

Article

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

Sin Hang Lee

Milford Molecular Diagnostics Laboratory, 2044 Bridgeport Avenue, Milford, CT 06460, USA; shlee01@snet.net

Abstract: Both SARS-CoV-2 and SARS-CoV 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 policy-makers. This article presented Sanger sequencing evidence to show that the RT-PCR diagnostic protocol established in 2003 for SARS-CoV can in fact detect SARS-CoV-2 accurately due to the well-known nonspecific PCR amplification of DNAs with similar 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 21 (42%) were false-positive. Although the other 29 positive isolates were categorized as Omicron variant by partial sequencing of the N gene, and the RBD and the NTD of the S gene, 9 (31%) showed focal to complete sequencing failure in the S gene segments due to multi-allelic SNPs. During the course of the study, an Omicron variant isolate containing a BA.1 NTD and a BA.2 RBD in its S gene was also detected. Timely routine partial S gene sequencing of all PCR-positive samples can discover multi-allelic SNPs and viral recombination in the circulating variants for investigation of their impacts on vaccine efficacies, therapeutics and diagnostics.

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

1. Introduction

The SARS-CoV-2 virus that causes the COVID-19 pandemic is genetically closely related to the SARS-CoV 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 8098 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.

Although the SARS-CoV-2 is less lethal than the SARS-CoV, in the past two years the COVID-19 pandemic has resulted in massive loss of life and an unprecedented economic crisis with far-reaching social impacts because the SARS-CoV-2 is a highly transmissible and more contagious virus [6,7]. Cell culture studies showed that SARS-CoV-2 spreads through cell-to-cell contact, which is mediated by the spike glycoprotein; and in comparison, the SARS-CoV-2 spike protein is more efficient in facilitating cell-to-cell transmission than is the SARS-CoV spike [8].

When RNA viruses are allowed to transmit from population to population, genetic changes invariably occur due to RNA polymerase copying errors, which may lead to single nucleotide nonsynonymous mutations and indel mutations, creating new variants; and the new variants may be more transmissible than their predecessors. The most recently emerging Omicron variant with its highest number of amino acid mutations is even more contagious than the earlier variants of concern [9]. According to the U.S. Centers for Disease Control and Prevention (CDC) surveillance data, as of the week ending January 22, 2022, Omicron was estimated to account for 99.2% of SARS-CoV-2 infections nationwide [10].

The appearance of the SARS-CoV-2 Alpha, Beta, Gamma, Delta and Omicron variants among many others since October 2020 confirms that allowing an RNA virus to continue circulating unchecked in the populations will eventually lead to mutations and affect virus transmissibility, disease severity, and capacity for immune evasion, which may in turn generate more mutations and variants. To break this vicious cycle, “we need targeted testing, contact tracing, and proper support for self-isolation. Without these seemingly obvious traditional public health steps, the pandemic will continue to worsen our longstanding social divides”, pointed out in July 2021 by the then BMJ editor in chief [11].

However, there is no diagnostic “target testing” for SARS-CoV-2 because in the United States the RT-qPCR assays were only granted emergency use authorization (EUA) by the Food and Drug Administration (FDA) “for the presumptive qualitative detection of nucleic acid from the 2019-nCoV in upper and lower respiratory specimens”, since February 4, 2020 [12]. By definition, a presumptive test result is not diagnostic of any diseases in medical practice because it implies that a final confirming test is pending, which may turn out to be different from the presumptive result in providing a negative or a positive answer.

The world dealt with the 2002/2003 SARS outbreak differently and successfully, without depending on a commercial presumptive diagnostic test kit, and had the SARS epidemic under control in six months. Public record shows that 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 definitive molecular diagnosis [13]. According to one report, the U.S. CDC-designed PCR primers directed to the polymerase gene of all coronaviruses and amplified a 405 bp fragment from the new agent. The fragment was then sequenced and compared with the GenBank reference sequences for molecular diagnosis [14]. In another document, the CDC developed a standard diagnostic protocol for SARS-CoV, which 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” [15]. With accurate diagnoses based on DNA sequencing, prompt isolation of patients and early treatment, the SARS outbreak ended in July [16] before a variant of concern was developed; the pandemic was stopped in 2003 by applying travel restrictions and isolating individuals infected by SARS-CoV [17].

As the Omicron variant becomes the dominant strain in the U.S., some of the probe-based RT-qPCR test kits designed to detect a short nucleic acid sequence of the Wuhan-Hu-1 prototype SARS CoV-2 (GenBank Sequence ID# NC_045512.2) are expected to fail. For example, the DNA sequence of the N1 probe of the CDC 2019-nCoV Real-Time RT-PCR Diagnostic Panel is FAM-ACC CCG CAT /ZEN/TAC GTT TGG TGG ACC-3IABkFQ [18]. However, all Omicron variant strains have an N gene P13L (CCC>CTC) mutation [19], which has changed the second and underlined nucleotide C of the N1 probe to T, creating a nucleotide mismatch in the probe between the ZEN™ Internal Quencher and the fluorophore reporter FAM. As a result, the 5'-end part of the probe including the internal quencher may not hybridize to its target DNA during the PCR cycling and escape the 5'-to-3' exonuclease activity of the Taq polymerase, leading to a false-negative test result.

On 27 December, 2021, the FDA officially announced that for the Omicron variant some RT-qPCR test kits for the detection of SARS-CoV-2 are expected to fail due to deletions at S protein amino acid positions 69-70 and mutations at nucleotide positions 23599 (T to G) and 23604 (C to A), and due to a nine-nucleotide deletion in the N-gene, spanning positions 28370-28362 [20].

If 99.2% of the current SARS-CoV-2 infections are truly caused by Omicron variants as the CDC estimated, all the RT-qPCR test kits previously granted EUA for the presumptive qualitative detection of the original SARS-CoV-2 Wuhan Hu-1 nucleic acid sequence [12, 21] must be re-evaluated and confirmed that they are in fact capable of detecting and diagnosing Omicron variants. Continued generation of false-negative and false-positive coronavirus test results will cause further confusion among policymakers as society attempts to return to pre-pandemic normalcy.

In the United States, there are no FDA-authorized, cleared, or approved diagnostic tests to specifically detect SARS-CoV-2 variants (Omicron or other variants). Currently, commercial COVID-19 test kits are designed and authorized by the FDA to check broadly for the SARS-CoV-2 virus, not for specific variants [22].

The purpose of this study was to introduce a generic RT-PCR with amplicon sequencing as recommended by the CDC in 2003 [15] for accurate diagnosis of SARS-CoV-2, including determination of its variants. As the SARS-CoV-2 Wuhan Hu-1 prototype and its subsequent variants have been allowed to circulate for more than 2 years from population to population, multi-allelic SNPs [23] have been accumulated in the virus and even showed up in some of the emerging variant isolates displaying genetic diversity within single infected hosts [24, 25]. The impact of these multi-allelic SNPs on variant diagnosis is explored along with the data generated by Sanger sequencing.

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 were verified to contain a 398-base segment of the SARS-CoV-2 N gene by Sanger sequencing [26], were used to evaluate the effectiveness of the SARS-CoV PCR primers [15] in detecting SARS-CoV-2 genomic RNA. These latter 16 specimens were the true-positives among 30, which were collected in October, 2020 and were previously tested positive for SARS-CoV-2 by an RT-qPCR assay [26].

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 U.S. Food and Drug Administration (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 [25-27]. 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 μ L of digestion buffer containing 1% sodium dodecyl sulfate, 20mM Tris-HCl (pH 7.6), 0.2M NaCl and 700 μ g/mL proteinase K, was added. The mixture was digested at 47°C for 1 hr in a shaker. An equal volume (400 μ L) 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). 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 μ molar forward primer in TE buffer, 1 μL of 10 μ molar reverse primer 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 μ molar forward primer, 1 μL of 10 μ molar 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 μ molar sequencing primer, 1 μL of BigDye® Terminator (v 1.1/Sequencing Standard Kit), 3.5 μL 5 \times 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 on-line BLAST. The sequences were also visually analyzed for nucleotide mutations and indels.

2.5. PCR primers

The sequences of the 3 heminested RT-PCR primers used to generate a 348-bp amplicon of the SARS-CoV-2 ORF1ab gene were listed in a CDC document [15]. 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 [26–28] 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 CDC 2003 heminested	Cor-p-F2 (+) F.			CTAACATGCTTAGGATAATGG	368
	Cor-p-R1 (-) R.			CAGGTAAGCGTAAACATCATC	
	Cor-p-F3 (+) F.			GCCTCTCTGTTCTGCTCGC	
	Cor-p-R1 (-) R.			CAGGTAAGCGTAAACATCATC	
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	CAATCCTGCTAACAAATGCTGC	
	Co3 nested R.	29097	29117	TTTGTCTGGACCACGTCTGC	
SARS-CoV-2 S gene RBD S9/S10 Nested	SS1 primary F.	22643	22663	TGTGTTGCTGATTATTCTGTC	460
	SS2 primary R.	23082	23102	AAAGTACTACTACTCTGTATG	
	S9 nested F.	22652	22672	GATTATTCTGCTCTATATAAT	
	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	TTTGAAATTACCCCTGTTTCC	
	SB7 nested F.	21628	22648	TCAATTACCCCTGCATACAC	
	SB8 nested R.	22097	22117	ATTACCCCTGTTTCCTTCAAG	

Table 1 summarized the 4 sets of PCR primers used in this study. The 3 heminested PCR primers initially designed for SARS-CoV were copied from a CDC publication [15]. The others were also previously published [26–28] but slightly modified to accommodate mutations in the S gene of emerging Omicron variants. The intended nested PCR amplicon size is underlines. 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 BA.2 subvariants. The sequences of the new general primer set for the S gene NTD amplification are:

SB11 5'-TCTCTAGTCAGTGTGTTAAC-3' Primary Forward

SB6 5'-TTGAAATTACCCCTGTTTCC-3' Primary Reverse

SB12 5'-TTAATCTTACAACCAAGAACTC-3' Nested Forward

SB8 5'-ATTACCCCTGTTTCCTTCAAG-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 [29–31].

