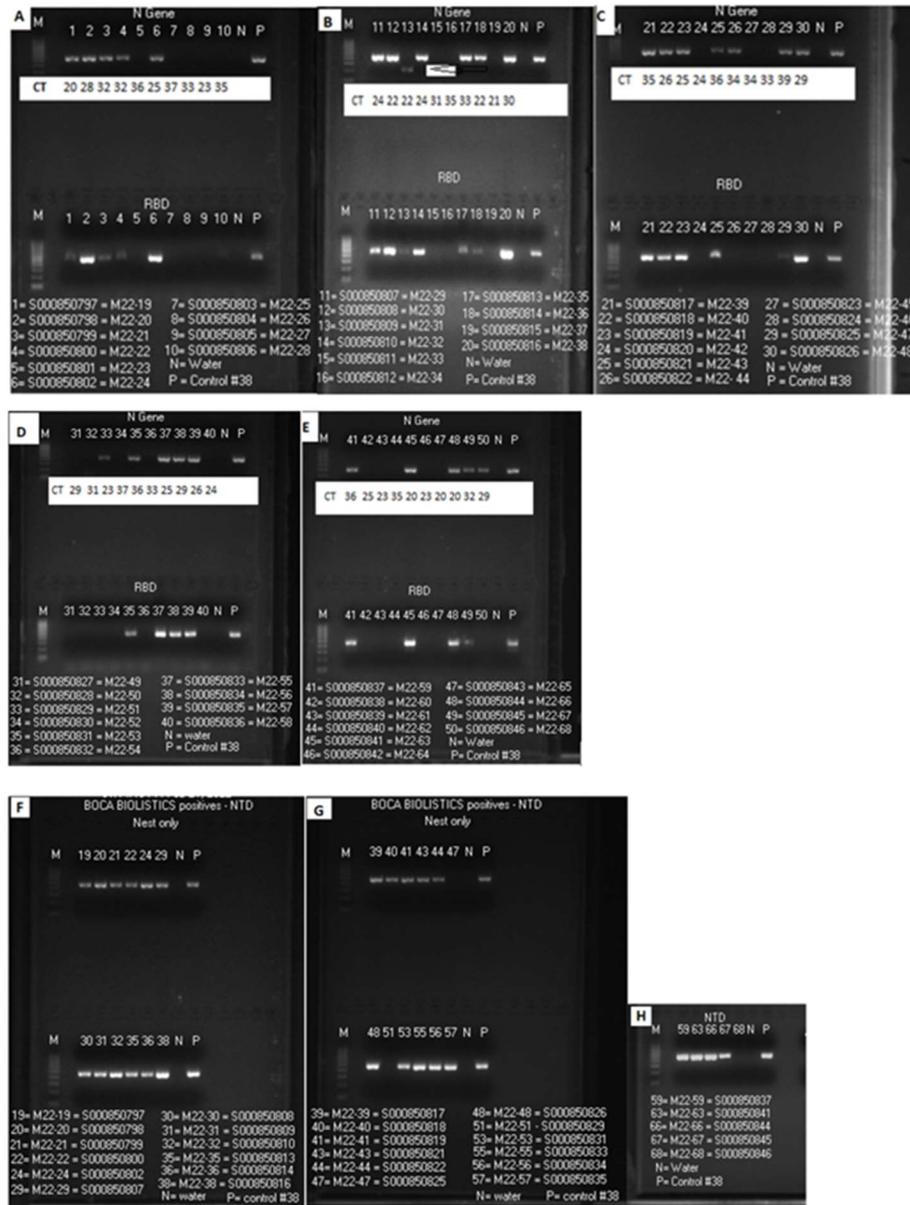


Figure 3. These are images of agarose gel electrophoresis of the SARS-CoV-2 N gene, RBD and NTD nested RT-PCR products. Panels A-E show a positive N gene band for 29 samples, M22-19, -20, -21, -22, -24, -29, -30, -31, -32, -35, -36, -38, -39, -40, -41, -43, -44, -47, -48, -51, -53, -55, -56, -57, -59, -63, -66, -67 and -68, in lanes 1, 2, 3, 4, 6, 11, 12, 13, 14, 17, 18, 20, 21, 22, 23, 25, 26, 29, 30, 33, 35, 37, 38, 39, 41, 45, 48, 49 and 50, respectively. These N gene PCR product bands were all about 398 bp in size except for that of sample M22-31 in lane 13, which was smaller in size and weak in fluorescence intensity (Panel B, lane 13 pointed by an arrowhead). The Ct values of the 50 RT-qPCR positive samples were listed in the N gene parts of the gel images.

RT-PCR and the RBD RT-PCR for each sample. The samples M22-44 (Figure 3, panel C, lane 26), M22-51 (Figure 3, panel D, lane 33) and M22-68 (Figure 3, panel E, lane 50) showed no RBD RT-PCR amplification. But an RT-PCR amplification of the NTD was successful on sample M22-44 (Figure 3, panel G, lane 44), indicating the presence of an S gene in this sample (also confirmed by DNA sequencing). All 29 samples found to be positive for N gene confirmed by DNA sequencing were subjected to an NTD nested RT-PCR amplification, and the images of the NTD nested RT-PCR results were presented in Figure 3, panels F, G and H, which show that except for samples M22-47, M22-51 and M22-68 (in Figure 3, panels G and H, lanes 47, 51 and 68), a robust NTD nested RT-PCR amplicon band similar to that of the control P was generated on the 26 samples that were also positive for a SARS-CoV-2 N gene RT-PCR amplification.

A special set of nested RT-PCR primers was designed in an attempt to amplify a segment of the

S gene upstream of the RBD on samples M22-47, M22-51 and M22-68 because the routine NTD nested RT-PCR failed to generate an amplicon from these 3 samples. Only 1 of the 3 samples, M22-51, yielded a nested RT-PCR amplicon for DNA sequencing.

All nested RT-PCR amplification products of the N gene, RBD and NTD were subjected to bi-directional Sanger sequencing, using the respective nested PCR primers as the sequencing primers. The results are summarized in Table 3 (overleaf).

3.3. Three RT-qPCR positive samples contained neither SARS-CoV-2 nor sufficient human cellular material

The nucleic acid extracts of the 21 samples, which were negative for N gene and RBD RT-PCR amplifications (Figure 3, panels A-E), were tested for the presence of human *BRCA* gene for sample adequacy. The results are presented in Figure 4.

adequacy. However, 3 samples, M22-42, M22-60 and M22-65, showed no human *BRCA* gene amplification, indicative of a lack of sufficient human cellular material in the samples. Notably, all these latter 3 samples had generated low Ct values (24, 25 and 20) although they did not contain detectable human cellular material or SARS-CoV-2.

BRCA gene has been shown to be a more stable indicator than the RNase P gene for the presence human cellular materials in archived nasopharyngeal swab specimens [15]. The fact that such low Ct values (24, 25 and 20) were generated by RT-qPCR testing on 3 clinical specimens, which had neither PCR-amplifiable *BRCA* gene nor RT-PCR-amplifiable SARS-CoV-2 nucleic acid, raised the possibility that the Ct values of the RT-qPCR may not always be a reliable yardstick for measuring SARS-CoV-2 viral loads in patient specimens. Numerous unidentified bacteria, fungi and viruses living in the normal nasal passageway can contribute nucleic acids to cause an unwanted positive quantitative PCR with a low Ct number.

Figure 4



Figure 4. This image of agarose gel electrophoresis of the nested PCR amplification products shows that 18 of the 21 samples, which were negative for SARS-CoV-2 N gene and RBD RT-PCR amplification, contained a segment of human *BRCA* gene, an indication of sample

Table 3. Correlation of the RT-PCR and the Sanger sequencing results of the 29 samples tested positive for SARS-CoV-2 by an EUA RT-qPCR assay and confirmed by Sanger sequencing

Sample No.	N gene			S gene RBD			S gene NTD			Special Comments
	PCR	FS(Co4)	RS(Co3)	PCR	FS(S9)	RS(S10)	PCR	FS(SB7)	RS(SB8)	
M22-19	+	+	+	+	+	+	+	+	+	
M22-20	+	+	+	+	+	+	+	+	+	
M22-21	+	+	+	+	+	+	+	+	+	
M22-22	+	+	+	+	+	+	+	+	+	
M22-24	+	+	+	+	+	+	+	+	+*	*Segment of multi-allelic SNPs in NTD.
M22-29	+	+	+	+	+	+	+	+	+	
M22-30	+	+	+	+	+	+	+	+	+	
M22-31	+ [#]	+ [#]	+ [#]	+	m-a	m-a	+	+	+	[#] N gene mutation and GGD deletion.
M22-32	+	+	+	+	+	+	+	+	+	
M22-35	+	+	+	+	+	+	+	+	+	
M22-36	+	+	+	+	+	ur ^a	+	+	+	^a Reverse primer sequencing unreadable.
M22-38	+	+	+	+	+	+	+	+	+	
M22-39	+	+	+	+	+	+	+	+	+	
M22-40	+	+	+	+	+	+	+	+	+	
M22-41	+	+	+	+	+	m-a	+	+	+	
M22-43	+	+	+	+	+	+	+	+	+	
M22-44	+	+	m-a	—	m-a ^b	—	+	+	+	^b Multi-allelic SNPs caused PCR failure.
M22-47	+	+ ^c	+ ^c	+	+	m-a	—	—	—	^c A competing Omicron with N gene S183P.
M22-48	+	+	+	+	+	m-a	+	+	+	
M22-51	+	+	+	—	— ^d	—	—	—	—	^d New primers needed for RBD amplicon.
M22-53	+	+	+	+	+	+	+	+	+	
M22-55	+	+	+	+	+	+	+	+	+	
M22-56	+	+	+	+	+	+	+	+	+	
M22-57	+	+	+	+	+	+	+	+	+	
M22-59	+	+	+	+	+	+	+	+	+	
M22-63	+	+	+	+	+	+	+	+	+	
M22-66	+	+	+	+	+	+	+	+	+	
M22-67	+	+	+	+	+	+	+	+	+	
M22-68	+	+	+	—	—	—	—	—	—	S gene PCR failure. N gene: R203K, G204R.

In Table 3, PCR = nested RT-PCR; the symbol “+” means a band was visible and the symbol “—” means a band was not visible at agarose gel electrophoresis.

FS(Co4) = Co4 forward sequencing primer;
 FS(S9) = S9 forward sequencing primer;
 FS(SB7) = SB7 forward sequencing primer;

RS(Co3) = Co3 reverse sequencing primer;
 RS(S10) = S10 reverse sequencing primer;
 RS(SB8) = SB8 reverse sequencing primer.

- + under FS(Co4) = R203K and G204R identified;
- + under RS(Co3) = R203K and G204R identified;
- + under FS(S9) = K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y and Y505H mutations identified in this sample;
- + under RS(S10) = T478K, S477N, G446S, N440K, K417N, S375F, S373P and S371L mutations identified in this sample;
- + under FS(SB7) = A67V, Δ69-70, T95I, G142D and Δ143-145 mutations identified in this sample;
- + under RS(SB8) = Δ143-145, G142D, T95I, Δ69-70 and A67V mutations identified in this sample.

3.4. Partial Sanger sequencing of the N gene and S gene as a diagnostic test for SARS-CoV-2 and Omicron variants

As summarized in Table 3, 21 of the 29 sequencing-confirmed positive samples, namely sample M22-19, M22-20, M22-21, M22-22, M22-24, M22-29, M22-30, M22-32, M22-35, M22-38, M22-39, M22-40, M22-43, M22-53, M22-55, M22-56, M22-57, M22-59, M22-63, M22-66 and M22-67, had R203K and G204R mutations in their N gene; S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y and Y505H mutations in their S gene RBD; and A67V, Δ 69-70, T95I, G142D and Δ 143-145 mutations in their S gene NTD. These mutations

were verified by bidirectional sequencing of a segment of the N gene, a segment of the RBD and a segment of the S gene NTD on each sample. However, 8 of the 29 samples, namely sample M22-31, M22-36, M22-41, M22-44, M22-47, M22-48, M22-51 and M22-68, which were confirmed to contain a segment of SARS-CoV-2 N gene by sequencing, failed to show R203K and G204R mutations in their N gene, or a complete set of bidirectional RBD and NTD sequences for definitive diagnosis of Omicron variant. A set of bidirectional sequencing electropherograms illustrating the Omicron variant mutations in the N gene, the RBD and the NTD of the S gene in the samples collected in January 2022 is presented in Figures 5-10.

Figure 6A

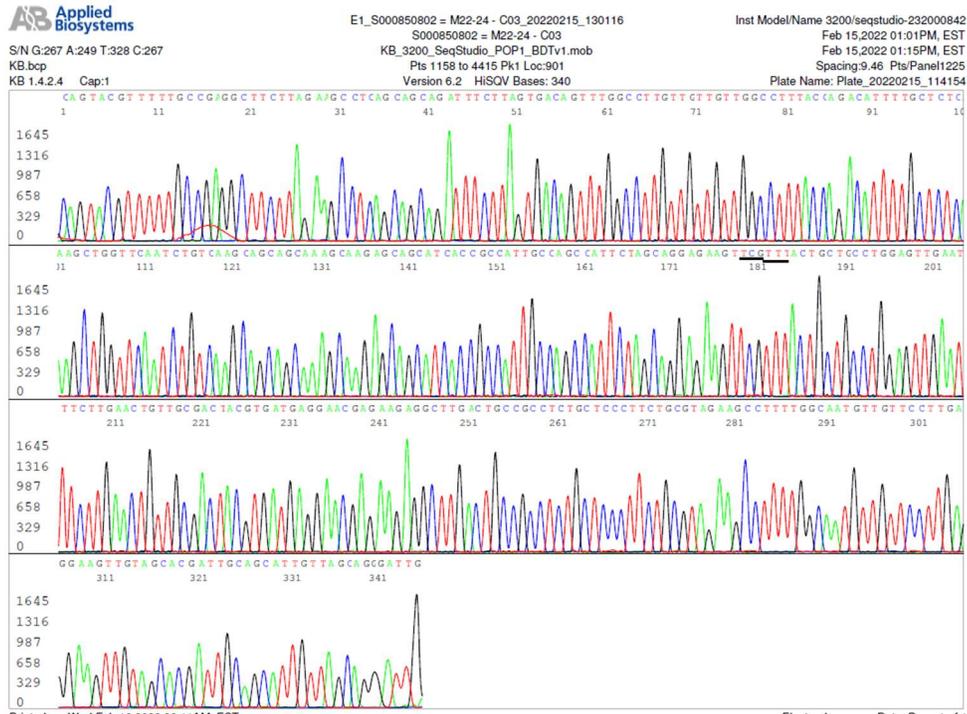
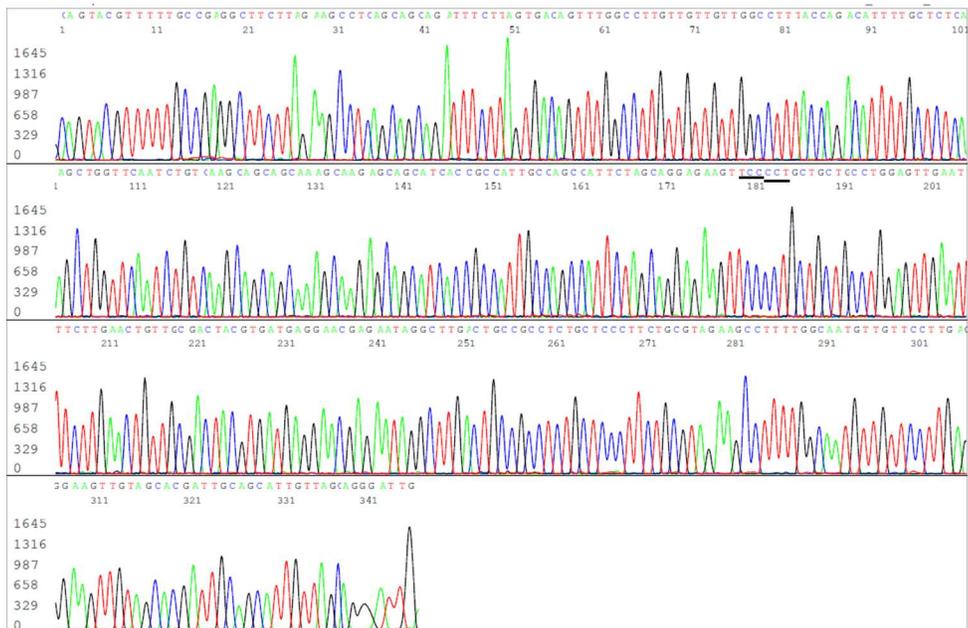


Figure 6B



Figures 6A and 6B. These two electropherograms showing the N gene G204R and R203K mutations in sample M22-24, using primer Co3 as the reverse sequencing primer (6A) and the wildtype SARS-CoV-2 Wuhan-Hu-1 control sequence for comparison (6B). Involved codons are underlined.

Figure 8A

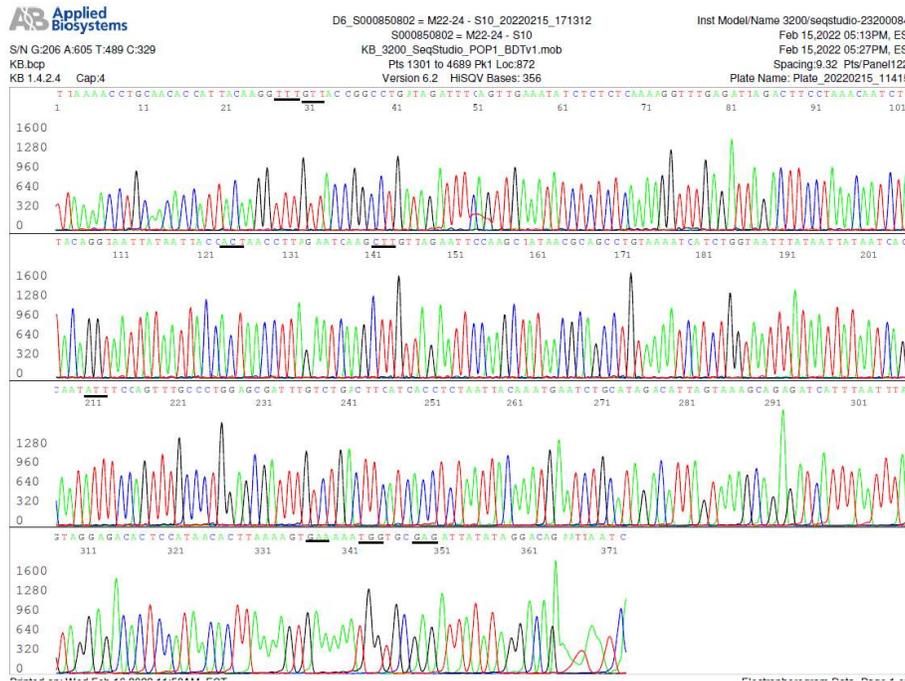
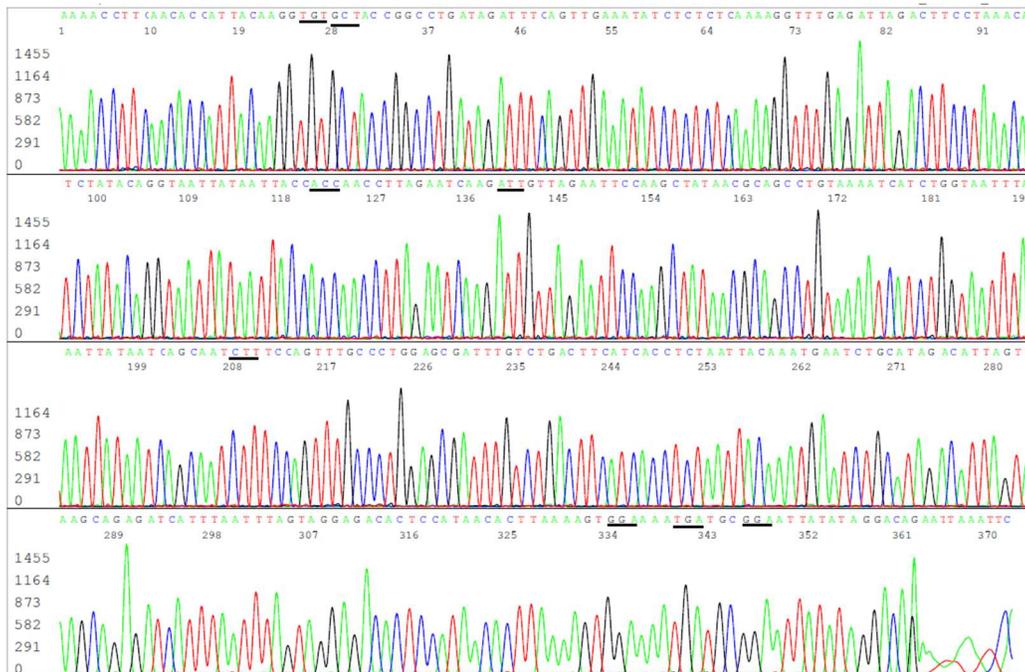


Figure 8B



Figures 8A and 8B. These two electropherograms show the S gene RBD T478K, S477N, G446S, N440K, K417N, S375F, S373P and S371L mutations in sample M22-24, using primer S10 as the reverse sequencing primer (8A) and the wildtype SARS-CoV-2 Wuhan-Hu-1 control sequence for comparison (8B). Involved codons are underlined.

Figure 9A

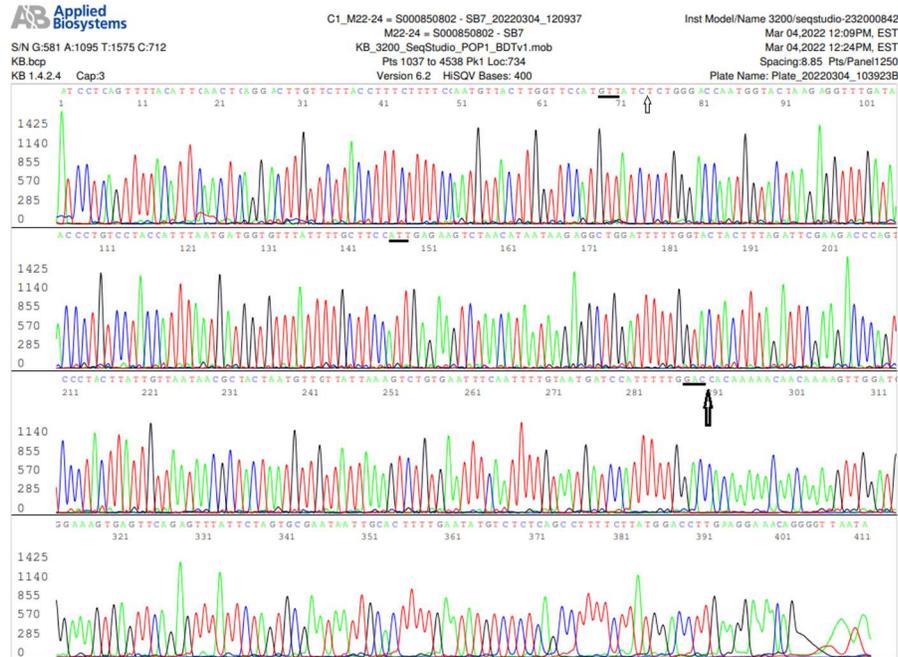
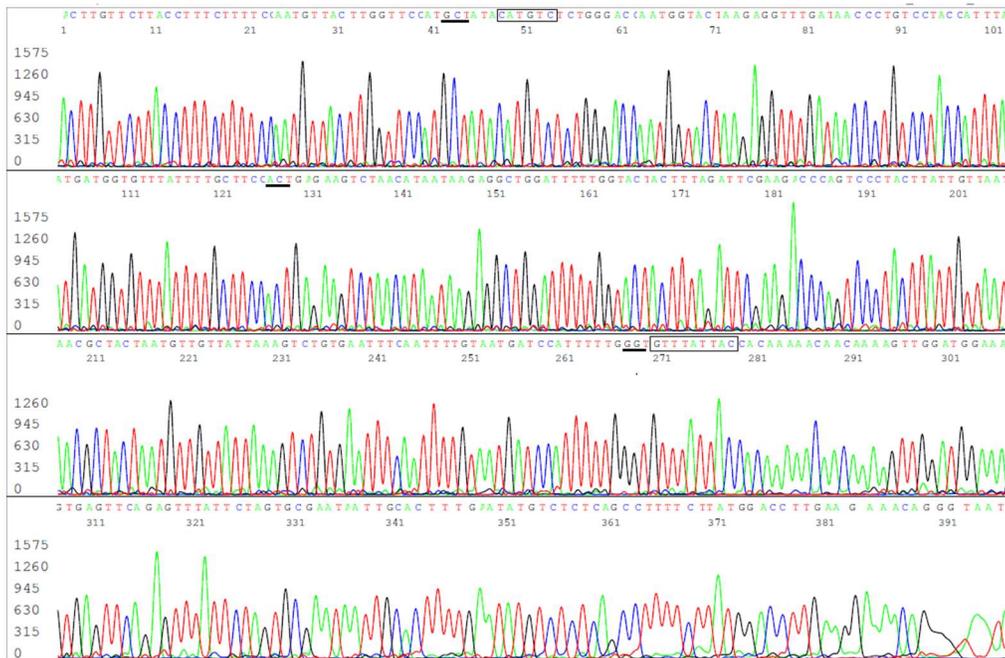


Figure 9B



Figures 9A and 9B show the S gene NTD A67V, Δ 69-70, T95I, G142D and Δ 143-145 mutations in sample M22-24, using primer SB7 as the forward sequencing primer (9A) and the wildtype SARS-CoV-2 Wuhan-Hu-1 control sequence for comparison (9B). Involved codons are underlined. The positions of Δ 69-70 and Δ 143-145 are indicated by a small arrow and a big arrow, respectively, in the M22 24 sequence (9A); and the corresponding nucleotides to be deleted for Omicron BA.1 are in two rectangular boxes in the control sequence (9B).

Figure 10A

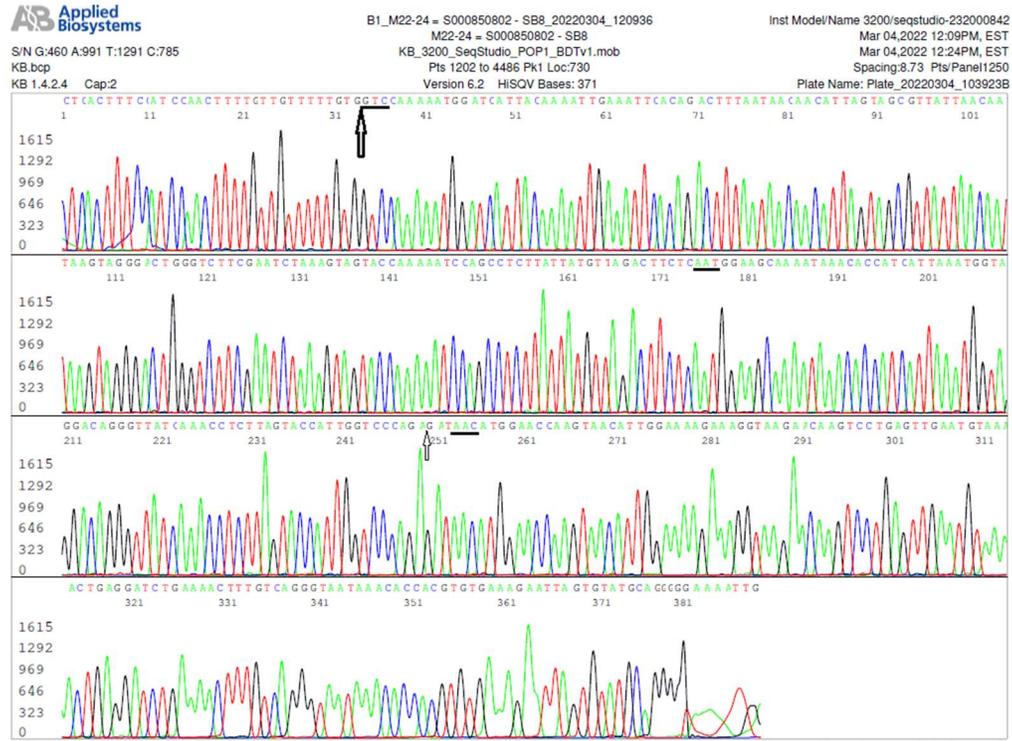
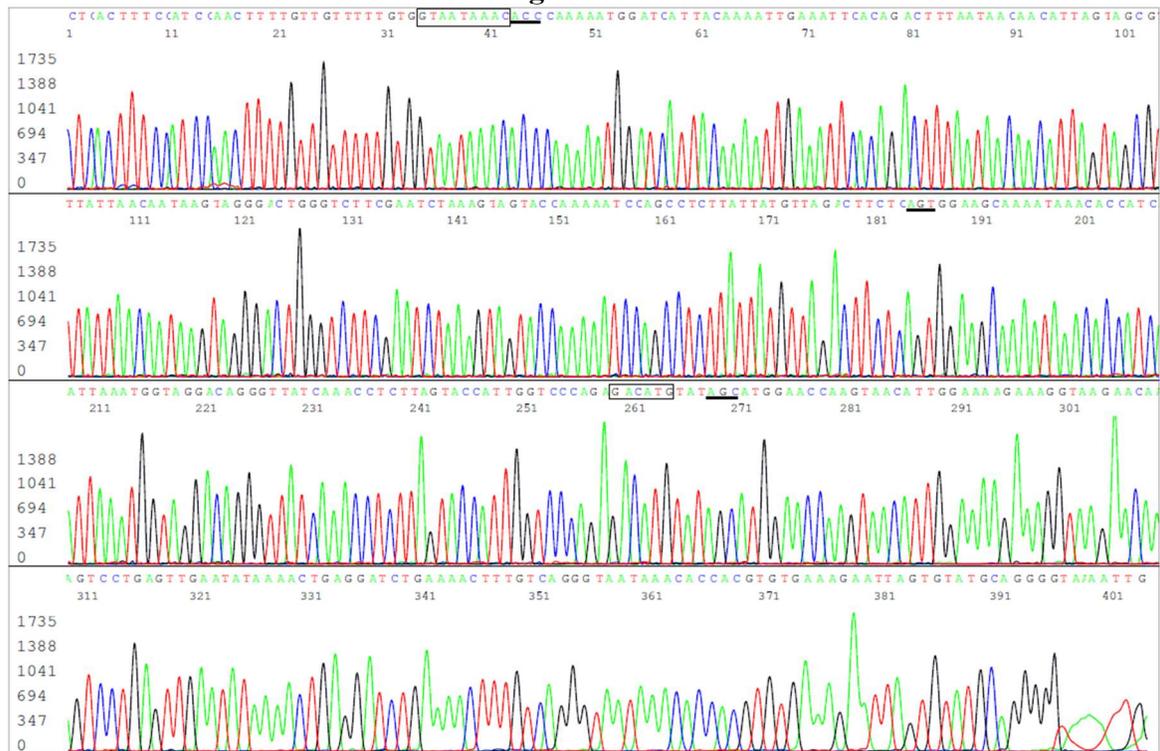


Figure 10B



Figures 10A and 10B show the S gene NTD Δ 143-145, G142D, T95I, Δ 69-70 and A67V mutations in sample M22-24, using primer SB8 as the reverse sequencing primer (A) and the wildtype SARS-CoV-2 Wuhan-Hu-1 control sequence for comparison (B). Involved codons are underlined. The positions of Δ 143-145 and Δ 69-70 are indicated by a big arrow and a small arrow, respectively, in the M22-24 sequence (A); and the corresponding nucleotides to be deleted for Omicron BA.1 are in two rectangular boxes in the control sequence (B).