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, G496S, 498R, 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, T75I	N. A.

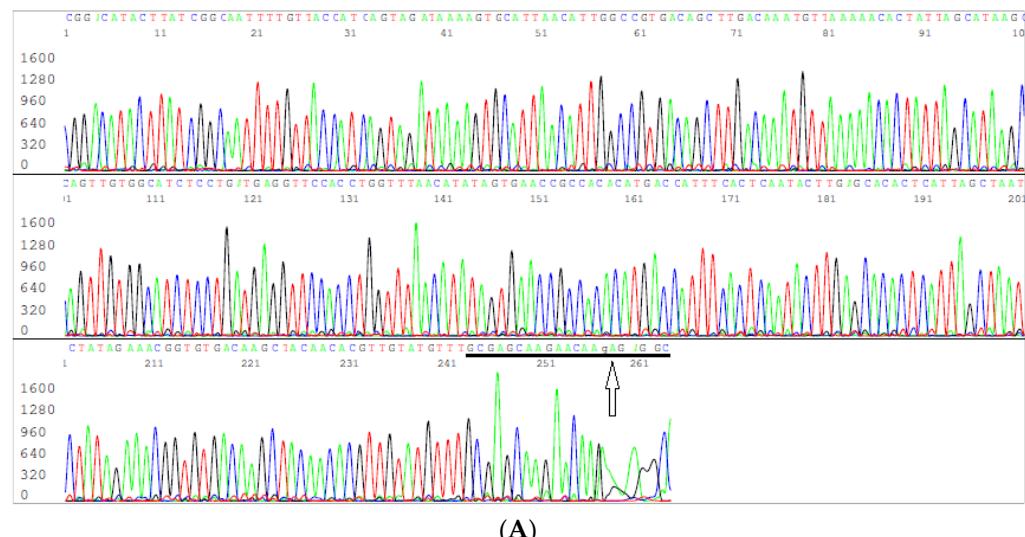
Table 2 shows that sequencing the 445-bp ACE2 RBD nested PCR amplicon can detect the key amino acid mutations from S371 to Y505. The combination patterns of these RBD mutations with additional information from the NTD sequencing can reliably diagnose all major variants of concern and variants of interest.

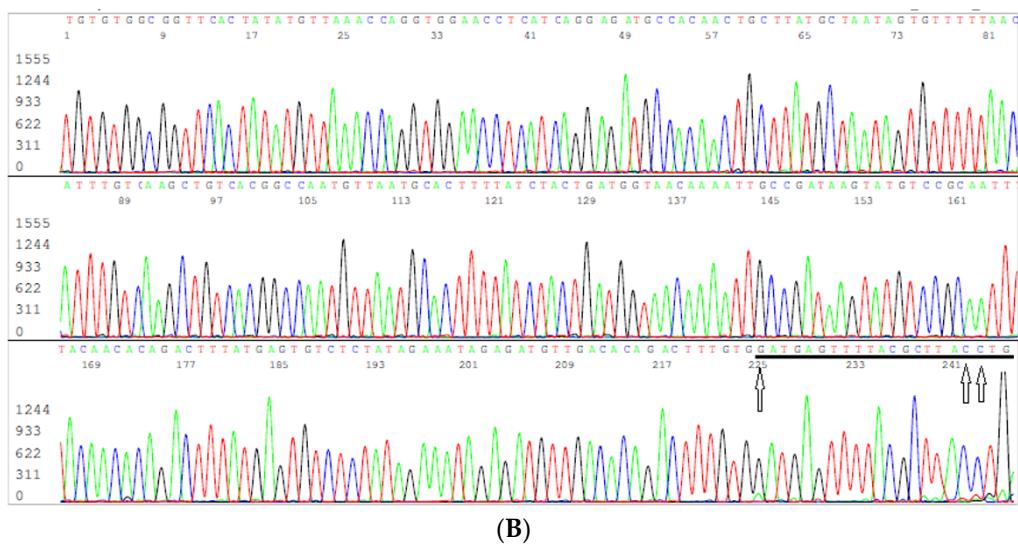
3. Results

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 section.

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

Sixteen (16) SARS-CoV-2 positive samples collected in October 2020 were selected for heminested RT-PCR amplification with the 3 PCR primers, which the CDC designed and recommended for SARS-CoV specific RT-PCR diagnosis in 2003 [15]. They all generated a 348-bp amplicon with an identical sequence. One of the 16 pairs of bidirectional sequencing electropherograms is presented in Figure 1 A and B for illustration.





(B)

Figure 1. A and B. The two computer-generated bidirectional sequencing electropherograms presented above show the 3'-5' sequence of a SARS-CoV-2 gene RT-PCR amplicon, using the CDC-recommended SARS-CoV Cor-p-R1 (-) reverse PCR primer 5'-CAGGTAAGCGTAAACATC-3' as the sequencing primer (Figure 1 A), and the 5'-3' sequence of the same amplicon, using the CDC-recommended forward PCR primer Cor-p-F3 (+) 5'-GCCTCTTGTCTGCTCG-3' as the sequencing primer (Figure 1 B), respectively. The sample was one of the 16 positive nasopharyngeal specimens collected in October, 2020. 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 1 A) and 3 mismatches are in the reverse primer (Figure 1 B). One of the mismatched nucleotides, a base G, is located in the 3' end of the reverse primer (Figure 1 B).

The 5'-3' composite sequence derived from the two electropherograms presented in Figure 1 is as follows:

GCCTCTTTGTTCTGCTCGAACATACAAACGTGTTAGCTGTCACACCGTTCTATAGATTAGCTAATGAGTGTGCTCAAGTA
TTGAGTGAATGGTCATGTGTGGCGGTTCACTATATGTTAACCCAGGTGGAACCTCATCAGGAGATGCCACAATGCTTATGCTAAT
AGTGTGTTAACATTGTCAAGCTGTCACGGCCAATGTTAATGCACCTTATCTACTGATGGTAACAAATTGCCATAAGTATGTC
CGCAATTACACACAGACTTATGAGTGTCTATAGAAATAGAGATGTTGACACAGACTTGTGGATGAGTTACGCTTACCTG

Submission of this 348-base sequence for BLAST alignment analysis showed that it matched numerous SARS-CoV-2 ORF1ab gene sequences recently deposited in the GenBank. One of the matches is presented in Figure 2A, a segment of SARS-CoV-2 ORF1ab gene sequence derived from a sample collected in Minnesota, U.S.A. on 30 January, 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.



Figure 2. (A) 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 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. (B) 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 inter-primer sequence presented in Figure 1 A and B. The 3 CDC-recommended SARS-CoV specific RT-PCR primer sequence sites are shaded gray or typed in red color. The mismatched nucleotides between the SARS-CoV 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). (C) is part of a SARS-CoV-2 ORF1ab gene sequence retrieved from the GenBank 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 Specific RT-PCR Primers recommended by the CDC in 2003 could have easily 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 were 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.