3.5. Minor multi-allelic SNPs in the S gene NTD of Omicron variant

When the first set of electropherograms was analyzed, it was noticed that there were inconsistent segmental losses of sequencing signal in some of the samples, for example, during sequencing of the NTD of sample M22-24. This kind of loss of signal was not observed during sequencing of the COVID-19 samples collected prior to November 2020 [15, 16, 25]. In order to rule out technical artefacts that might be introduced from run-to-run sequencing variations, small aliquots ($\sim 0.2\mu\text{L}$) were transferred from one single tube of nested RT-PCR products into several Sanger reactions with either forward (SB7) or reverse (SB8) sequencing primer in one single run to generate several electropherograms, including those presented in Figure 9A, Figure 10A, Figure 11 and Figure 12, for comparison.

The presence of impure templates or multiple templates in one Sanger reaction is a well-known cause for loss of signal in DNA sequencing. Since the unreadable segments in the electropherograms presented in Figure 11 and Figure 12 are flanked by perfect SARS-CoV-2 sequences in both ends, these interfering DNAs must be parts of the target templates, which have mutated to form multi-allelic SNPs without an indel. An indel would have caused sequencing frameshift after the site of an indel [16,29].

3.6. Omicron variant with major multi-allelic SNPs in the S gene and N gene

The nested RT-PCR on sample M22-44 did not generate a visible RBD amplicon (see Figure 3, panel C, lane 26). But there was a clear NTD nested RT-PCR amplicon on this sample (see Figure 3, panel G, lane 44). Bidirectional DNA sequencing of the NTD RT-nested PCR products showed typical A67V, Δ 69-70, T95I, G142D and Δ 143-145 mutations, confirming the presence of an S gene in the sample.

Using the forward S9 PCR primer as the sequencing primer, Sanger sequencing of the RBD nested PCR products, which did not form a visible DNA band at gel electrophoresis (Figure 3, panel C, lane 26), showed small stretches of SARS-CoV-2 S gene RBD sequence in the background of an unreadable electropherogram, indicating that the usually dominant RBD sequence was being overshadowed by different species of RBD sequences with multi-allelic SNPs (Figure 13). However, base mutations of the RBD cannot be determined.

Figure 11

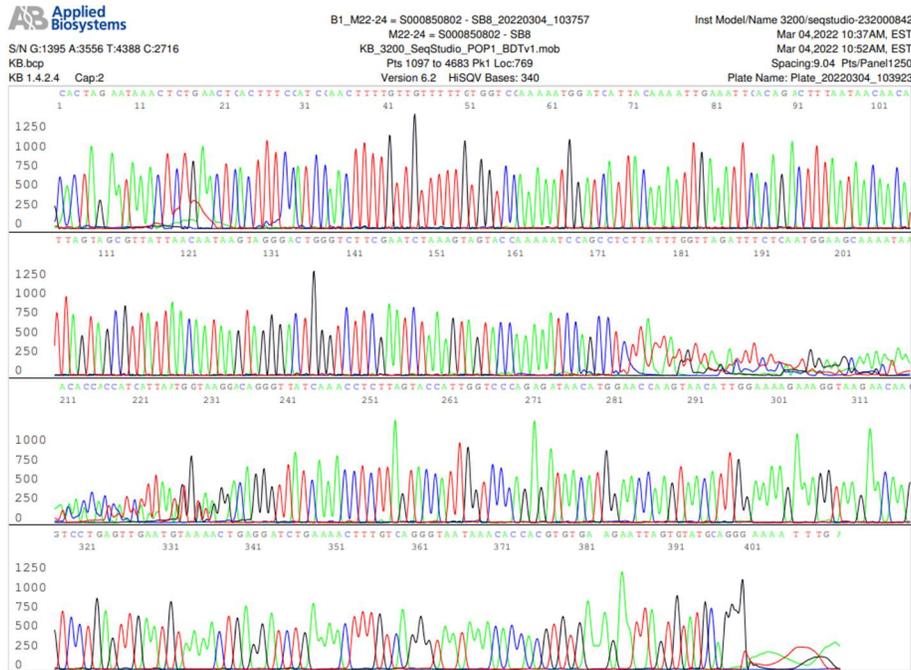
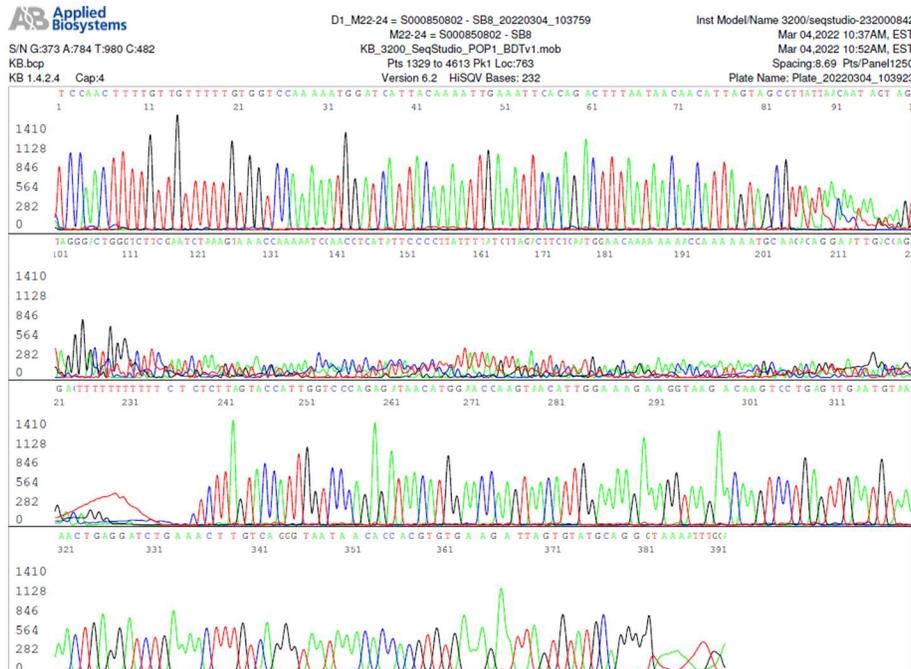


Figure 12



Figures 11 and 12. These electropherograms show loss of sequencing signal in the NTD reverse primer sequencing from base position 180 to base position 230 (Figure 11) and from base position 90 to base position 238 (Figure 12) although the template came from the same nested RT-PCR products, which were used as the template to generate Figures 9A and 10A.

Figure 13

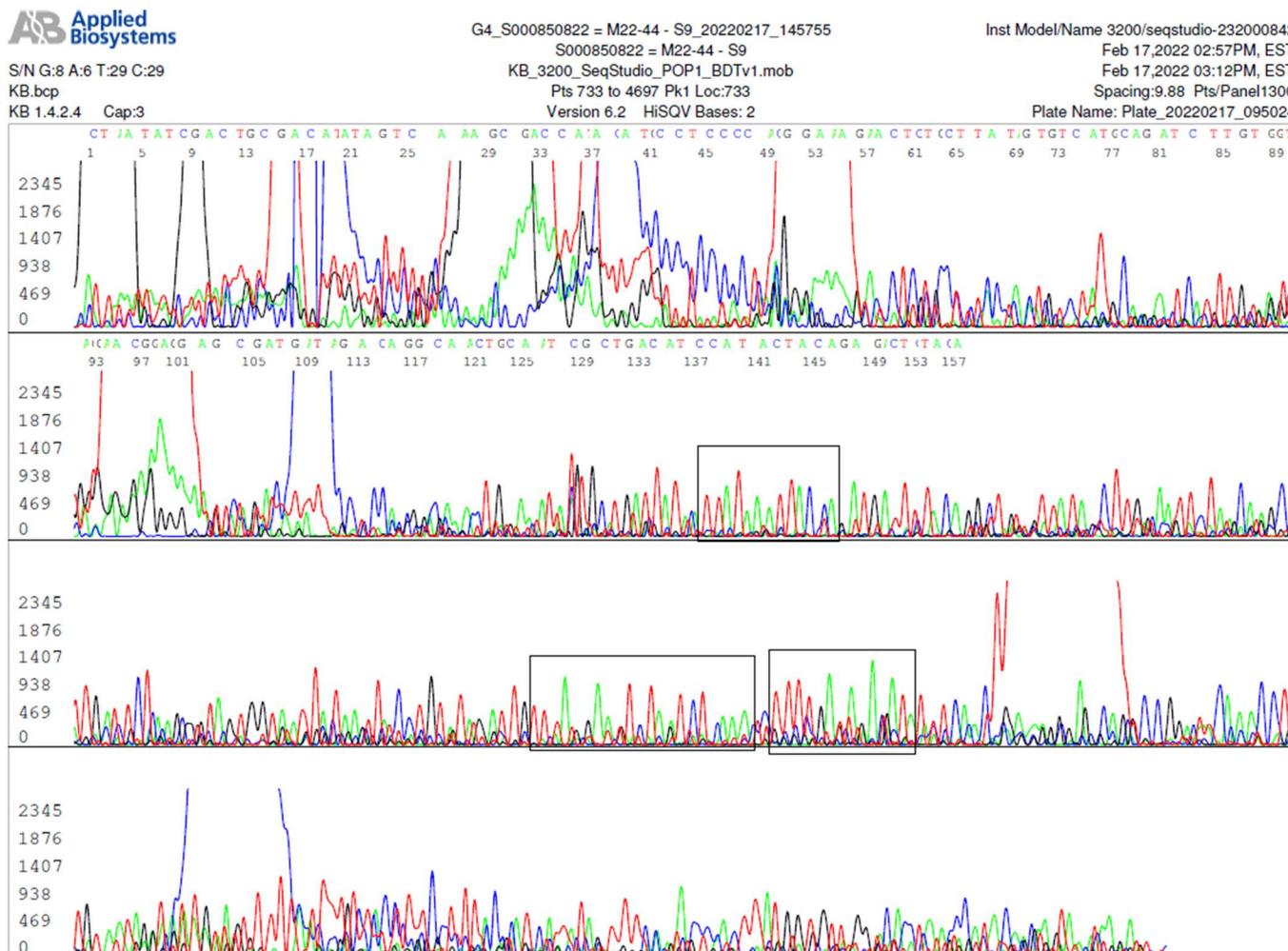


Figure 13. This is an electropherogram of forward primer sequencing of the RBD nested PCR products of sample M22-44 although a band of the PCR products was not visible to the naked eye (Figure 3, panel C, lane 26). Accurate base calling on this electropherogram was not possible due to multiple overlapping sequences. But the electropherogram showed one stretch of sequence “TTATAAATTACCA” in a single rectangle and another stretch of sequence “TCTAATCTCAAACCTTTTGAGAGAGAT” identified by two rectangles located about 97 bases downstream. These two stretches of sequences in their respective positions are characteristic of an S gene RBD of SARS-CoV-2 (compare these two sequences with that illustrated in Figure 7 A). The lack of a dominant PCR amplicon might account for the absence of an RBD nested RT-PCR product band for sample M22-44 (Figure 3, panel C, lane 26).

After the emergence of the Omicron variants in November 2021, SARS-CoV-2 genomes with many undetermined nucleic acid sequences in the RBD and the NTD of the S gene have been entered

in the GenBank database. One of these examples similar to the unreadable segment of RBD sequence (Figure 13 M22-44) is illustrated in Figure 14.

Figure 14

```

22441 ttcactgtag aaaaaggaat ctatcaaaact tctaacttta gagtccaacc aacagaatct
22501 attgtagat ttcctaatat tacaaacttg tgnntnnnn nnnnnnnnnn nnnnnnnnnn
22561 nnnnnnnnnn nnnnnnnnnn nnngaacagg aagagaatca gcaactnnnn nnnnnnnnat
22621 tctgtcctat ataattccgc atcattttcc actttaagt gtnnnnnnnn nnnnnnnnnt
22681 aaattaaatg atctctgctt tnctaagtgc tatgcagatt catttgnnnn nnnnnnnnat
22741 naagtcagac aaatcgctcc agggcaaact ggaaagatnn ctgattataa ttataaatta|
22801 ccnnntgatt ttacaggctn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnctaaggtt
22861 ggtggtaann nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn
22921 nnnnnnnnnn nnnnnnnaan nnnnnnnnnn nnnnnnnnnn nnnnnntgg tgttgaaggt
22981 tnnnattggt actttccttt acaatcatat ggnnnnnnnn nnnnnnnnnn nnnnnnnnnn
23041 nnnnnnnnnn nnnnngtagt actttctttt gannnnnnnn nnnnnnnnnn nnnnnnnnnn
23101 nnnnnnnnnn nnnnnnctaa tttggttaa aacaaatgtg tcaatttcaa cttcaatnnn
    
```

Figure 14. This is an S gene RBD nucleotide sequence excised from GenBank Seq ID# OL898842. The nucleotide positions 22615-22635 and 23039-23059 typed in red represent the positions of the sequences of the S9 forward nested PCR primer and the S10 reverse nested PCR primer, respectively. The sites for the primary RT-PCR primers are shaded gray. The letter “n” means that the base in that position can be a, c, g or t, undetermined due to multi-allelic SNPs. Although the sequences of the N gene and the S gene NTD of the GenBank Seq ID# OL898842 showed an amino acid mutation profile commonly associated with the Omicron variant, the profile of its amino acid mutations in the RBD remains unknown due to multi-allelic SNPs in this region, as illustrated in the sequence shown in Figure 14.

The reverse primer sequencing of the N gene nested PCR products on sample M22-44 generated a sequence with a large ~168-base unreadable segment between two perfectly deciphered sequences (Figure 15), while the forward primer sequencing showed a fully expected N gene sequence with R203K and G204R mutations commonly seen in an Omicron variant (Figure 16).

Loss of signal in diagnostic N gene sequencing is unusual [15]. A search of the GenBank database revealed that a group of SARS-CoV-2 sequences submitted to the GenBank after October 2021 contained a 117-base segment gap (Figure 17), which partially overlapped on the 168-base sequence framed in the two rectangles in Figure 16.

An identical 117-base gap is also found in the N gene of other SARS-CoV-2 genomes, such as those listed in GenBank Seq ID# OV086560 and Seq ID# OV080807. No translation was annotated in the GenBank database for these isolates. In addition to the 117-base gap, the green-highlighted 97-base sequence in Figure 17 shares only partial identity with the sequence in the rectangles in Figure 16. The findings of multi-allelic SNPs in the N gene and in the S gene RND in M22-44 suggest that at least some of the Omicron variant isolates harbor diverse genomic populations in one host [30-33].