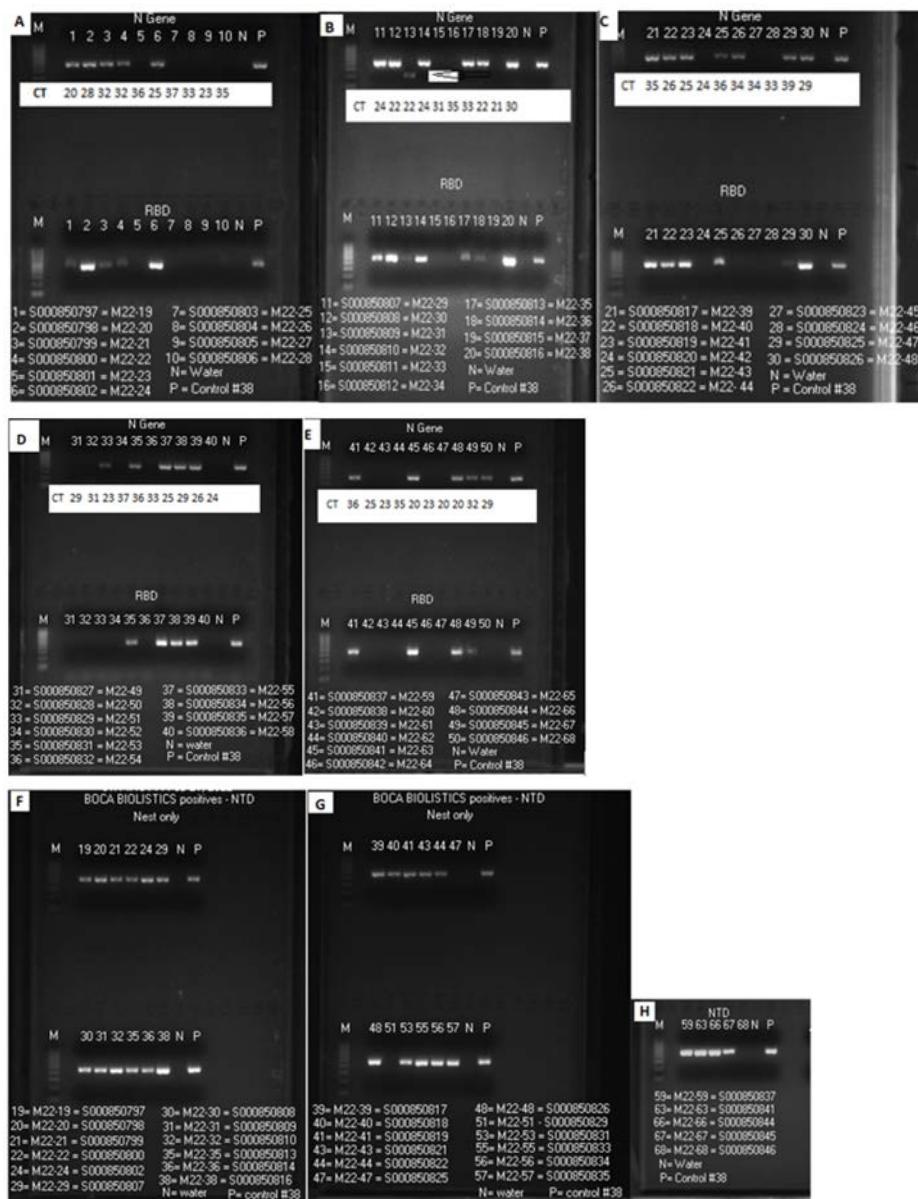


Figure 3. These are images of agarose gel electrophoresis of the SARS-CoV-2 N gene, RBD and NTD nested RT-PCR products. Panels A-E show a positive N gene band for 29 samples, M22-19, -20, -21, -22, -24, -29, -30, -31, -32, -35, -36, -38, -39, -40, -41, -43, -44, -47, -48, -51, -53, -55, -56, -57, -59, -63, -64, -65, -66, -67, -68, -69, -70, -71, -72, -73, -74, -75, -76, -77, -78, -79, -80, -81, -82, -83, -84, -85, -86, -87, -88, -89, -90, -91, -92, -93, -94, -95, -96, -97, -98, -99, -100, -101, -102, -103, -104, -105, -106, -107, -108, -109, -110, -111, -112, -113, -114, -115, -116, -117, -118, -119, -120, -121, -122, -123, -124, -125, -126, -127, -128, -129, -130, -131, -132, -133, -134, -135, -136, -137, -138, -139, -140, -141, -142, -143, -144, -145, -146, -147, -148, -149, -150, -151, -152, -153, -154, -155, -156, -157, -158, -159, -160, -161, -162, -163, -164, -165, -166, -167, -168, -169, -170, -171, -172, -173, -174, -175, -176, -177, -178, -179, -180, -181, -182, -183, -184, -185, -186, -187, -188, -189, -190, -191, -192, -193, -194, -195, -196, -197, -198, -199, -200, -201, -202, -203, -204, -205, -206, -207, -208, -209, -210, -211, -212, -213, -214, -215, -216, -217, -218, -219, -220, -221, -222, -223, -224, -225, -226, -227, -228, -229, -230, -231, -232, -233, -234, -235, -236, -237, -238, -239, -240, -241, -242, -243, -244, -245, -246, -247, -248, -249, -250, -251, -252, -253, -254, -255, -256, -257, -258, -259, -260, -261, -262, -263, -264, -265, -266, -267, -268, -269, -270, -271, -272, -273, -274, -275, -276, -277, -278, -279, -280, -281, -282, -283, -284, -285, -286, -287, -288, -289, -290, -291, -292, -293, -294, -295, -296, -297, -298, -299, -300, -301, -302, -303, -304, -305, -306, -307, -308, -309, -310, -311, -312, -313, -314, -315, -316, -317, -318, -319, -320, -321, -322, -323, -324, -325, -326, -327, -328, -329, -330, -331, -332, -333, -334, -335, -336, -337, -338, -339, -340, -341, -342, -343, -344, -345, -346, -347, -348, -349, -350, -351, -352, -353, -354, -355, -356, -357, -358, -359, -360, -361, -362, -363, -364, -365, -366, -367, -368, -369, -370, -371, -372, -373, -374, -375, -376, -377, -378, -379, -380, -381, -382, -383, -384, -385, -386, -387, -388, -389, -390, -391, -392, -393, -394, -395, -396, -397, -398, -399, -400, -401, -402, -403, -404, -405, -406, -407, -408, -409, -410, -411, -412, -413, -414, -415, -416, -417, -418, -419, -420, -421, -422, -423, -424, -425, -426, -427, -428, -429, -430, -431, -432, -433, -434, -435, -436, -437, -438, -439, -440, -441, -442, -443, -444, -445, -446, -447, -448, -449, -450, -451, -452, -453, -454, -455, -456, -457, -458, -459, -460, -461, -462, -463, -464, -465, -466, -467, -468, -469, -470, -471, -472, -473, -474, -475, -476, -477, -478, -479, -480, -481, -482, -483, -484, -485, -486, -487, -488, -489, -490, -491, -492, -493, -494, -495, -496, -497, -498, -499, -500, -501, -502, -503, -504, -505, -506, -507, -508, -509, -510, -511, -512, -513, -514, -515, -516, -517, -518, -519, -520, -521, -522, -523, -524, -525, -526, -527, -528, -529, -530, -531, -532, -533, -534, -535, -536, -537, -538, -539, -540, -541, -542, -543, -544, -545, -546, -547, -548, -549, -550, -551, -552, -553, -554, -555, -556, -557, -558, -559, -560, -561, -562, -563, -564, -565, -566, -567, -568, -569, -570, -571, -572, -573, -574, -575, -576, -577, -578, -579, -580, -581, -582, -583, -584, -585, -586, -587, -588, -589, -590, -591, -592, -593, -594, -595, -596, -597, -598, -599, -600, -601, -602, -603, -604, -605, -606, -607, -608, -609, -610, -611, -612, -613, -614, -615, -616, -617, -618, -619, -620, -621, -622, -623, -624, -625, -626, -627, -628, -629, -630, -631, -632, -633, -634, -635, -636, -637, -638, -639, -640, -641, -642, -643, -644, -645, -646, -647, -648, -649, -650, -651, -652, -653, -654, -655, -656, -657, -658, -659, -660, -661, -662, -663, -664, -665, -666, -667, -668, -669, -670, -671, -672, -673, -674, -675, -676, -677, -678, -679, -680, -681, -682, -683, -684, -685, -686, -687, -688, -689, -690, -691, -692, -693, -694, -695, -696, -697, -698, -699, -700, -701, -702, -703, -704, -705, -706, -707, -708, -709, -710, -711, -712, -713, -714, -715, -716, -717, -718, -719, -720, -721, -722, -723, -724, -725, -726, -727, -728, -729, -730, -731, -732, -733, -734, -735, -736, -737, -738, -739, -740, -741, -742, -743, -744, -745, -746, -747, -748, -749, -750, -751, -752, -753, -754, -755, -756, -757, -758, -759, -760, -761, -762, -763, -764, -765, -766, -767, -768, -769, -770, -771, -772, -773, -774, -775, -776, -777, -778, -779, -770, -771, -772, -773, -774, -775, -776, -777, -778, -779, -780, -781, -782, -783, -784, -785, -786, -787, -788, -789, -790, -791, -792, -793, -794, -795, -796, -797, -798, -799, -800, -801, -802, -803, -804, -805, -806, -807, -808, -809, -8010, -8011, -8012, -8013, -8014, -8015, -8016, -8017, -8018, -8019, -8020, -8021, -8022, -8023, -8024, -8025, -8026, -8027, -8028, -8029, -8030, -8031, -8032, -8033, -8034, -8035, -8036, -8037, -8038, -8039, -8040, -8041, -8042, -8043, -8044, -8045, -8046, -8047, -8048, -8049, -8050, -8051, -8052, -8053, -8054, -8055, -8056, -8057, -8058, -8059, -8060, -8061, -8062, -8063, -8064, -8065, -8066, -8067, -8068, -8069, -8070, -8071, -8072, -8073, -8074, -8075, -8076, -8077, -8078, -8079, -8080, -8081, -8082, -8083, -8084, -8085, -8086, -8087, -8088, -8089, -8090, -8091, -8092, -8093, -8094, -8095, -8096, -8097, -8098, -8099, -80100, -80101, -80102, -80103, -80104, -80105, -80106, -80107, -80108, -80109, -80110, -80111, -80112, -80113, -80114, -80115, -80116, -80117, -80118, -80119, -80120, -80121, -80122, -80123, -80124, -80125, -80126, -80127, -80128, -80129, -80130, -80131, -80132, -80133, -80134, -80135, -80136, -80137, -80138, -80139, -80140, -80141, -80142, -80143, -80144, -80145, -80146, -80147, -80148, -80149, -80150, -80151, -80152, -80153, -80154, -80155, -80156, -80157, -80158, -80159, -80160, -80161, -80162, -80163, -80164, -80165, -80166, -80167, -80168, -80169, -80170, -80171, -80172, -80173, -80174, -80175, -80176, -80177, -80178, -80179, -80180, -80181, -80182, -80183, -80184, -80185, -80186, -80187, -80188, -80189, -80190, -80191, -80192, -80193, -80194, -80195, -80196, -80197, -80198, -80199, -80200, -80201, -80202, -80203, -80204, -80205, -80206, -80207, -80208, -80209, -80210, -80211, -80212, -80213, -80214, -80215, -80216, -80217, -80218, -80219, -80220, -80221, -80222, -80223, -80224, -80225, -80226, -80227, -80228, -80229, -80230, -80231, -80232, -80233, -80234, -80235, -80236, -80237, -80238, -80239, -80240, -80241, -80242, -80243, -80244, -80245, -80246, -80247, -80248, -80249, -80250, -80251, -80252, -80253, -80254, -80255, -80256, -80257, -80258, -80259, -80260, -80261, -80262, -80263, -80264, -80265, -80266, -80267, -80268, -80269, -80270, -80271, -80272, -80273, -80274, -80275, -80276, -80277, -80278, -80279, -80280, -80281, -80282, -80283, -80284, -80285, -80286, -80287, -80288, -80289, -80290, -80291, -80292, -80293, -80294, -80295, -80296, -80297, -80298, -80299, -80300, -80301, -80302, -80303, -80304, -80305, -80306, -80307, -80308, -80309, -80310, -80311, -80312, -80313, -80314, -80315, -80316, -80317, -80318, -80319, -80320, -80321, -80322, -80323, -80324, -80325, -80326, -80327, -80328, -80329, -80330, -80331, -80332, -80333, -80334, -80335, -80336, -80337, -80338, -80339, -80340, -80341, -80342, -80343, -80344, -80345, -80346, -80347, -80348, -80349, -80350, -80351, -80352, -80353, -80354, -80355, -80356, -80357, -80358, -80359, -80360, -80361, -80362, -80363, -80364, -80365, -80366, -80367, -80368, -80369, -80370, -80371, -80372, -80373, -80374, -80375, -80376, -80377, -80378, -80379, -80380, -80381, -80382, -80383, -80384, -80385, -80386, -80387, -80388, -80389, -80390, -80391, -80392, -80393, -80394, -80395, -80396, -80397, -80398, -80399, -80400, -80401, -80402, -80403, -80404, -80405, -80406, -80407, -80408, -80409, -80410, -80411, -80412, -80413, -80414, -80415, -80416, -80417, -80418, -80419, -80420, -80421, -80422, -80423, -80424, -80425, -80426, -80427, -80428, -80429, -80430, -80431, -80432, -80433, -80434, -80435, -80436, -80437, -80438, -80439, -80440, -80441, -80442, -80443, -80444, -80445, -80446, -80447, -80448, -80449, -80450, -80451, -80452, -80453, -80454, -80455, -80456, -80457, -80458, -80459, -80460, -80461, -80462, -80463, -80464, -80465, -80466, -80467, -80468, -80469, -80470, -80471, -80472, -80473, -80474, -80475, -80476, -80477, -80478, -80479, -80480, -80481, -80482, -80483, -80484, -80485, -80486, -80487, -80488, -80489, -80490, -80491, -80492, -80493, -80494, -80495, -80496, -80497, -80498, -80499, -80500, -80501, -80502, -80503, -80504, -80505, -80506, -80507, -80508, -80509, -80510, -80511, -80512, -80513, -80514, -80515, -80516, -80517, -80518, -80519, -80520, -80521, -80522, -80523, -80524, -80525, -80526, -80527, -80528, -80529, -80530, -80531, -80532, -80533, -80534, -80535, -80536, -80537, -80538, -80539, -80540, -80541, -80542, -80543, -80544, -80545, -80546, -80547, -80548, -80549, -80550, -80551, -80552, -80553, -80554, -80555, -80556, -80557, -80558, -80559, -80560, -80561, -80562, -80563, -80564, -80565, -80566, -80567, -80568, -80569, -80570, -80571, -80572, -80573, -80574, -80575, -80576, -80577, -80578, -80579, -80580, -80581, -80582, -80583, -80584, -80585, -80586, -80587, -80588, -80589, -80590, -80591, -80592, -80593, -80594, -80595, -80596, -80597, -80598, -80599, -80600, -80601, -80602, -80603, -80604, -80605, -80606, -80607, -80608, -80609, -80610, -80611, -80612, -80613, -80614, -80615, -80616, -80617, -80618, -80619, -80620, -80621, -80622, -80623, -80624, -80625, -80626, -80627, -80628, -80629, -80630, -80631, -80632, -80633, -80634, -80635, -80636, -80637, -80638, -80639, -80640, -80641, -80642, -80643, -80644, -80645, -80646, -80647, -80648, -80649, -80650, -80651, -80652, -80653, -80654, -80655, -80656, -80657, -80658, -80659, -80660, -80661, -80662, -80663, -80664, -80665, -80666, -80667, -80668, -80669, -80670, -80671, -80672, -80673, -80674, -80675, -80676, -80677, -80678, -80679, -80680, -80681, -80682, -80683, -80684, -80685, -80686, -80687, -80688, -80689, -80690, -80691, -80692, -80693, -80694, -80695, -80696, -80697, -80698, -80699, -80700, -80701, -80702, -80703, -80704, -80705, -80706, -80707, -80708, -80709, -80710, -80711, -80712, -80713, -80714, -80715, -80716, -80717, -80718, -80719, -80720, -80721, -80722, -80723, -80724, -80725, -80726, -80727, -80728, -80729, -80730, -80731, -80732, -80733, -80734, -80735, -80736, -80