Figure 16

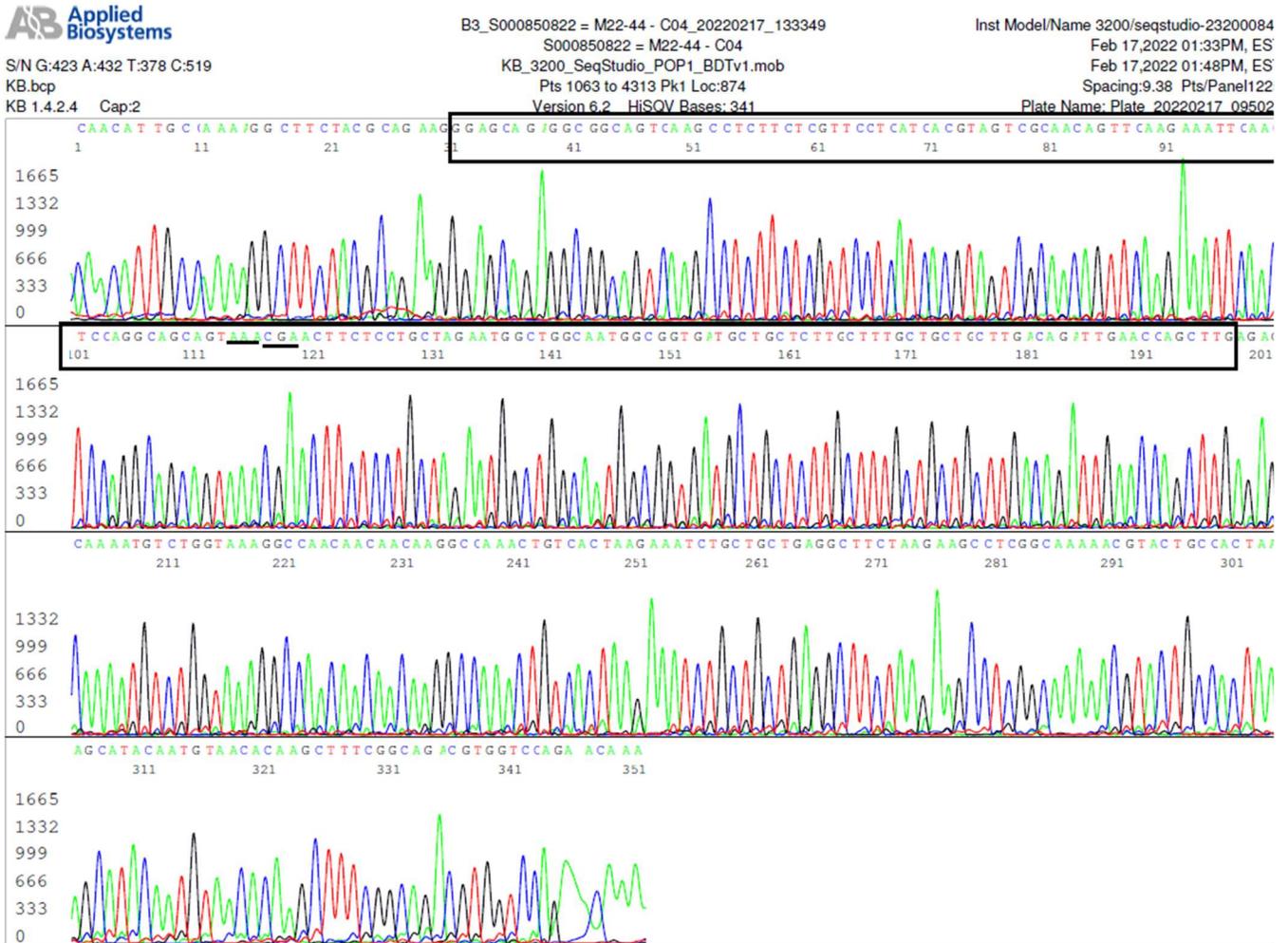


Figure 16 is an electropherogram showing an expected DNA sequence for an Omicron isolate when the same N gene nested PCR products, which were used to generate the sequence presented in Figure 15, were sequenced using the forward Co4 primer as the sequencing primer. As shown in Figure 16, the template sequence has the R203K and G204R mutations (codons underlined), usually present in the Omicron variants. The 168-base stretch of 5’-3’ sequence, which was unreadable in Figure 15, is now framed by two rectangles in Figure 16.

Figure 17

```

28561 atgaaagatc tcagtccaag atggattttc tactacctag gaactgggcc agaagctgga
28621 cttccctatg gtgctaacaa agacggcatc atatgggttg caactgaggg agccttgaat
28681 acaccaaaaag atcacattgg caccgcgaat cctgctaac
      [gap 117 bp]      Expand Ns
28837                                     agtt caagaaatc aactccaggg
28861 agcagtatgg gaacttctcc tgctagtgat gctgctcttg ctttgctgct gcttgacaga
28921 ttgaaccage ttgagagcaa aatgtctggt aaaggccaac aacaacaagg ccaaactgtc
28981 actaagaaat ctgctgctga ggcttctaag aagcctcggc aaaacgtac tgccactaaa
    
```

Figure 17 is a segment of the N gene nucleotide sequence excised from GenBank Seq ID# OV146725, showing a 117-base gap, in which the nucleotide bases could not be determined by DNA sequencing.

3.7. Nontarget PCR amplification of the N gene sequence due to a GGD deletion

On sample M22-31, the N gene nested RT-PCR product formed a weak fluorescent band at agarose gel electrophoresis. The molecular size of the band

was smaller than the others (Figure 3, panel B, lane 13). The results of bidirectional Sanger sequencing of the N gene nested PCR product are presented in Figures 18 and 19.

Figure 18

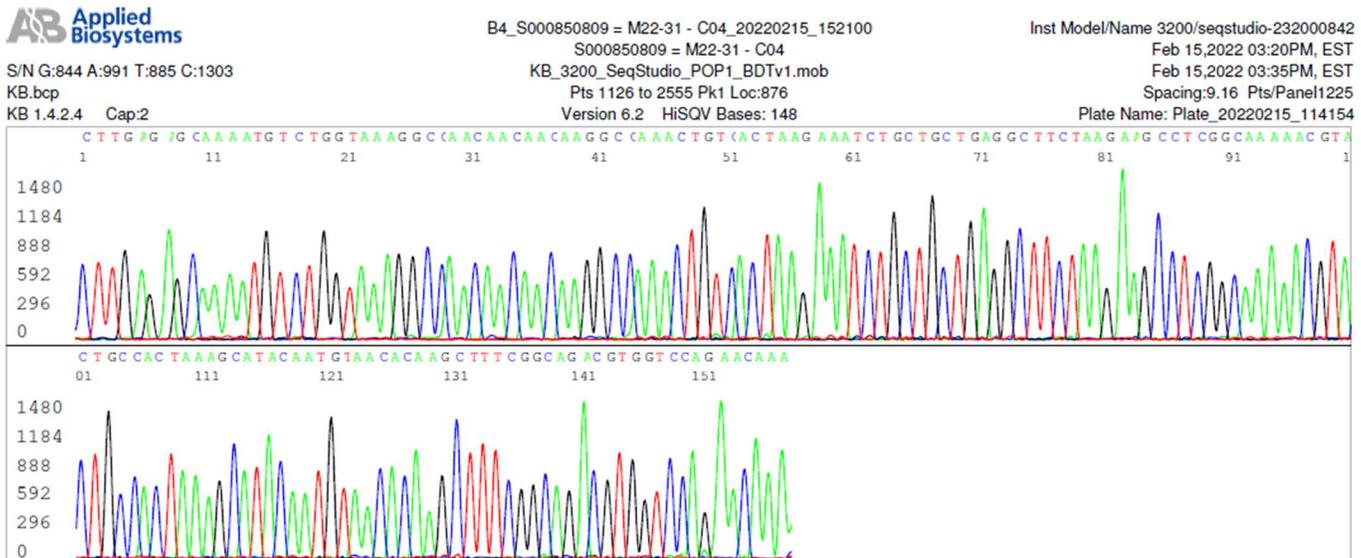
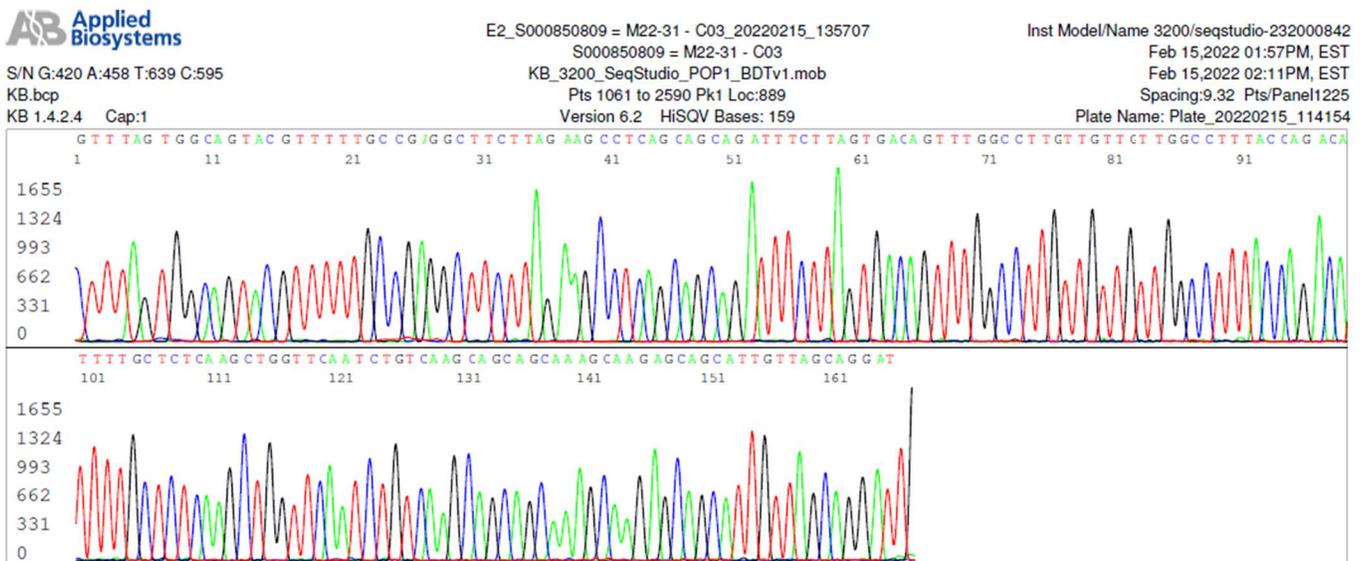


Figure 19



Figures 18 and 19 are electropherograms of the forward (18) and reverse (19) sequencing of the N gene nested PCR products of sample M22-31. The R203 and G204 codons were not included in the PCR amplicon (see Figures 5 and 6).

The 5’-3’ reading composite sequence derived from the electropherograms of Figures 18 and 19 is a 212 bp PCR amplicon with a sequence:

CAATCCTGCTAACAATGCTGCTCTTGCTTT
 GCTGCTGCTTGACAGATTGAACCAGCTTGA
 GAGCAAATGTCTGGTAAAGGCCAACAAAC
 AACAAAGGCCAACTGTCACTAAGAAATCT
 GCTGCTGAGGCTTCTAAGAAGCCTCGGCA
 AAAACGTACTGCCACTAAAGCATAACAATG
 TAACACAAGCTTTCGGCAGACGTGGTCCA
 GAACAAA

Submission of this sequence to the GenBank for BLAST analysis induced a re-turned report shown in Figure 20.

A search of the GenBank database revealed a group of recently submitted SARS-CoV-2 genomic sequences that harbor a 214-216 GGD deletion (Δ 214-216) in the N gene. The deletion of the 214-216 GGD codons created a new 9-base sequence that fully matched the 9-base 3’ terminal sequence of the nested PCR Co4 forward primer (see Figure 21).

The N gene 214-216 GGD deletion is often reported in SARS-CoV-2 isolates with T95I, G142D, E156del, F157del and R158G, the S gene NTD mutations associated with the Delta variant, for example, in GenBank Sequence ID# OL891989, OL451208 and ID# OL553744. The finding of an N gene 214-216 GGD deletion in sample M22-31

raised the possibility of its being a Delta variant, especially when multi-allelic SNPs prevented generation of an unambiguous RBD sequence.

However, a segment of 141-base sequence in the reverse primer sequence of the RBD confirmed that sample M22-31 was indeed an Omicron variant as demonstrated in Figure 22. After this sequence was converted to the 5’-3’ format, it read:

5’-AAACTGGAAATATTGCTGATTATAATT
ATAAATTACCAGATGATTTTACAGGCTGCG
TTATAGCTTGGAAATTCTAACAAGCTTGATT
CTAAGGTTAGTGGTAATTATAATTACCTGT
ATAGATTGTTTAGGAAGTCTAATC

The underlined 138-base sequence encodes amino acids 415-460 of the SARS-CoV-2 S protein TGNIADYNYKLPDDFTGCVIAWNSNKLDSK VSGNYNYLYRFRKSN with K417N, N440K and G446S mutations (underlined) that are characteristic of an Omicron variant.

In addition, the bidirectional sequencing of the NTD confirmed the presence of A67V, Δ 69-70, T95I, G142D and Δ 143-145. One of the sequencing panels showing A67V and Δ 69-70 is presented in Figure 23. Therefore, M22-31 was interpreted as an unusual Omicron BA.1 variant with a 214-216 GGD deletion in its N gene based on information retrieved from the GenBank.

Figure 20

Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/CO 2101858642/2021

Sequence ID: [OL891989.1](#) Length: 29472 Number of Matches: 1

Range 1: 28554 to 28753 [GenBank](#) [Graphics](#) [▼ Next Match](#)

Score	Expect	Identities	Gaps	Strand
370 bits(200)	4e-98	200/200(100%)	0/200(0%)	Plus/Plus
Query 13	CAATGCTGCTCTTGCTTTGCTGCTTGACAGATTGAACCAGCTTGAGAGCAAAATGTC	72		
Sbjct 28554	CAATGCTGCTCTTGCTTTGCTGCTTGACAGATTGAACCAGCTTGAGAGCAAAATGTC	28		
Query 73	TGGTAAAGGCCCAACAACAAGGCCAACTGTCACTAAGAAATCTGCTGCTGAGGCTTC	13		
Sbjct 28614	TGGTAAAGGCCCAACAACAAGGCCAACTGTCACTAAGAAATCTGCTGCTGAGGCTTC	28		
Query 133	TAAGAAGCCTCGGCCAAAAACGTACTGCCACTAAAGCATACAATGTAACACAAGCTTTTCGG	19		
Sbjct 28674	TAAGAAGCCTCGGCCAAAAACGTACTGCCACTAAAGCATACAATGTAACACAAGCTTTTCGG	28		
Query 193	CAGACGTGGTCCAGAACAAA 212			
Sbjct 28734	CAGACGTGGTCCAGAACAAA 28753			

Figure 20. This BLAST report indicates that there is no 100% ID match with the submitted 212-base sequence in the GenBank database. The closest match with the submitted sequence is a 200-base segment of the N gene of a SARS-CoV-2 isolate, GenBank Sequence ID# OL891989, if the first 12 nucleotides of the Co4 forward nested PCR primer were excluded for the sequence alignment.