-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.

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 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 bidirectional Sanger sequencing, using the respective nested PCR primers as the sequencing primers. The results are summarized in Table 3.

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	+	+	+	+	+	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;

RS(Co3) = Co3 reverse sequencing primer.

FS(S9) = S9 forward sequencing primer;

RS(S10) = S10 reverse sequencing primer;

FS(SB7) = SB7 forward 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.

3.3. 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 were presented in Figure 4.

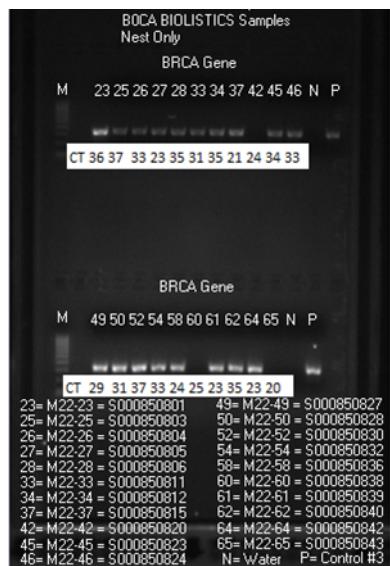


Figure 4. This image of agarose gel electrophoresis of the nested PCR amplification products shows that 18 of the 21 samples, which were negative for SARS-CoV-2 N gene and RBD RT-PCR amplification, contained a segment of human *BRCA* gene, an indication of sample adequacy. However, 3 samples, M22-42, 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 [27]. 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.

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, Δ69-70, T95I, G142D and Δ143-145 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. A set of bidirectional sequencing electropherograms illustrating these mutations is presented in Figures 5-10 as follows.

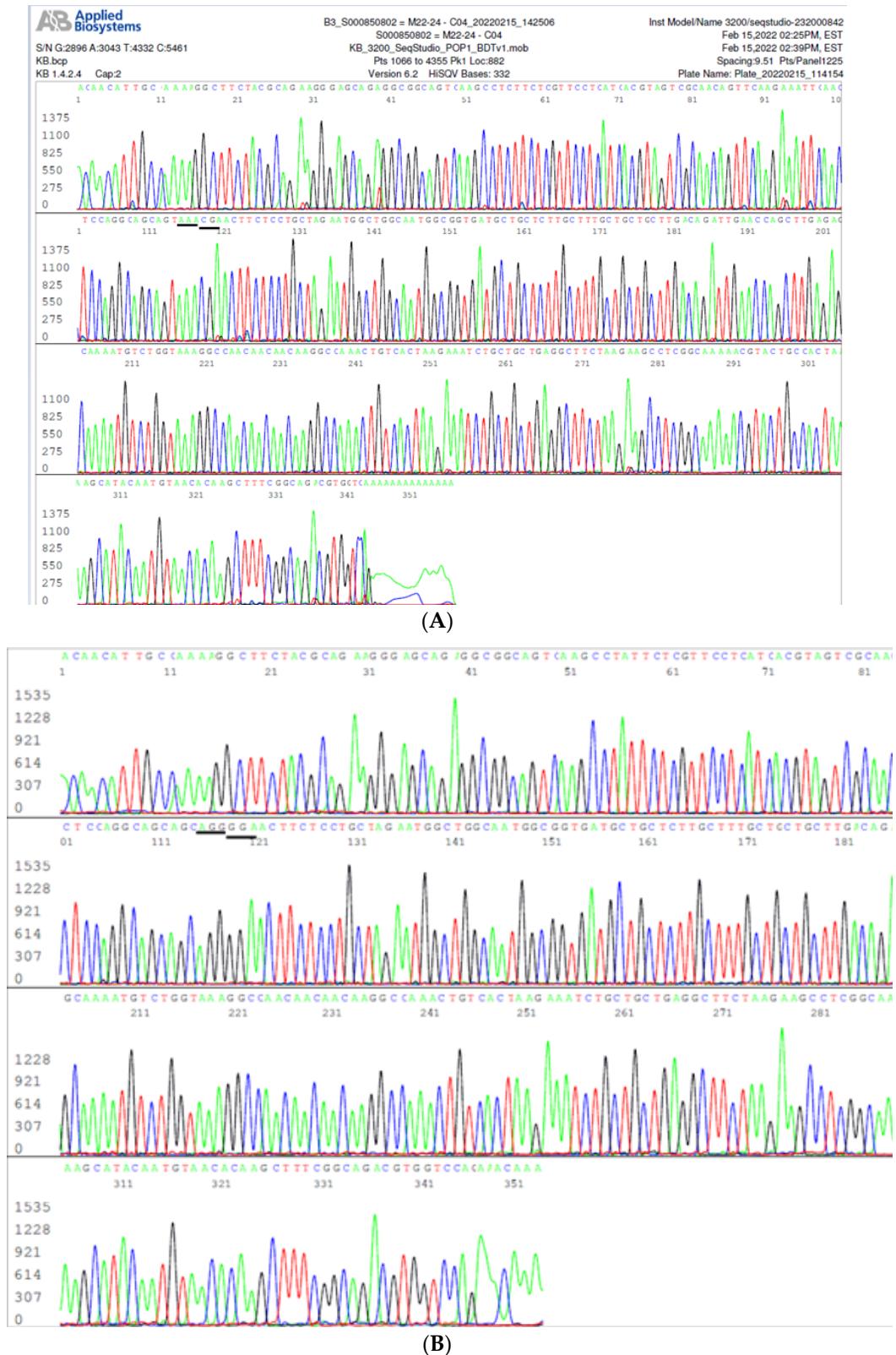


Figure 5. These two electropherograms showing the N gene R203K and G204R mutations in sample M22-24, using primer Co4 as the forward sequencing primer (A) and the wildtype SARS-CoV-2 Wuhan-Hu-1 control sequence for comparison (B). Involved codons are underlined.