Figure 21

```

NC_045512.2 N gene sequence
28681 gggagccttg aatacaccaa aagatcacat tggcaccgcg aatectgcta acaatgctgc
28741 aatcgtgcta caacttcctc aaggaacaac attgccaaaa ggcttctacg cagaagggag
28801 cagagggcgc agtcaagcct cttctcgttc ctcatcacgt agtcgcaaca gtccaagaaa
28861 ttcaactcca ggcagcagta ggggaacttc tcctgctaga atggctggga atggcgggtga
28921 tggctgctctt gctttgctgc tgcttgacag attgaaccag cttgagagca aaatgtctgg
28981 taaaggccaa caacaacaag gccaaactgt cactaagaaa tctgctgctg aggcttctaa
29041 gaagcctcgg caaaaacgta ctgccactaa agcatacaat gtaacacaag ctttcggcag
29101 acgtgggtcca gaacaaacc aaaggaaattt tggggaccag gaactaatca gacaaggaac

OL891989.1 N gene sequence with GGD deletion
28321 actgagggag ccttgaatac accaaaagat cacattggca cccgcaatcg tgetaaacaat
28381 gctgcaatcg tgctacaact tcctcaagga acaacattgc caaaaggcct ctacgcagaa
28441 gggagcagag gggcagtc aacctcttct cgttcctcat cacgtagtcg caacagttca
28501 agaaattcaa ctccaggcag cagtatggga acttctcctg ctagaatggc tggcaatgct
28561 gctcttgctt tgctgctgct tgacagattg aaccagcttg agagcaaaaat gtctggtaaa
28621 ggccaacaac aacaaggcca aactgtcact aagaaatctg ctgctgaggc ttctaagaag
28681 cctcggcaaa aacgtactgc cactaaagca tacaatgtaa cacaagcttt cggcagaagt
28741 ggtccagAAC aaaccCAagg aaatTTTggg gaccaggaac taatcagaca aggaactgat
    
```

Figure 21 lists two SARS-CoV-2 N gene segments, one excised from the SARS-CoV-2 Wuhan Hu-1 reference Sequence ID# NC_045512.2 (upper) and the other from Sequence ID# OL891989 (lower). For position identification, the forward and reverse primary RT-PCR primers are highlighted blue, and the forward and reverse nested RT-PCR primers are typed in red on the inner sides of the blue-highlighted primary PCR primers. As shown in the upper sequence, the intended nested PCR amplicon is 398 bp in size, defined by the Co4/Co3 nested PCR primers. The 9-base codons for GGD are shaded gray in the upper sequence. Theoretically, when a 9-base deletion occurs in a template between two PCR primers, the expected amplicon should have reduced by 9 bases to 389 bp in size. However, for sample M-22 31, a 212 bp amplicon was generated instead. That is because a new 9-base sequence, caatgctgc (highlighted green in the lower sequence), fully matching the 3' end sequence of the nested PCR forward primer, was created. After acquiring a new 9-base sequence fully matching the 3' terminus of a primer, a new primer template duplex was formed to initiate a PCR. Given a choice, PCR always favors amplification of a shorter template [34]

Figure 22

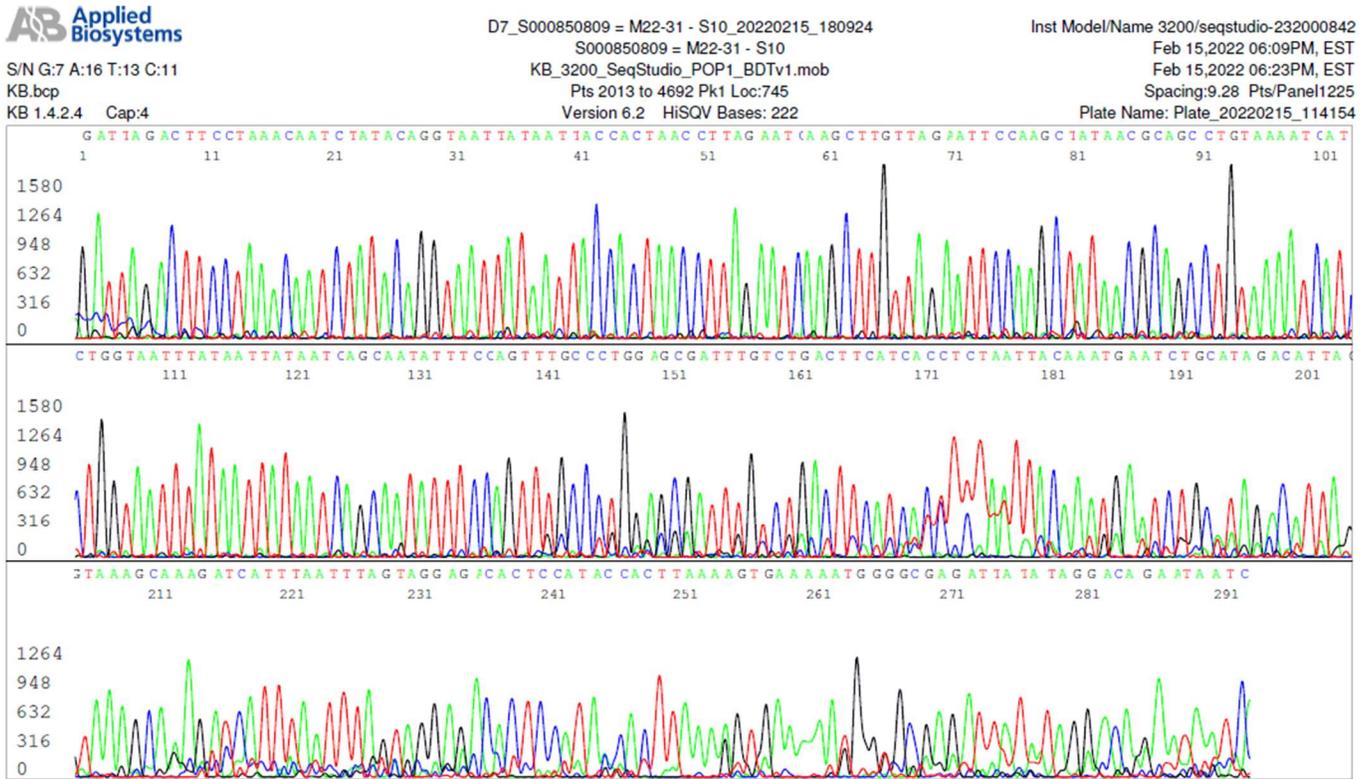


Figure 22. This reverse primer sequencing electropherogram was generated by at least two homologous gene templates, which shared a 141-base common sequence before the heterogeneous base-calling peaks overlapped. The homologous 141-base sequence reads:

3’—GATTAGACTTCCTAAACAATCTATACAGGTAATTATAATTACCACTAACCTTAGAATCAAGCTTGT TAGAATTCCAAGCTATAACGCAGCCTGTAAAATCATCTGGTAATTTATAATTATAATCAGCAATATTC CAGTTT-5’.

Figure 23

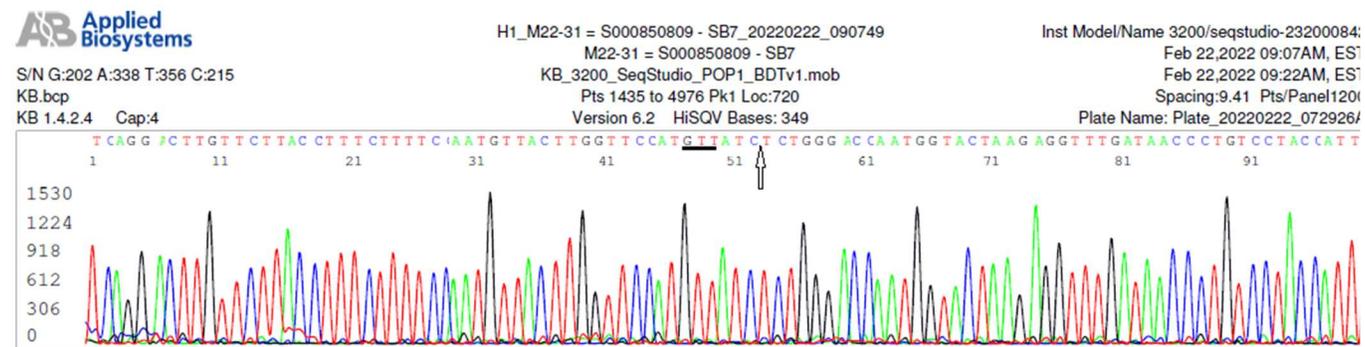


Figure 23. This is an electropherogram showing A67V and Δ69-70, part of the NTD mutations characteristic of an Omicron variant of SARS-CoV-2 in sample M22-31.

3.8. Existence of two competing viruses as cause of S gene sequencing failure

In sample M22-47, there were two competing SARS-CoV-2 viruses, which were demonstrated by bidirectional sequencing of the N gene nested PCR products in Figures 24 and 25.

A search of the GenBank database revealed a group of recently deposited SARS-CoV-2 genomic sequences with R203K, G204R and S183P

mutations in the N gene, such as Sequences ID: OM917790, OM807710, OM657831, OM512484 and OM508240. These isolates all have multiple undetermined stretches of sequences in the S gene. Sample M22-47 harbored at least two competing populations of SARS-CoV-2 Omicron variant, one with a S183P mutation in the N gene that may have multi-allelic SNPs in or around the RBD of the S gene, as shown in Figures 26 and 27.

Figure 24

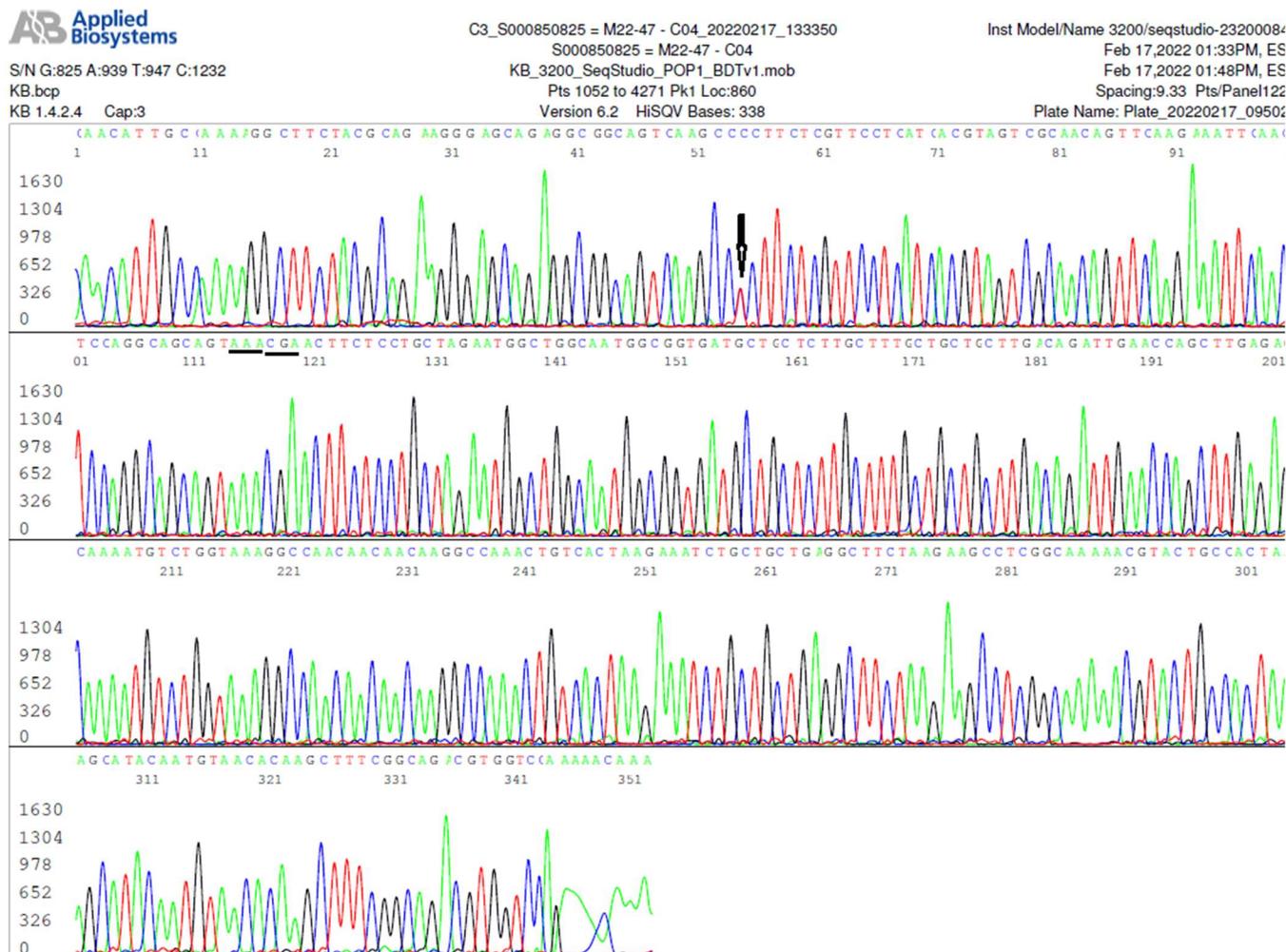


Figure 24 is a forward N gene sequencing electropherogram on sample M-22 47 generated by two competing templates. One of the 2 templates has a T to C mutation at reference position 28820, indicated by an arrow (the computer read the combined T/C peaks as a “C”). A nucleotide T>C mutation in this position changes the codon TCT (serine) to CCT (proline), creating an amino acid mutation S183P. The R203K and G204R mutations for an Omicron variant are underlined.

Figure 25

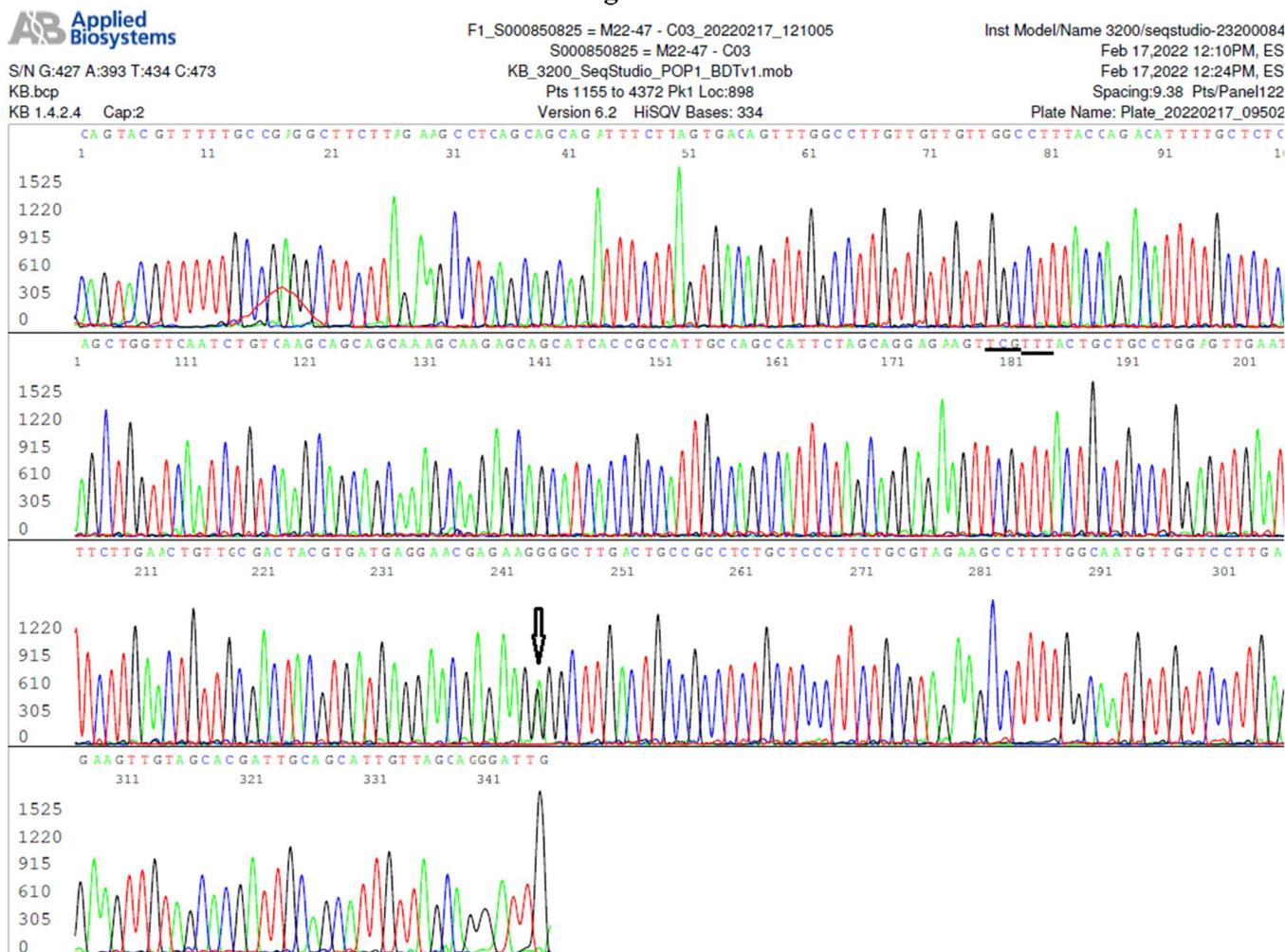


Figure 25 is an electropherogram of the reverse N gene sequencing of the same nested PCR product that was used to generate the electropherogram presented in Figure 24. The mutated nucleotide G peak in the competing template is superimposed on the “A” peak of the parental sequence, pointed by an arrow. The G204R and R203K mutations are underlined.