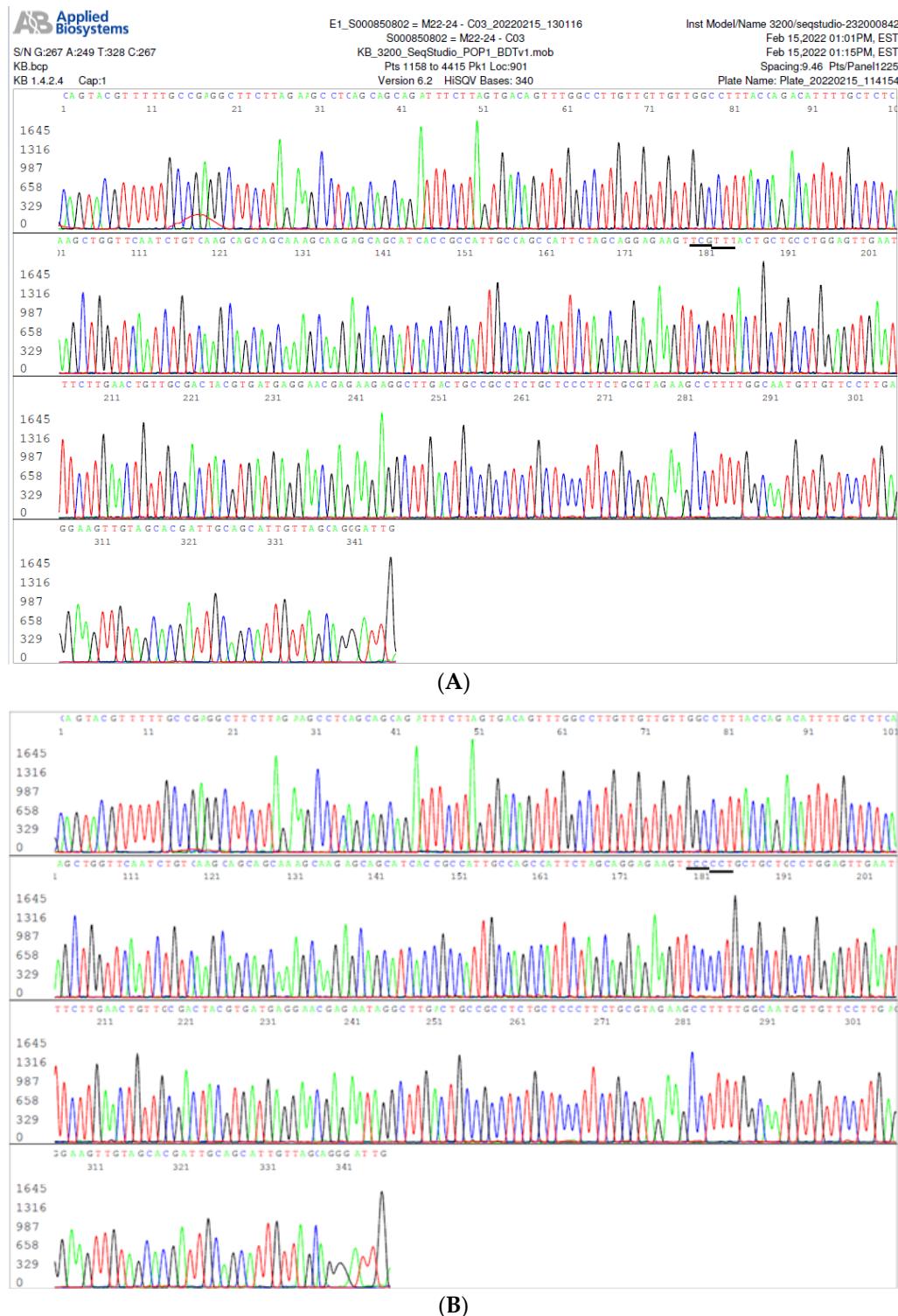


Figure 6. These two electropherograms showing the N gene G204R and R203K mutations in sample M22-24, using primer Co3 as the reverse sequencing primer (A) and the wildtype SARS-CoV-2 Wuhan-Hu-1 control sequence for comparison (B). Involved codons are underlined.

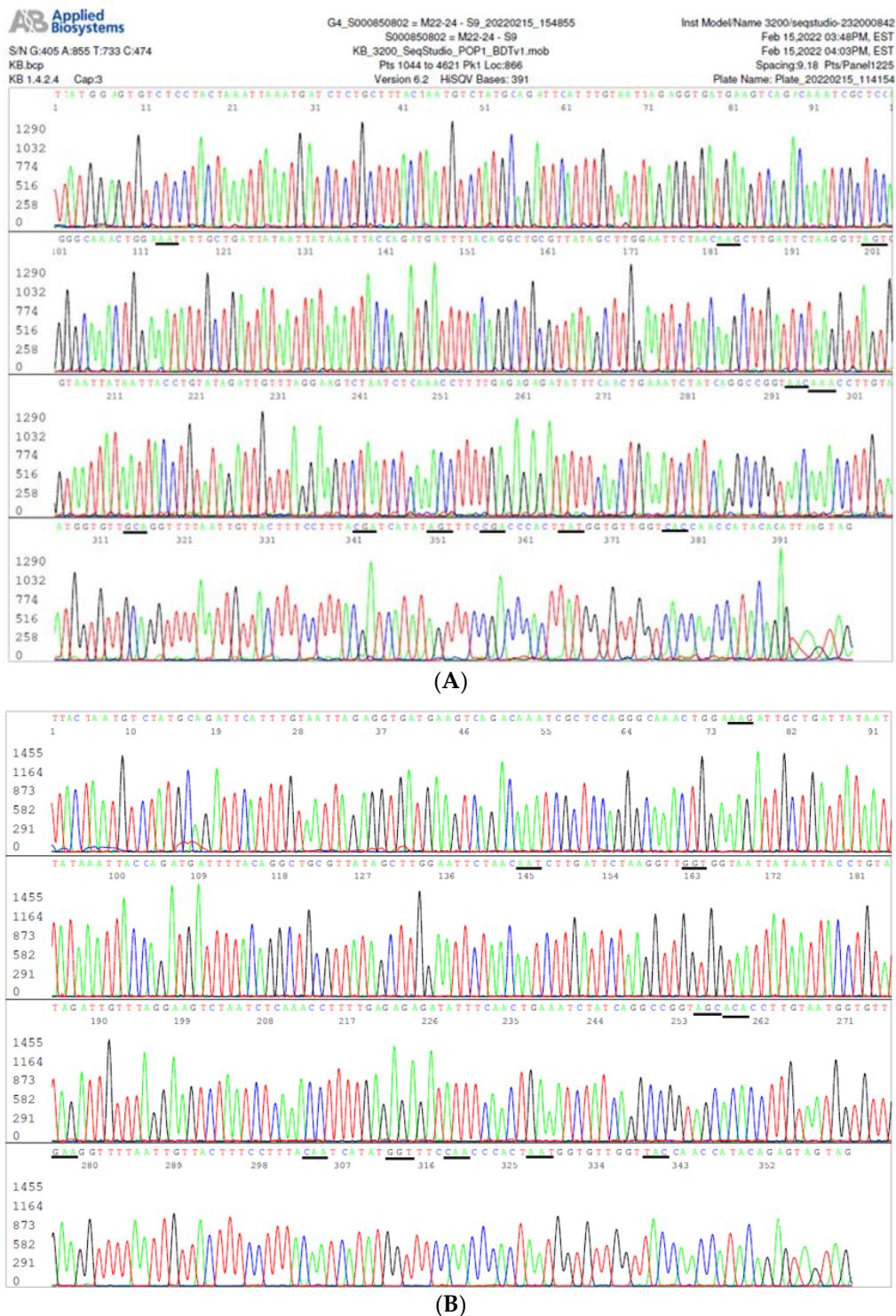


Figure 7. These two electropherograms showing the S gene RBD K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y and Y505H mutations in sample M22-24, using primer S9 as the forward sequencing primer (A) and the wildtype SARS-CoV-2 Wuhan-Hu-1 control sequence for comparison (B). Involved codons are underlined.



Figure 8. These two electropherograms showing 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 (A) and the wildtype SARS-CoV-2 Wuhan-Hu-1 control sequence for comparison (B). Involved codons are underlined.

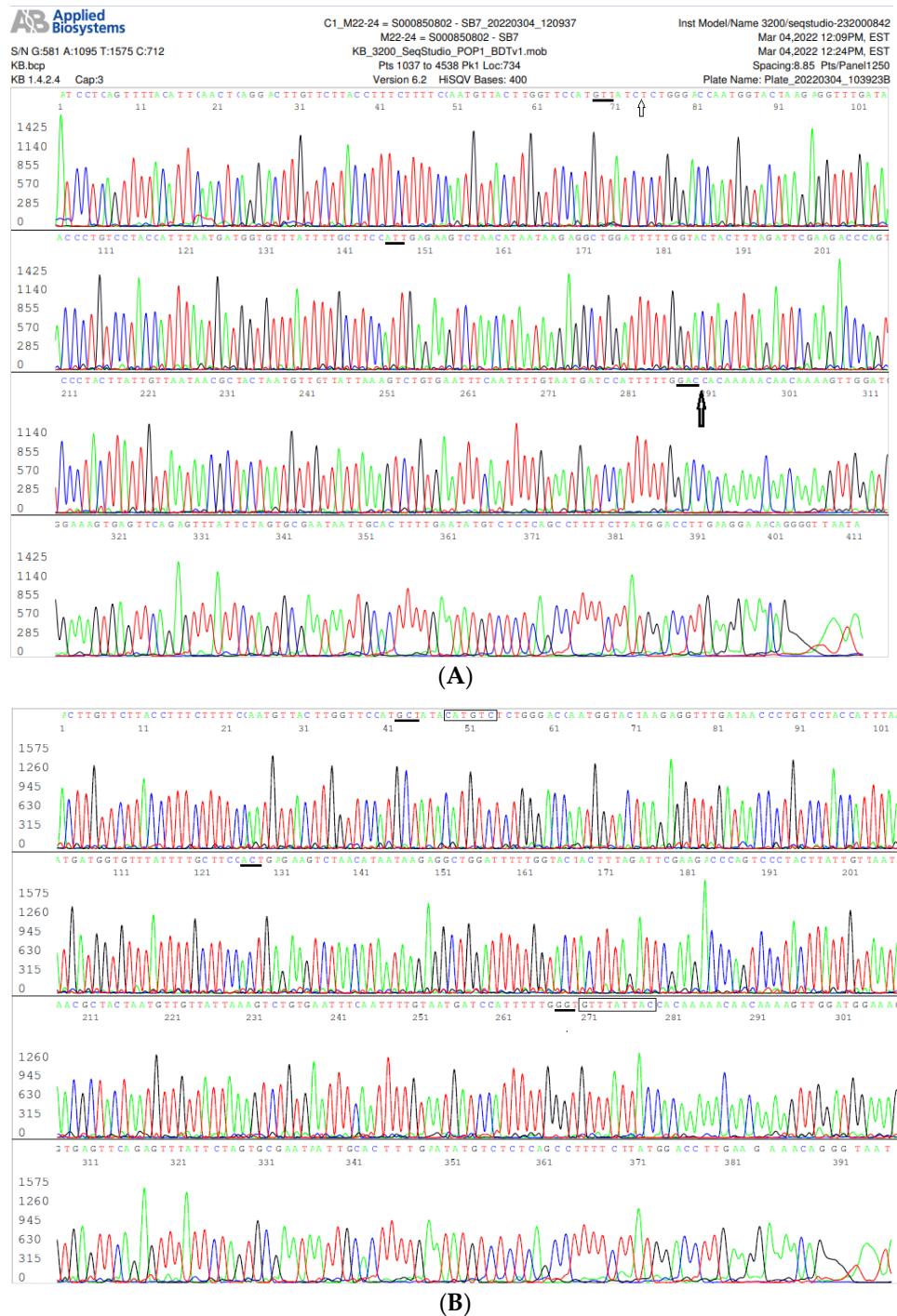


Figure 9. These two electropherograms showing the S gene NTD A67V, Δ69-70, T95I, G142D and Δ143-145 mutations in sample M22-24, using primer SB7 as the forward sequencing primer (A) and the wildtype SARS-CoV-2 Wuhan-Hu-1 control sequence for comparison (B). Involved codons are underlined. The positions of Δ69-70 and Δ143-145 are indicated by a small arrow and a big 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).

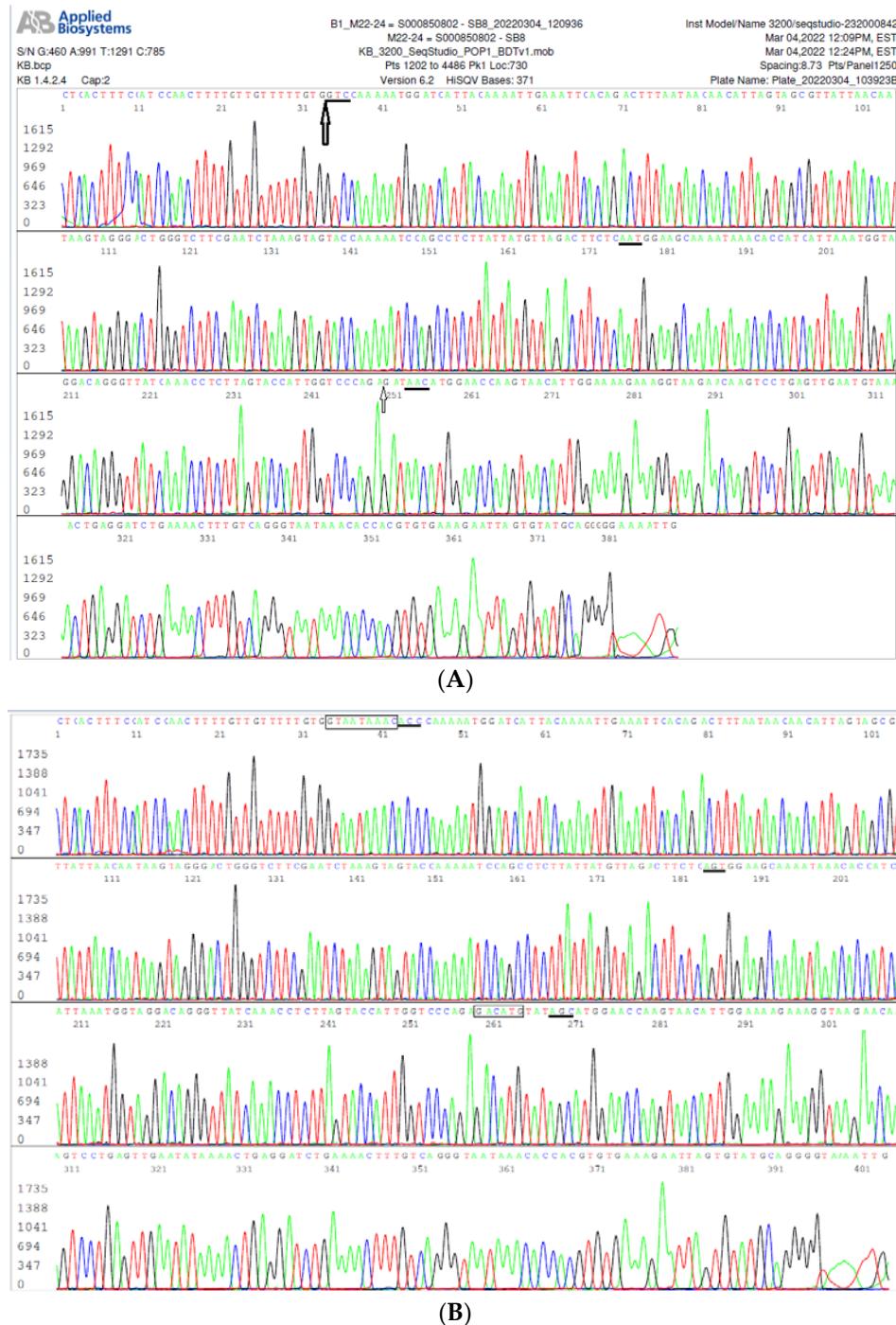


Figure 10. These two electropherograms showing the S gene NTD $\Delta 143$ - 145 , $G142D$, $T95I$, $\Delta 69$ - 70 and $A67V$ mutations in sample M22-24, using primer SB8 as the reverse sequencing primer (A) and the wildtype SARS-CoV-2 Wuhan-Hu-1 control sequence for comparison (B). Involved codons are underlined. The positions of $\Delta 143$ - 145 and $\Delta 69$ - 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 [26-28]. 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 9 A, Figure 10A, Figure 11 and Figure 12 for comparison.

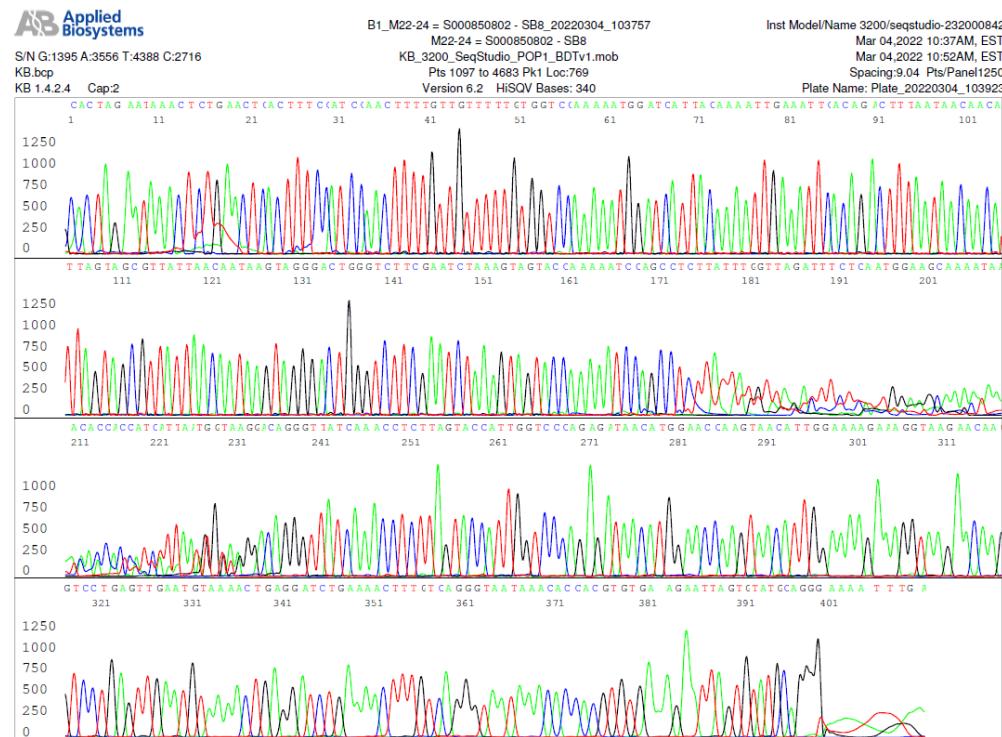


Figure 11. This electropherogram showing loss of sequencing signal in the NTD reverse primer sequencing from base position 180 to base position 230 although the template came from the same nested RT-PCR products, which were used as the template to generate Figure 9 A and Figure 10A.