Figure 26

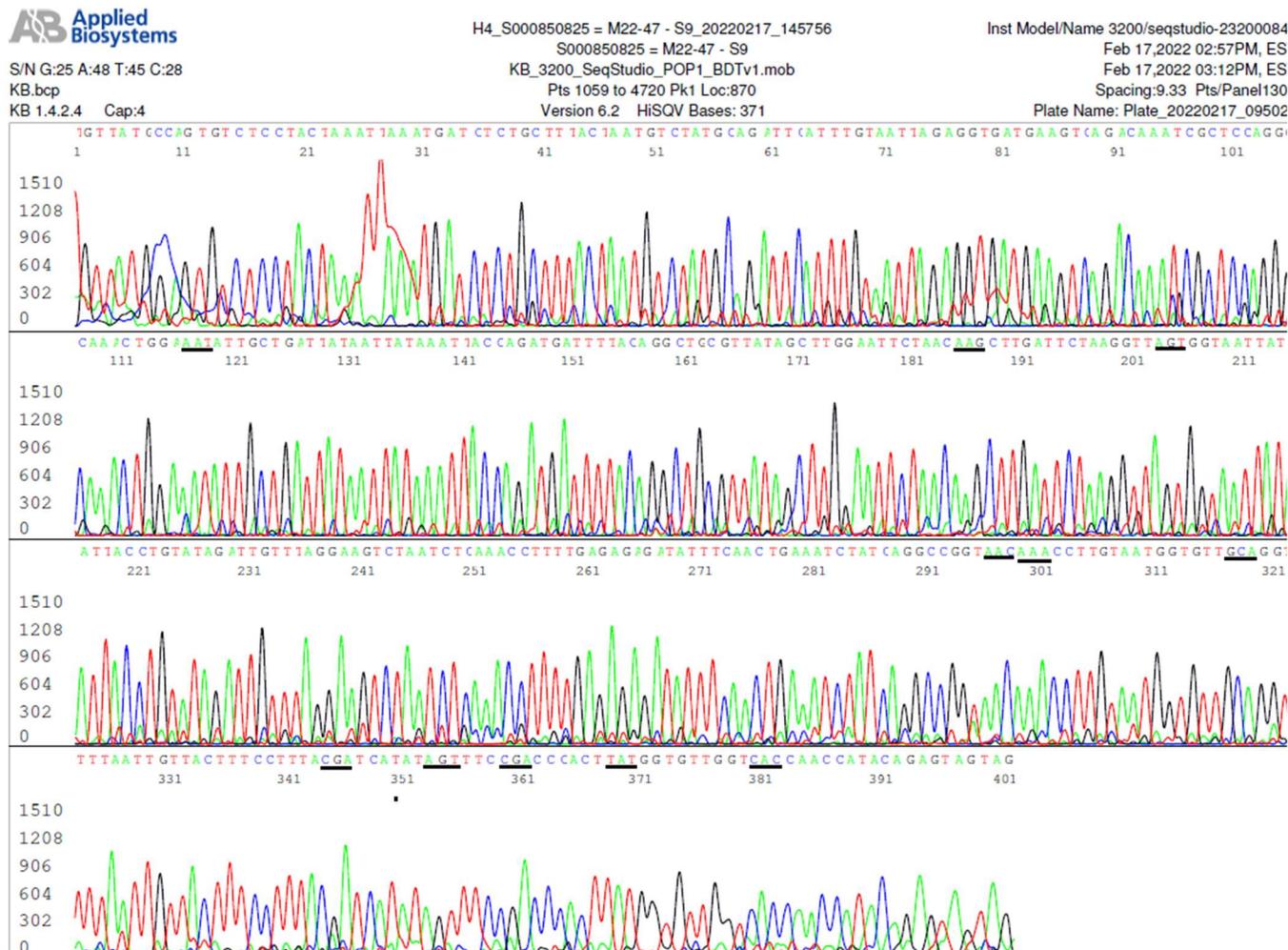


Figure 26 is an electropherogram of the forward primer sequencing of the S gene RBD nested PCR products of sample M22-47 (Figure 3, panel C, lane 29). It shows K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y and Y505H mutations in the dominant sequence, which is diagnostic of an Omicron variant BA.1.

Figure 27

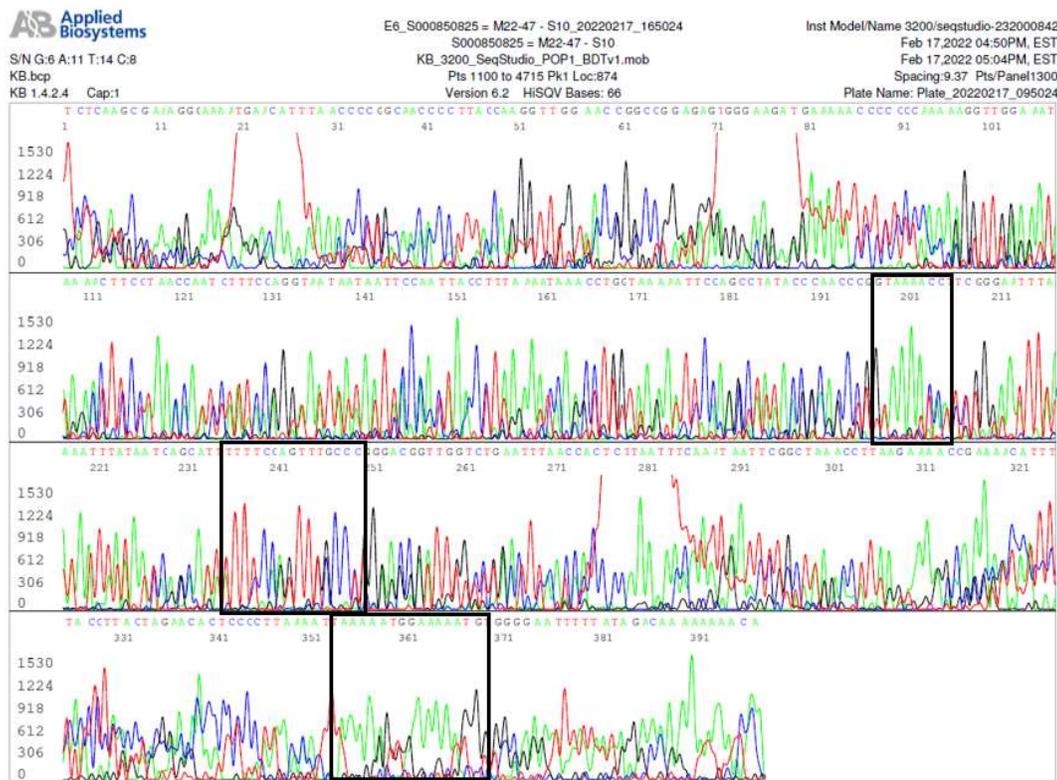


Figure 27 is an electropherogram of the S gene RBD reverse sequencing of the same nested PCR product that was used to generate the electropherogram presented in Figure 26. Accurate base calling was not possible due to multiple overlapping sequences. But the electropherogram showed at least 3 short stretches of sequence (in rectangles) which are characteristic of an S gene RBD of SARS-CoV-2. (Compare this electropherogram with that illustrated in Figure 8A.)

3.9. Unpredictable multi-allelic SNPs prevented S gene RT-PCR amplification

As shown in Figure 3, panels F, G and H, the S gene NTD RT-PCR was negative for samples M22-47, M22-51 and M-68 although the forward sequencing of the RBD cDNA amplicon showed a typical profile of mutations for Omicron variant for sample M22-47 (see Figure 26). To prove that the samples with “non-visible” gel electrophoresis results are in fact free of amplicons, the nested PCR products displaying no visible NTD amplicon band at gel electrophoresis (Figure 3, panels F, G and H) were also sequenced. The results of sequencing the NTD nested PCR products on sample M22-51 are shown in Figure 28.

A new set of nested RT-PCR primers, referred to as the NTD1 primers, was designed in an attempt to amplify a 445-base segment of the S gene immediately upstream of the RBD on samples M22-47, M22-51 and M22-68. The sequence of the primary RT-PCR forward primer is PF1: 5'-TTATGTGGGTTATCTTCAACC; the primary RT-PCR reverse primer is PR2: 5'-AGTTTGCCCTGGAGCGATTG; the nested PCR forward primer is NF3: 5'-GTGGGTTATCTTCAACCTAGG; and the nested PCR reverse primer is NR4: 5'-TTTGCCCTGGAGCGATTGTC. The NTD1 primer RT-PCR conditions were identical to those used for routine testing. The RT-PCR results are presented in a gel image labeled NTD1 (Figure 29).

Figure 28A

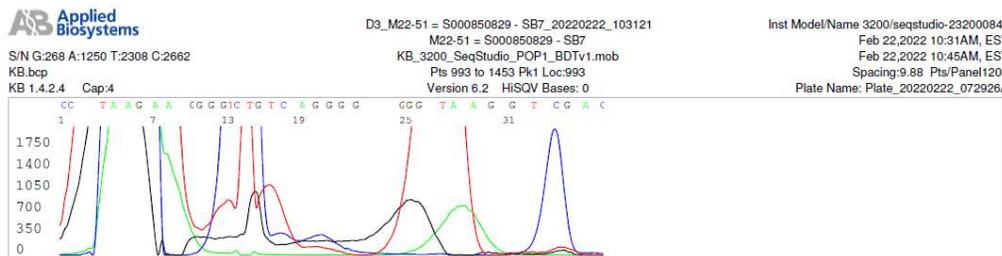
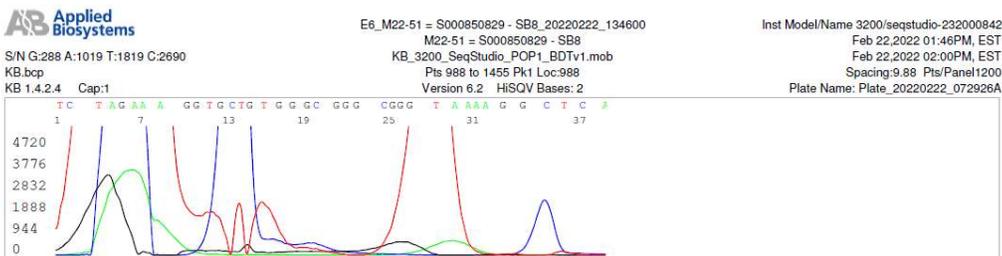


Figure 28B



Figures 28A and 28B. These two bidirectional sequencing electropherograms confirmed that there was no NTD SB7/SB8 nested PCR amplicon on sample M22-51, as shown in Figure 3, Panel G, Lane 51.

Figure 29

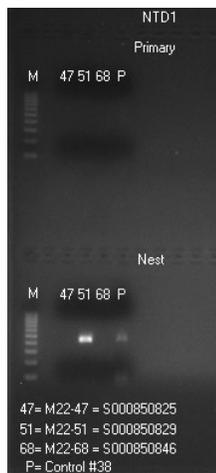


Figure 29 is an image of agarose gel electrophoresis of the RT-PCR products showing that the new set of NTD1 PCR primers was able to amplify a 445-bp segment of the S gene immediately upstream of the RBD on sample M22-51, but not on samples M22-47 and M22-68. A forward primer sequencing verified the authenticity of the RT-PCR product from sample M22-51 (Figure 30).

Three sets of nested RT-PCR primers were used and failed to generate a cDNA amplicon of the RBD or the NTD of the S gene for Sanger sequencing from sample M22-68. Without sequencing information of the S gene RBD or NTD, sample M22-68 was considered as a “presumptive” Omicron variant based on the N gene R203K and G204R mutations only.

In the GenBank sequence database, there are numerous Omicron look-alike isolates that harbor the N gene mutations and the S gene NTD mutations commonly seen in the Omicron variants without the characteristic Omicron mutations in the RBD of the S gene. One of such examples is illustrated by GenBank Sequence ID# OL898842, a specimen collected on 4 December 2021 in Texas, U.S.A. This isolate had the P13L, Δ31- 33, R203K and G204R mutations in the N gene, and the A67V, Δ69-70, T95I, Δ211, L212I, and ins214EPE mutations in the S gene NTD, but not the mutations in the RBD to qualify for an Omicron variant (Figure 31).

Figure 30

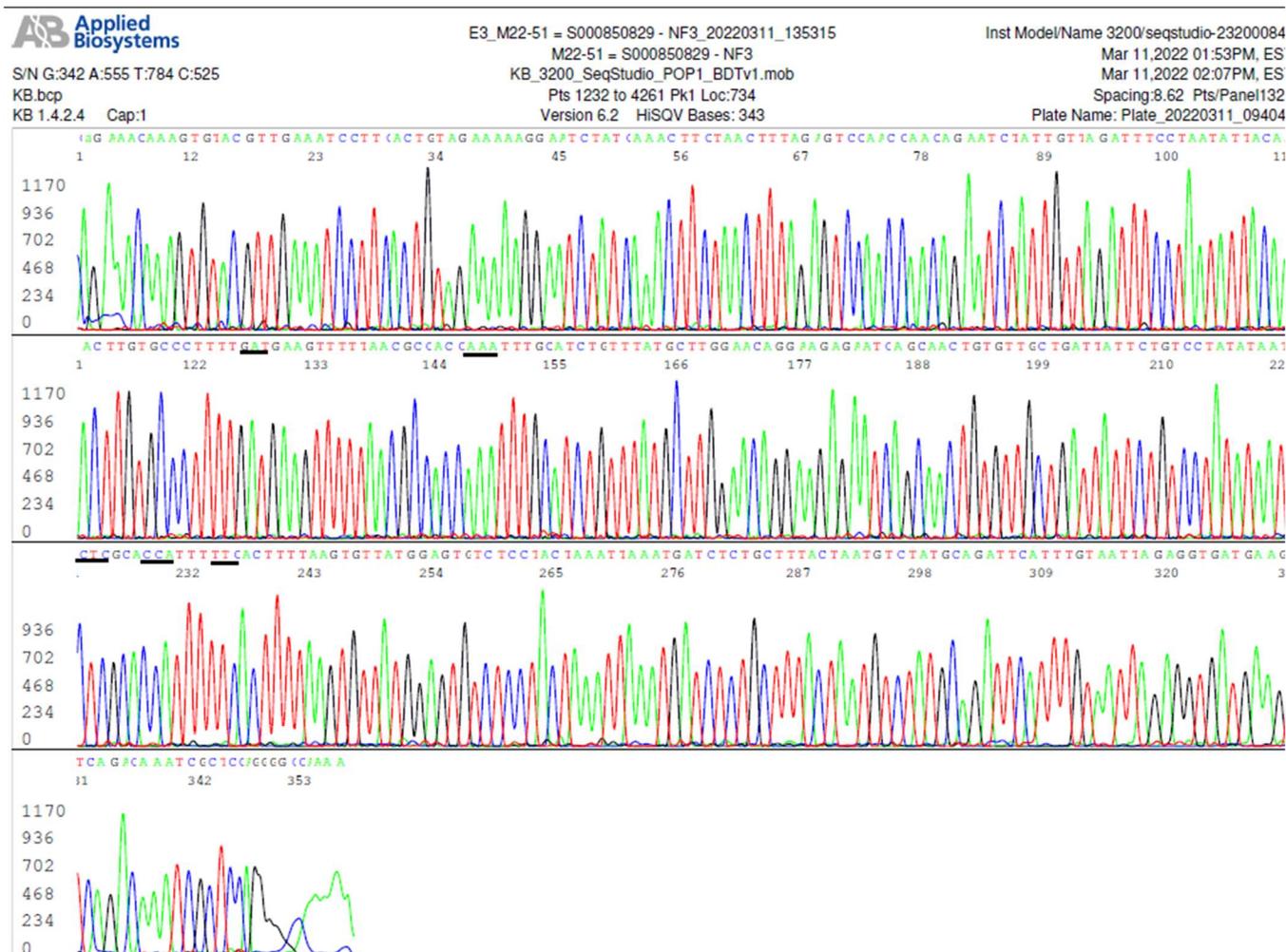


Figure 30 is an electropherogram of the forward sequencing of the sample M22-51 nested RT-PCR amplicon illustrated in Figure 29, using the forward nested PCR NF3 primer as the sequencing primer. It shows G339D (GAT), R346K(AAA), S371L(CTC), S373P(CCA) and S375F(TTC) mutations (codons underlined), which are suggestive of an Omicron variant BA.1 with an additional R346K mutation. However, since the routine RT-PCR primers failed to amplify the key segments of the RBD and NTD in this sample, accurate diagnosis of the subvariant is not possible.