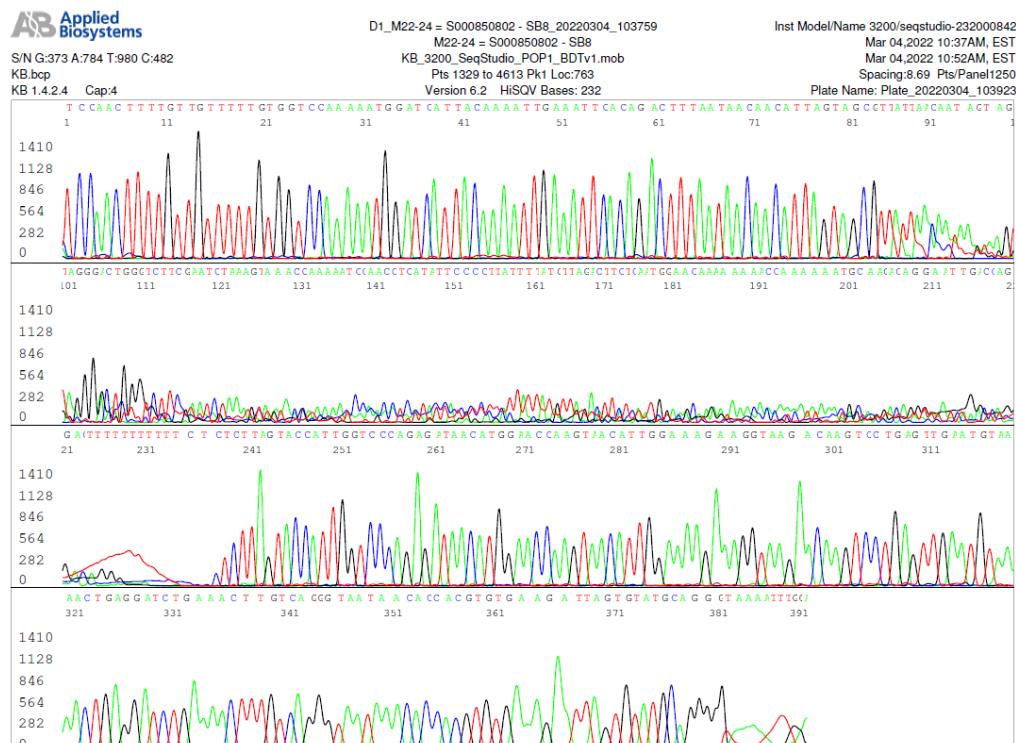


Figure 12. This electropherogram showing loss of sequencing signal in NTD reverse primer sequencing from base position 90 to base position 238 although the template came from the same nested RT-PCR products, which were used as the template to generate Figure 9 A and Figure 10A.

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 [27, 32].

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, Δ69-70, T95I, G142D and Δ143-145 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.

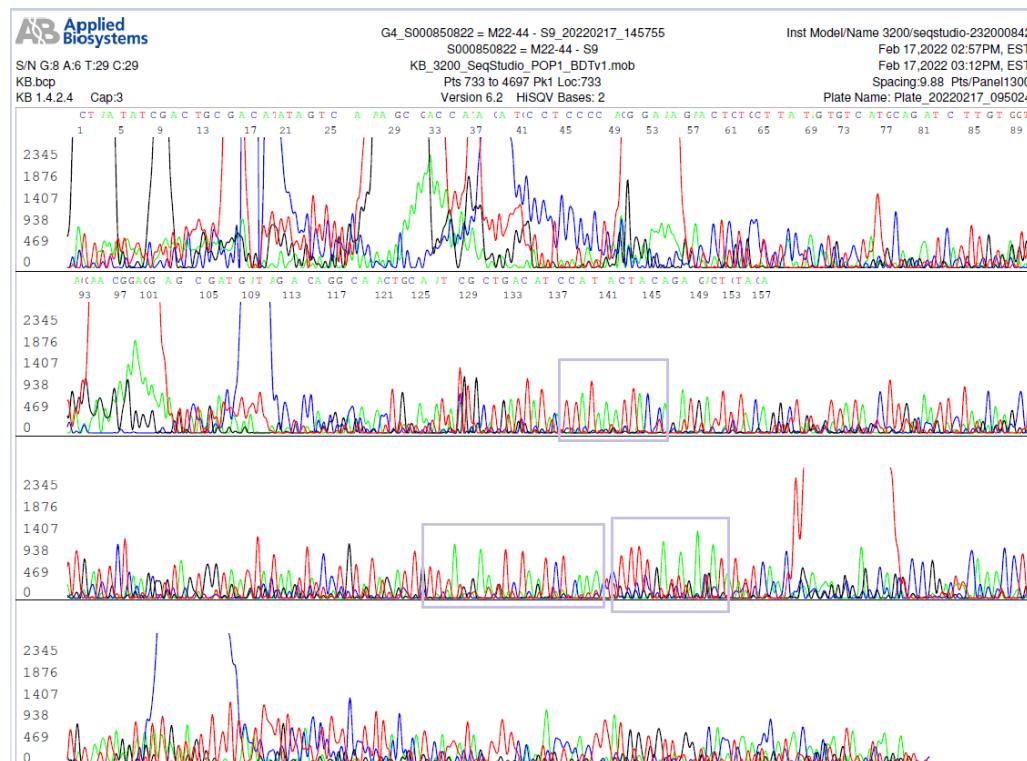


Figure 13. This is an electropherogram of forward primer sequencing of the RBD nested PCR products of sample M22-44 although a band of the PCR products was not visible to the naked eye (Figure 3, panel C, lane 26). Accurate base calling on this electropherogram was not possible due to multiple overlapping sequences. But the electropherogram showed one stretch of sequence "TTATAAATTACCA" in a single rectangle and another stretch of sequence "TCTAAATCTCAAAC-CTTTGAGAGAGAT" 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.

```
22441 ttcaactgttag aaaaaggaat ctatcaaact tctaaacttta gagtccaaacc aacagaatct
22501 attgttagat ttccataat tacaacttg tgnnntnnnn nnnnnnnnnnn nnnnnnnnnnn
22561 nnnnnnnnnnn nnnnnnnnnnn nnngaacagg aagagaatca gcaactnnnn nnnnnnnnnat
22621 tctgtccatat ataatccgc atcatttcc acttttaagt gttnnnnnnnn nnnnnnnnnnt
22681 aaattaaatg atctctgctt tnctaattgtc tatgcagatt catttgtnnn nnnnnnnnnat
22741 naagtcaagac aaatcgctcc agggcaaact gaaaaagatnn ctgattataa ttataaaattt
22801 ccnnntgatt ttacaggctn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnctaaggtt
22861 ggtggtaann nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn nnnnnnnnnn
22921 nnnnnnnnnnn nnnnnnnnaan nnnnnnnnnnn nnnnnnnnnnn nnnnnnnntgg tgttgaaggt
22981 tnnnnattgtt actttccctt acaatcatat ggnnnnnnnnn nnnnnnnnnnn nnnnnnnnnnn
23041 nnnnnnnnnn nnnnngtagt actttccctt gannnnnnnn nnnnnnnnnnn nnnnnnnnnn
23101 nnnnnnnnnnn nnnnnnctaa tttggtaaa aacaaatgtg tcaattcaa 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).

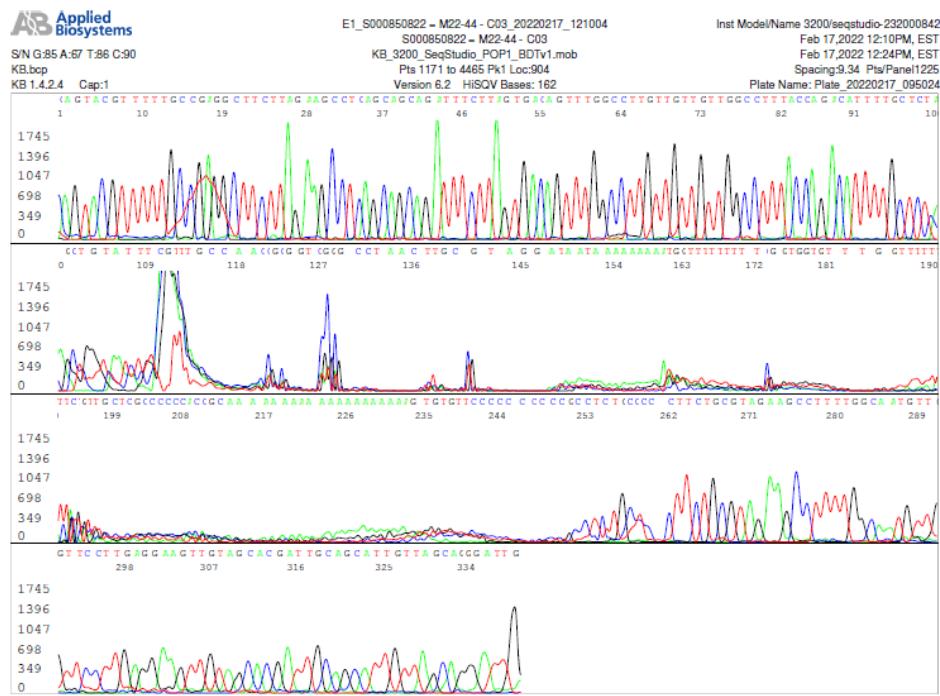


Figure 15. This is the only N gene sequencing electropherogram among a total of 58 (Table 3) showing loss of signal in a segment of DNA sequence. It was generated using a reverse sequencing primer. Since the beginning and the ending parts of this sequence are accurately deciphered, the intervening segments of the templates must harbor multi-allelic SNPs without insertions or deletions.

A forward primer sequencing of the same N gene nested PCR products generated a fully expected sequence with R203K and G204R mutations (Figure 16).

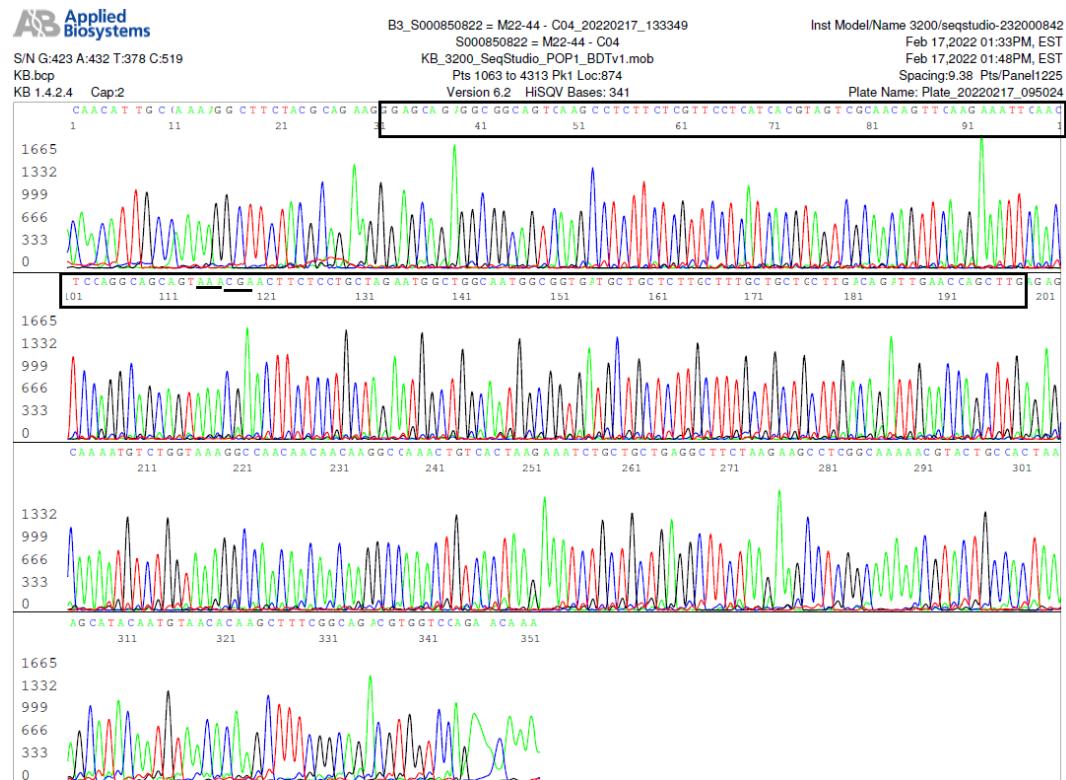


Figure 16. This is an electropherogram showing an expected DNA sequence for an Omicron isolate when the same N gene nested PCR products, which were used to generate the sequence presented

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.

Loss of signal in diagnostic N gene sequencing is unusual [26]. 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.