Figure 31

```

21526..25347
/gene="S"
/codon_start=1
/product="surface glycoprotein"
/protein_id="UHB39767.1"

/translation="MFVFLVLLPLVSSQCVNLTTRTQLPPAYTNSFTRGVYYPDKVFR
SSVLHSTQDLFLPFFSNVTWFHVISGTNGTKRFDNPVLPFNDGVYFASIEKSNIIIRGW
IFGTTLDSTQSLIIVNNXXXXVIKXCEFQFCNDPFLGVXYHKNNKSWMEXXRVXXX
XXXXFEYVSQPFXXDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPIIVREPEDLPQ
GFSALEPLVDLPIGINITRFQTLLLALHRSYLTPGDSSSGWTAGAAAYVGYLQPRTFL
LKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITN
LXXXXXXXXXXXXXXXXXXXXNRKRISNXXXXSVLYNSASFSTFKCXXXXXXKLNDLCF
XNVYADSFVXXXXVRQIAPGQTGKXXDYNYKLPXDFTGXXXXXXXXXXXXXXXXKVGGXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXGVEGXXCYFPLQSYGXXXXXXXXXXXX
XXVVLSFXXXXXXXXXXXXXXXXXXXXNLVKNKCVNFNFNXXTGTGVLTESNKKFLPFQQFG
RDIADTTDAVRDPQTLEILDITPCSFGGVSVITPGTNTSNQVAVLYQGVNCTEVPAI
HADQLTPTWRVYSTGSNVFQTRAGCLIGAEYVNNSYECDPIGAGICASYQTQKSHR
RARSVASQSIIAYTXXLGAENSVAYSNNSXXXXNFTIXXTTEILPVXXXXTSVXXXX

```

Figure 31 is an S protein NTD/RBD amino acid sequence retrieved from GenBank Sequence ID# OL898842. The underlined bold letters “VIS”, “I”, “II” and “EPE” marked the sites of mutations “A67V, Δ69-70”, “T95I”, “Δ211, L212I”, and “ins214EPE”, respectively. In the GenBank database, the letter X (typed in red here) is used to highlight the presence of undetermined or variable amino acids, an indication of multi-allelic SNPs in these nucleic acid sequence positions. If these X codon sequences have replaced those in the primer-binding site of the template for the 3’ terminus of a PCR primer, the RT-PCR process will fail.

4 Discussion

PCR was invented to replicate, or to amplify, a target segment of DNA for DNA sequencing without going through a laborious bacterial cloning [35]. PCR needs a pair of primers, single-stranded DNAs of about 20 bases long, to define the segment of target DNA to be replicated. But PCR primer/template hybridization is not fully sequence-specific because PCR primers may attach to non-target DNAs and amplify unwanted DNAs if these DNAs are present and partially match the primers in nucleotide sequence. As a result, relying on PCR, especially the qPCR technology using Ct numbers as the surrogate for actual PCR product analysis, for disease diagnosis is bound to generate false positives. The experimental results of this work emphasize that while RT-qPCR is generating a significant number of false-positive test results at the current stage of the COVID-19 pandemic, the very nature of PCR lacking specificity can be

exploited for designing useful diagnostics for all SARS-related coronaviruses in general if the PCR products are routinely monitored by DNA sequencing. The key points are discussed as follows.

4.1. The COVID-19 pandemic could have been avoided or curtailed by using the SARS-CoV-1 specific RT-PCR primers in early 2020

PCR is a chemical process of primer-initiated template-directed exponential enzymatic polymerization of deoxynucleoside triphosphates (dNTPS) in the test tube. The specificity of the PCR DNA amplification depends on the fidelity of the enzyme, the DNA polymerase whose function is to extend the length of the primer by adding only the correctly matched dNTP to the 3’ end of the primer according to the direction of the template sequence. The binding of a primer to the template, commonly referred to as annealing, is based on hybridization

of two ssDNA fragments, which is a nonspecific process in that a primer can actually bind to a segment of ssDNA with mismatched nucleotides and initiate a PCR. The present study has presented experimental evidence to support the claim that the world could have taken advantage of the partially specific nature of PCR amplification by using the CDC-recommended SARS-CoV-1 specific RT-PCR primers and diagnostic protocol [10] for accurate detection of SARS-CoV-2 at the early stage of the COVID-19 outbreak to avoid or to curtail a pandemic and to lower the death toll. The history of SARS epidemic control in 2003 clearly shows that early detection of positives correctly is of paramount importance to suppress the spread of coronaviruses, ending the SARS epidemic in six months without developing a variant of concern. A set of RT-PCR primers targeting a highly conserved genomic segment of SARS coronaviruses, such as the CDC-recommended SARS-CoV-1 specific RT-PCR primers [10] or the N gene RT-PCR primers presented in this paper, should be available to all major community hospital laboratories in the world in preparation for a timely accurate diagnosis in the next SARS coronavirus outbreak. The hospital laboratories dealing with patients should not wait for the commercial companies to develop an approved test kit to diagnose another emerging SARS coronavirus for early patient treatment and isolation.

It is noteworthy to point out that while the 306-base inter-primer ORF1ab gene sequences defined by primer Cor-p-F3 (+) and primer Cor-p-R1 (-) (Figure 1) in the 16 specimens collected in October 2020 were identical to that of the corresponding segment of the ORF1ab gene sequence of the Wuhan-Hu-1 prototype (GenBank Sequence ID: NC_045512.2), the 398-base N gene sequences defined by the Co4/Co3 primer pair in these 16 samples all showed single nucleotide mutations [15].

4.2. PCR needs DNA sequencing to verify the authenticity of its products in molecular diagnosis

The general assumption that PCR only extends a matched, but not mismatched, nucleotide at the 3' end of a primer is incorrect [36-39]. Using real-time Taqman™ PCR as a model to investigate the effects of primer-template mismatches, a group of investigators showed that a few base mismatches between the primer and the template were well tolerated by the PCR process. Even a nucleotide mismatch at the 3'-terminal position of a primer did not prevent initiation of a real-time PCR but led to an increase of the Ct value by 5.19, on average. Mismatch impact rapidly declined at positions further away from the 3'-terminal position, although there were exceptions [39].

The Sanger sequencing results presented in this paper confirm that the CDC-recommended SARS-CoV-1 Cor-p-R1 (-) reverse PCR primer is able to amplify a corresponding 348-bp target cDNA of the SARS-CoV-2 gene for diagnostic purposes even when there were 3 mismatches in a primer, one of them located at the 3'-terminal position (Figure 1B). But this principle does not apply to RT-qPCR diagnostics, because a 3'-terminal nucleotide mismatch in a primer may boost the Ct value to “negative” territory, a common problem when turning a quantitative test into a qualitative “Yes or No” test. The flaw of the RT-qPCR as a diagnostic assay is that it depends on a number, which may vary from laboratory to laboratory and from test run to test run, to distinguish between the positives and the negatives of a test result. The analyte of PCR is a segment of target DNA, the presence of which can only be verified by demonstrating its nucleotide sequence.

Comparing the N gene reverse nested PCR primer used for this study with the corresponding N gene segment of SARS-CoV-1 (GenBank Seq. ID# AY508724) showed only 1 mismatch located 1 base

away from the 3' terminus of the primer. And there were 2 mismatches located 12 bases away from the 3' terminus in the forward nested PCR primer. Therefore, it is expected that the N gene nested RT-PCR primer set used in this study can also amplify a corresponding 398-bp N gene of the SARS-CoV-1, or of another emerging SARS coronavirus, because these regions of the N gene are highly conserved in this group of viruses.

In the absence of a preferred target template, the DNA polymerase may extend a PCR primer which has attached to a non-target DNA with at least 6 matching bases in its 3' end [40]. For example, the SARS-CoV-2 N gene reverse nested PCR primer has been shown to initiate a PCR amplification of a segment of human chromosome 1 gene due to a 6-base match in its 3' terminus with a human genomic sequence [15], a mechanism that may contribute to the 21 RT-qPCR false-positive reference specimens (Figure 3, panels A-E). According to the FDA advice, false results generated by RT-qPCR assays can be investigated using Sanger sequencing [41].

Non-target DNA amplification by PCR was clearly demonstrated in Figures 18-21, in which a set of PCR primers was found to amplify a shorter DNA segment instead of the fully matched longer target template when the shorter DNA segment offered a 9-base sequence matching the 3' terminal sequence of a PCR primer (Figure 21). PCR always prefers amplification of shorter templates when there is such an option [34].

4.3. The N gene is a more reliable target for RT-PCR detection while partial S gene sequencing is needed for variant determination

Of the 29 specimens collected from patients in the month of January 2022 that were confirmed to be positive for SARS-CoV-2 by partial N gene sequencing, there were 2 from which neither an RBD nor an NTD RT-PCR product band could be generated by a set of PCR primers routinely used

for partial S gene sequencing. Another 2 of the 29 positive samples yielded either a positive RBD RT-PCR product or a positive NTD RT-PCR product, not both (Table 3). These results indicate that 4/29 (13.8%) of the positive samples might be missed if a segment of the S gene were chosen as the only RT-PCR target for COVID-19 diagnosis. The S gene mutation rate is probably much higher than that of the N gene among the Omicron strains.

However, some SARS-CoV-2 isolates with an N gene harboring P13L, Δ 31-33, R203K and G204R mutations may not have a demonstrable RBD mutation profile to support an Omicron variant diagnosis as shown in the GenBank sequences ID# OL898842, OL901854, OL902308 and OL920485 even when the NTD of the S gene in these isolates has been sequenced to show the presence of A67V, Δ 69-70, T95I, G142D and Δ 143-145 mutations, as shown in Figure 31. The N gene R203K and G204R mutations are not reliable for Omicron variant diagnosis because they were already found in the SARS-CoV-2 strains circulating in early 2020 [42] long before the Omicron variant emerged. In the current series, 2 (M22-44 and M22-68) of 29 positive samples did not yield an RBD sequence for a definitive diagnosis of an Omicron variant.

4.4. Multi-allelic SNPs found in Omicron variants

When RNA viruses are allowed to transmit from population to population, genetic change invariably occurs due to RNA polymerase copying errors. In any given SARS-CoV-2 infection, there are probably thousands of viral particles each with unique single-letter mutations [43]. However, only a small fraction of these intra-host single-nucleotide variants become fixed [44], to be passed to the next generation to infect another host. Epidemiological studies often employ per-patient consensus sequences, which summarize each patient's virus population into a single sequence and ignore minor variants. This paper has presented Sanger sequencing

evidence (Figures 11, 12, 13, 15, 22 and 27) for these minor variants, which co-exist with a dominant Omicron variant in single hosts. Although little attention was directed to these minor variants of SARS-CoV-2, intra-host diversity has been shown to affect disease progression [45], transmission risk [46], and treatment outcome [47] in other RNA viruses. The existence of these multi-allelic SNPs involving the RBD of SARS-CoV-2 warrants further investigation.

This study shows that Omicron subvariant sequences with multi-allelic SNPs are commonly found in the S gene RBD and NTD, but only rarely found in the N gene. A high frequency of multi-allelic SNPs may even lower the PCR efficiency to a level at which the S gene PCR products could not form a visible band at electrophoresis but was demonstrated by Sanger sequencing (Figure 13). As previously reported, there were no demonstrable multi-allelic SNPs in the N gene [15] or in the S gene RBD and NTD [25] of the SARS-CoV-2 isolates collected in October 2020. Sequencing of the N gene nested PCR contents without a visible band at agarose gel electrophoresis invariably showed no evidence of an amplification product [15].

4.5. A 42% false discovery rate of RT-qPCR assays for SARS-CoV-2 RNA detection

Real-time or quantitative PCR (qPCR) was first described in 1993 to monitor the accumulation of double-stranded DNA (dsDNA) being generated in each PCR cycle. Results obtained with this approach can quantitate very small numbers of a known dsDNA in the mixture [48] when there are no other interfering DNAs in the system. The analyte is measured relative to a set of standards used to construct a standard curve [49]. However, when qPCR is adapted into a “plus/minus” or a “yes/no” assay for the purpose of detecting genomic DNA of an infectious agent in a complex clinical specimen, it needs to distinguish zero from non-

zero in a standard curve. But in chemical quantitative analysis, the spacing between the zero calibrator and the lowest limit of quantitation of an analyte is extremely difficult to determine [50].

Using qPCR for the diagnosis of infectious diseases, such as Monkeypox virus infections, the CDC requires the testing laboratories to establish their own positive control Ct cut-off value or to prepare a standard curve in order to identify the samples that are truly positive for Monkeypox virus DNAs [51]. However, no such requirement is set for the SARS-CoV-2 RT-qPCR assays [52]. As a result, the diagnostic laboratories do not have a validated quantitative standard curve or a verified Ct cut-off value for SARS-CoV-2 RT-qPCR tests; cut-off values differ from laboratory to laboratory. In some circumstances, the distinction between background noise and actual presence of the target virus is difficult to ascertain [53] in these RT-qPCR assays; a 42% false discovery rate in SARS-CoV-2 RT-qPCR assays is not unexpected. The need for a confirmatory test with 100% specificity was already recognized by the current CDC director 2 years ago [54]. Using RT-qPCR tests with false-positive results to evaluate the endpoint in COVID-19 vaccine development might have artificially inflated the vaccine effectiveness. For example, the COVID-19 vaccine efficacy in the clinical trials was primarily assessed by the results of RT-qPCR testing of placebo participants with minor symptoms [55]. Without confirmatory Sanger sequencing of the RT-qPCR products, the claim of the BNT162b2 vaccine being 95% effective against COVID-19 [56] becomes questionable.

4.6. Limitations of diagnostic testing for SARS-CoV-2 Omicron subvariants

Sanger sequencing of the Spike protein gene RBD and NTD segments has been recommended as a practical means for SARS-CoV-2 variant diagnosis by the European CDC and the WHO [57]. However, there are more than 500 amino acids

encoded by more than 1,500 nucleotides in this region of the S gene, spanning from the beginning of the NTD to the end of the RBD. Since there is a high mutation rate in the RBD and the NTD of the Omicron strains, an enormous number of subvariants have been reported in the literature, with uncertain or unproven clinical significance. Mutations affecting the primer-binding sites may cause S gene RT-PCR failures, as demonstrated in specimens M22-47, M22-51 and M22-68 in this report, although the N gene of these samples can be amplified and sequenced. Moving the S gene PCR primers to another region may amplify an alternative segment. But the alternative sequence may not show the exact anticipated mutation profile for a rigid variant classification, as demonstrated in Figure 30 for M22-51. Figure 30 shows G339D, S371L, S373P, and S375F mutations indicative of an Omicron BA.1 subvariant, but also an additional R346K mutation, which is one of the key mutations in a recently emerging Omicron BF.7 subvariant [58]. Bidirectional sequencing electropherograms confirming the presence of R346K mutation in specimen M22-51 and a novel L84I mutation in the S gene NTD in another BA.4/BA.5 subvariant sample have been previously published [59]. These Sanger sequencing data suggest that the circulating Omicron viruses cannot always be pigeonholed into a rigid subvariant. Despite our desperate, eternal attempt to separate, contain, and mend, categories always leak (Trinh 1989:94) [60].

5 Conclusion

The widely used RT-qPCR assay relying on a Ct number as the surrogate for the physical presence of SARS-CoV-2 nucleic acid in clinical specimens is flawed. This study shows that there are at least 42% false positives in the nasopharyngeal swab samples that were collected and tested in January 2022 and labeled as RT-qPCR positives. However, the nonspecific binding of PCR primers to closely

related nucleic acids can be exploited by using a set of consensus PCR primers to amplify all SARS coronaviruses, including those emerging in the future, provided the PCR products are routinely verified by DNA sequencing. All PCR-positive specimens should be sequenced for verification of the PCR products and for variant determination. Routine sequencing of the RBD and NTD of the S gene can timely discover significant amino acid mutations that have impacts on vaccine efficacies and therapeutics.

6 References

1. Lu R, Zhao X, Li J, Niu P, Yang B, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: Implications for virus origins and receptor binding. *Lancet*. 2020; 395(10224):565-74.
2. Zhou P, Yang X L, Wang X G, Hu B, Zhang L, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020; 579(7798), 270–3.
3. Shang J, Wan Y, Luo C, Ye G, Geng Q, et al. Cell entry mechanisms of SARS-CoV-2. *Proceedings of the National Academy of Science*. USA. 2020; 26;117(21):11727-34.
4. Coronavirus Updates. Worldometer. Available online: <https://www.worldometers.info/coronavirus/>
5. CDC. Severe Acute Respiratory Syndrome (SARS). Available online: <https://www.cdc.gov/sars/about/faq.html> (accessed on 4 March 2022).
6. Abdelrahman Z, Li M, Wang X. Comparative review of SARS-CoV-2, SARS-CoV, MERS-CoV, and Influenza A respiratory viruses. *Frontiers in Immunology*. 2020; 11:552909. <https://doi.org/10.3389/fimmu.2020.552909>

7. Johansson M A, Quandelacy T M, Kada S, Prasad P V, Steele M, et al. SARS-CoV-2 transmission from people without COVID-19 symptoms. *JAMA Network Open*. 2021; 4(1): e2035057. <https://doi.org/10.1001/jamanetworkopen.2020.35057>
8. Zhong NS, Zheng BJ, Li YM, Poon, Xie ZH, et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February 2003. *Lancet*. 2003;362(9393):1353–8. [https://doi.org/10.1016/s0140-6736\(03\)14630-2](https://doi.org/10.1016/s0140-6736(03)14630-2)
9. Drosten C, Preiser W, Günther S, Schmitz H, Doerr HW. Severe acute respiratory syndrome: identification of the etiological agent. *Trends in Molecular Medicine*. 2003;9(8):325-7.
10. CDC. SARS-CoV Specific RT-PCR Primers. Available online: <https://www.who.int/publications/m/item/sars-cov-specific-rt-pcr-primers> (accessed on 8 April 2022).
11. CDC SARS Response Timeline. Available online: <https://www.cdc.gov/about/history/sars/timeline.htm> (accessed on 4 June 2022).
12. Taleghani N, Taghipour F. Diagnosis of COVID-19 for controlling the pandemic: A review of the state-of-the-art. *Biosensors & Bioelectronics*. 2021; 174: 112830. <https://doi.org/10.1016/j.bios.2020.112830>
13. FDA. Nucleic Acid Amplification Assay for the Detection of Enterovirus RNA – Class II Special Controls Guidance for Industry and FDA Staff. <https://www.fda.gov/medical-devices/guidance-documents-medical-devices-and-radiation-emitting-products/nucleic-acid-amplification-assay-detection-enterovirus-rna-class-ii-special-controls-guidance>
14. FDA to CDC. Letter Dated 1 December 2020. <https://www.fda.gov/media/134919/download>
15. Lee SH. Testing for SARS-CoV-2 in cellular components by routine nested RT-PCR followed by DNA sequencing. *International Journal of Geriatrics and Rehabilitation*. 2020; 2: 69–96.
16. Lee SH. qPCR is not PCR just as a straightjacket is not a jacket – The truth revealed by SARS-CoV-2 false-positive test results. *COVID-19 Pandemic: Case Studies & Opinions*. 2021; 2: 230–78. <https://researchinfotext.com/article-details/qPCR-is-not-PCR-Just-as-a-Straightjacket-is-not-a-Jacket-the-Truth-Revealed-by-SARS-CoV-2-False-Positive-Test-Results>
17. Kevadiya BD, Machhi J, Herskovitz J, Oleynikov MD, Blomberg, et al. Diagnostics for SARS-CoV-2 infections. *Nature Materials*. 2021; 20(5):593-605. <https://www.doi.org/10.1038/s41563-020-00906-z>
18. Loeffelholz MJ, Tang YW. Laboratory diagnosis of emerging human coronavirus infections – The state of the art. *Emerging Microbes & Infections*. 2020; 9: 747–756.
19. Zheng L, Wang X, Zhou C, Liu Q, Li S, et al. Analysis of the infection status of healthcare workers in Wuhan during the COVID-19 outbreak: A cross-sectional study. *Clinical Infectious Diseases*. 2020; 71:2109-2113. <https://www.doi.org/10.1093/cid/ciaa588>
20. Liu R, Han H, Liu F, Lv Z, Wu K, et al. Positive rate of RT-PCR detection of SARS-CoV-2 infection in 4880 cases from one hospital in Wuhan, China, from Jan to Feb 2020. *Clinica Chimica Acta*. 2020; 505:172-175. <https://www.doi.org/10.1016/j.cca.2020.03.009>

21. Meng B, Abdullahi A, Ferreira IATM, Goonawardane N, Saito A, et al. Altered TMPRSS2 usage by SARS-CoV-2 Omicron impacts infectivity and fusogenicity. *Nature*. 2022; 603: 706–714. <https://doi.org/10.1038/s41586-022-04474-x>
22. McCallum M, Czudnochowski N, Rosen LE, Zepeda SK, Bowen JE, et al. Structural basis of SARS-CoV-2 Omicron immune evasion and receptor engagement. *Science*. 2022; 375 (6583):864-868. <https://www.doi.org/10.1126/science.abn8652>
23. VanBlargan L A, Errico JM, Halfmann P J, Zost S J, Crowe J E Jr, et al. An infectious SARS-CoV-2 B.1.1.529 Omicron virus escapes neutralization by therapeutic monoclonal antibodies. *Nature Medicine*. 2022; 28:490–495. <https://doi.org/10.1038/s41591-021-01678-y>
24. McCarthy K R, Rennick L J, Nambulli S, Robinson-McCarthy L R, Bain W G, et al. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science*. 2021; 371: 1139–1142. <https://doi.org/10.1126/science.abf6950>
25. Lee SH. A routine Sanger sequencing target specific mutation assay for SARS-CoV-2 variants of concern and interest. *Viruses*. 2021; 13: 2386. <https://doi.org/10.3390/v13122386>
26. CDC. SARS-CoV-2 Variant Classifications and Definitions. Available online: <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html>
27. ECDC. Implications of the emergence and spread of the SARS-CoV-2 B.1.1. 529 variant of concern (Omicron) for the EU/EEA. 26 November 2021. <https://www.ecdc.europa.eu/sites/default/files/documents/Implications-emergence-spread-SARS-CoV-2%20B.1.1.529-variant-concern-Omicron-for-the-EU-EEA-Nov2021.pdf>
28. CoV-lineages Github. Proposal to split B.1.1.529 to incorporate a newly characterised sibling lineage #361. <https://github.com/cov-lineages/pango-designation/issues/361>
29. Lee SH. Lyme disease caused by *Borrelia burgdorferi* with two homeologous 16S rRNA genes: A case report. *International Medical Case Reports Journal*. 2016; 9: 101–106. <https://doi.org/10.2147/IMCRJ.S99936>
30. Walker A, Houwaart T, Wienemann T, Vasconcelos MK, Strelow D, et al. Genetic structure of SARS-CoV-2 reflects clonal superspreading and multiple independent introduction events, North-Rhine Westphalia, Germany, February and March 2020. *Eurosurveillance*. 2020; 25(22):2000746.
31. Gupta K, Toelzer C, Williamson MK, Shoemark DK, Oliveira ASF, et al. Structural insights in cell-type specific evolution of intra-host diversity by SARS-CoV-2. *Nature Communications*. 2022; 13(1): 222. <https://doi.org/10.1038/s41467-021-27881-6>
32. Wang Y, Wang D, Zhang L, Sun W, Zhang Z, et al. Intra-host variation and evolutionary dynamics of SARS-CoV-2 populations in COVID-19 patients. *Genomic Medicine*. 2021; 13(1): 30. <https://doi.org/10.1186/s13073-021-00847-5>
33. Mushegian AA, Long SW, Olsen RJ, Christensen PA, Subedi S, et al. Within-host genetic diversity of SARS-CoV-2 in the context of large-scale hospital-associated genomic surveillance. medRxiv preprint. 2022. <https://doi.org/10.1101/2022.08.17.22278898>
34. Shagin DA, Lukyanov KA, Vagner LL, Matz MV. Regulation of average length of complex PCR product. *Nucleic Acids Research*. 1999; 27(18):e23.
35. Appenzeller, T. Democratizing the DNA sequence. *Science*. 1990; 247: 1030-1032.

36. Kwok S, Kellogg DE, McKinney N, Spasic D, Goda L, et al. Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Research*. 1990; 18(4): 999-1005.
37. O'Dell SD, Humphries SE, Day IN. PCR induction of a TaqI restriction site at any CpG dinucleotide using two mismatched primers (CpG-PCR). *Genome Research*. 1996; 6:558-68. <https://doi.org/10.1101/gr.6.6.558>
38. Ruiz-Villalba A, van Pelt-Verkuil E, Gunst QD, Ruijter JM, van den Hoff MJ. Amplification of nonspecific products in quantitative polymerase chain reactions (qPCR). *Biomolecular Detection and Quantification*. 2017; 14:7-18. <https://doi.org/10.1016/j.bdq.2017.10.001>
39. Stadhouders R, Pas SD, Anber J, Voermans J, Mes TH, et al. The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5' nuclease assay. *Journal of Molecular Diagnostics*. 2010; 12(1):109-17. <https://doi.org/10.2353/jmoldx.2010.090035>
40. Ryu KH, Choi SH, Lee JS. Restriction primers as short as 6-mers for PCR amplification of bacterial and plant genomic DNA and plant viral RNA. *Molecular Biotechnology*. 2000; 14: 1-3.
41. FDA. In Vitro Diagnostics EUAs. Molecular Diagnostic Template for Laboratories. <https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizationsmedical-devices/vitro-diagnostics-euas> (accessed on 19 January 2021).
42. Narayanan S, Ritchey JC, Patil G, Narasaraju T, More S, et al. SARS-CoV-2 Genomes from Oklahoma, United States. *Frontiers in Genetics*. 2021;11:612571. <https://doi.org/10.3389/fgene.2020.612571>
43. Callaway E. Beyond Omicron: What's next for COVID's viral evolution. *Nature*. 2021; 600 (7888):204-207. <https://doi.org/10.1038/d41586-021-03619-8>
44. Li J, Du P, Yang L, Zhang J, Song C, et al. Two-step fitness selection for intra-host variations in SARS-CoV-2. *Cell Reports*. 2022; 38(2): 110205. <https://www.doi.org/10.1016/j.celrep.2021.110205>
45. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature*. 2006; 439(7074): 344-8. <https://www.doi.org/10.1038/nature04388>
46. Poon LL, Song T, Rosenfeld R, Lin X, Rogers MB, et al. Quantifying influenza virus diversity and transmission in humans. *Nature Genetics*. 2016; 48(2):195-200. <https://www.doi.org/10.1038/ng.3479>
47. To KK, Chan JF, Chen H, Li L, Yuen KY. The emergence of influenza A H7N9 in human beings 16 years after influenza A H5N1: A tale of two cities. *Lancet Infectious Diseases*. 2013; 13(9):809-21. [https://www.doi.org/10.1016/S1473-3099\(13\)70167-1](https://www.doi.org/10.1016/S1473-3099(13)70167-1)
48. Higuchi R, Fockler C, Dollinger G, Watson R. Kinetic PCR analysis: Real-time monitoring of DNA amplification reactions. *Biotechnology*. 1993; 11: 1026-1030.
49. Svec D, Tichopad A, Novosadova V, Pfaffl MW, Kubista M. How good is a PCR efficiency estimate: Recommendations for precise and robust qPCR efficiency assessments. *Biomolecular Detection and Quantification*. 2015; 3:9-16.

- <https://www.doi.org/10.1016/j.bdq.2015.01.005>
50. Azadeh M, Sondag P, Wang Y, Raines M, Sailstad J. Quality controls in ligand binding assays: Recommendations and best practices for preparation, qualification, maintenance of lot to lot consistency, and prevention of assay drift. *AAPS Journal*. 2019; 21(5):89. <https://www.doi.org/10.1208/s12248-019-0354-6>
51. Centers for Disease Control & Prevention, Poxvirus & Rabies Branch (PRB). Test procedure: Monkeypox virus generic real-time PCR test. <https://www.cdc.gov/poxvirus/monkeypox/pdf/pcr-diagnostic-protocol-508.pdf>
52. CDC. 2019–Novel Coronavirus (2019-nCoV) real-time RT-PCR diagnostic panel: Instructions for use. <https://www.fda.gov/media/134922/download>
53. WHO. Information notice for IVD users – Nucleic acid testing (NAT) technologies that use real-time polymerase chain reaction (RT-PCR) for detection of SARS-CoV-2. <https://www.who.int/news/item/14-12-2020-who-information-notice-for-ivd-users> (Accessed 20 February, 2021.) Archived at: <https://bit.ly/3zIYJVZ>
54. Paltiel AD, Zheng A, Walensky RP. Assessment of SARS-CoV-2 screening strategies to permit the safe reopening of college campuses in the United States. *JAMA Network Open*. 2020; 3(7):e2016818. Published 2020, 1 Jul. <https://www.doi.org/10.1001/jamanetworkopen.2020.16818>
55. Pfizer Inc. PF-07302048 (BNT162 RNA-Based COVID-19 Vaccines) Protocol C4591001 https://cdn.pfizer.com/pfizercom/2020-11/C4591001_Clinical_Protocol_Nov2020.pdf
56. Pfizer Inc. Pfizer and BioNTech conclude Phase 3 study of COVID-19 vaccine candidate, meeting all primary efficacy endpoints. <https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-conclude-phase-3-study-covid-19-vaccine>
57. ECDC and WHO Regional Office for Europe. Methods for the detection and characterisation of SARS-CoV-2 variants – First update. <https://www.ecdc.europa.eu/en/publications-data/methods-detection-and-characterisation-sars-cov-2-variants-first-update>
58. WHO. Tracking SARS-CoV-2 variants. <https://www.who.int/activities/tracking-SARS-CoV-2-variants>
59. Lee, S. Sanger sequencing for molecular diagnosis of SARS-CoV-2 Omicron subvariants and its challenges. *Journal of Biosciences and Medicines*. 2022; 10: 182-223. <https://www.doi.org/10.4236/jbm.2022.109015>
60. Clarke AE, Casper MJ. From simple technology to complex arena: Classification of Pap smears, 1917-90. *Medical Anthropology Quarterly*. 1996; 10:601-23.

7 Author Statements

Funding: This study was funded in part by The Institute for Pure and Applied Knowledge via the Nucleic Acid Assay Technology Evaluation Consortium.

Institutional Review Board Statement: Material supplier, Boca Biolistics, LLC (Pompano Beach, FL, USA) has provided a statement of Independent Investigational Review Board, Inc. (Columbia, MD, USA) SOP 10-00414 Rev E (De-Linking Specimens).

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The author thanks Wilda Garayua for her technical assistance.

Potential conflicts of interest: Dr. Lee served as a technical advisor to the IPAK NAATEC Consortium.

EDITOR’S NOTE: This article has been updated to change the incorrect reference to “false positive rate” to “false discovery rate”. This change does not alter the conclusions or interpretation of the study.