```

28561 atgaaagatc tcagtccaag atggtatttc tactacctag gaactggcc agaagctgga
28621 ctccctatg gtgctaacaa agacggcatc atatgggtt caactgaggg agccttgaat
28681 acacaaaag atcacattgg caccgcata cctgctaac
[gap 117 bp] Expand Ns
28837 agtt caagaaattc aactccaggc
28861 agcagtatgg gaacttctcc tgctagtat gctgcttgc ctggctgct gcttgacaga
28921 ttgaaccacgtt tagagagcaa aatgtcttgtt aaaggccaaac aacaacaagg ccaaactgtc
28981 actaagaat ctgctgtga ggcttcaag aagcctcgcc aaaaacgtac tgccactaaa

```

Figure 17. This 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.

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 [23-25].

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 were presented in Figures 18 and 19.

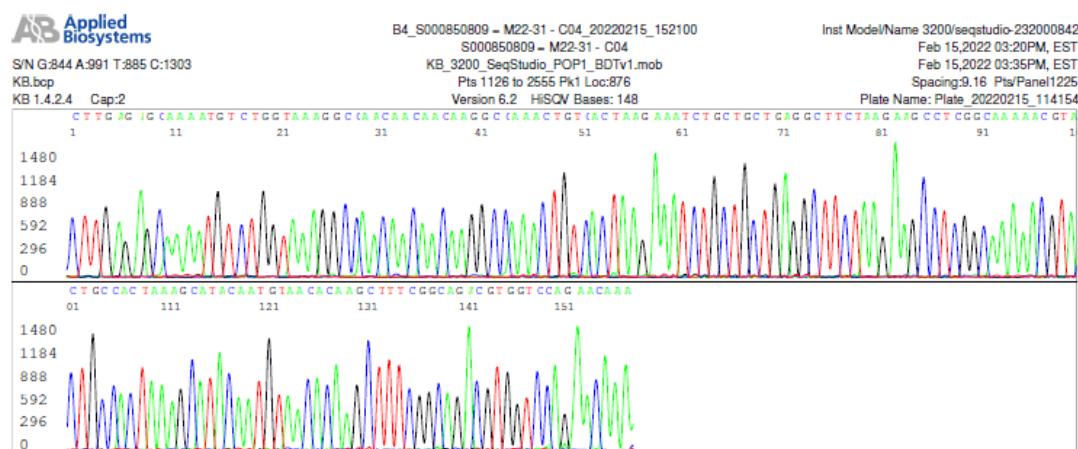


Figure 18. This is an electropherogram of the forward sequencing of the N gene nested PCR products of sample M22-31. The R203 and G204 codons were not included in the PCR amplicon (see Figure 5).

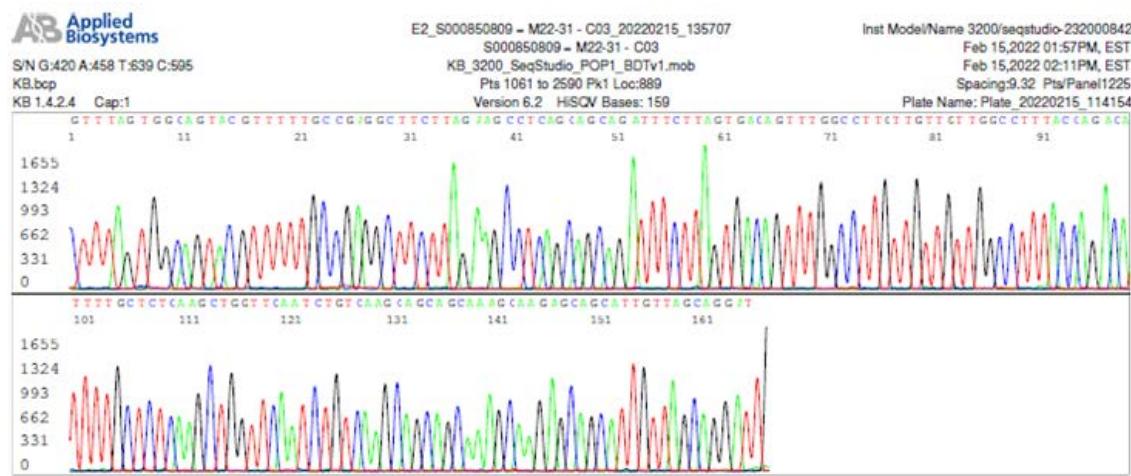


Figure 19. This is an electropherogram of the reverse sequencing of the N gene nested PCR products of sample M22-31. The R203 and G204 codons were not included in the PCR amplicon (see Figure 6).

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

CAATCCTGCTAACAACTGCTGCTTGTGCTGCTGAGACAGATTGAACCAGCTTGAGAGCAAAATGCTGGTAAAGGCCAACA
ACAACAAGGCCAAACTGTCACTAAGAAATCTGCTGCTGAGGCTTCTAAGAAGCCTCGCAAAACGTACTGCCACTAAAGCATACAA
TGTAACACAAAGCTTCCGGCAGACGTGGTCCAGAACAAA

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

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

Range 1: 28554 to 28753 GenBank Graphics					▼ Next Match	▲ Previous Match
Score 370 bits(200)	Expect 4e-98	Identities 200/200(100%)	Gaps 0/200(0%)	Strand Plus/Plus		
Query 13	CAATGCTGCTCTGCTTGTGCTGCTTGTGACAGATTGAAACAGCTTGGAGAGCAAATGTC				72	
Sbjct 28554	CAATGCTGCTCTGCTTGTGCTGCTTGTGACAGATTGAAACAGCTTGGAGAGCAAATGTC					28613
Query 73	TGGTAAAGGCCAACAACAAAGGCCAACACTGTCACTAAAGAAATCTGCTGCTGAGGCCCTTC				132	
Sbjct 28614	TGGTAAAGGCCAACAACAAAGGCCAACACTGTCACTAAAGAAATCTGCTGCTGAGGCCCTTC					28673
Query 133	TAAGAACGCTCGGCAAAACAGTCACTGCCACTAAAGCATACATGTAACACAAGCTTCCG				192	
Sbjct 28674	TAAGAACGCTCGGCAAAACAGTCACTGCCACTAAAGCATACATGTAACACAAGCTTCCG					28733
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.

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).

NC_045512.2 N gene sequence

```

28681 gggagcccttg aatacaccaa aagatcacat tggcaccggc aatcctgcta acaatgctgc
28741 aatcgtgcta caacttcctc aaggaaacaac atggccaaaa ggcttctacg cagaaggggag
28801 cagaggccgc agtcaagcct ctttcgttc ctcatcacgt agtcgcaaca gttcaagaaa
28861 ttcaactcca ggcagcagta ggggacttc tcctgctaga atggctggca atggcggtga
28921 tgtgtctctt gctttgctgc tgcttgacag attgaaccag cttgagagca aatgtctgg
28981 taaaaggccaa caacacaacg gccaactgt cactaagaaa tctgtgtctg aggcttctaa
29041 gaaggcctcg caaaaacgta ctgcactaa agcataacaat gtaacacaag ctttcgcag
29101 acgtggtcca gaacaaacc aaggaaattt tggggaccag gaactaatca gacaaggAAC

```

OL891989.1 N gene sequence with GGD deletion

```

28321 actgaggag ccttgaatac accaaaagat ccatggca cccgaaatcc tgttaacaat
28381 gctgcaatcg tgctacaact tcctcaagga acaacattgc caaaaggctt ctacgcagaa
28441 gggagcagag gccggcagtca agcctttct cgttccatcacgtatcg caacagttc
28501 agaaaattcaa ctccaggcag cagtatggga acttctctg ctagaatggc tggaaatgt
28561 gttcttgctt tgctgtgt tgacagattt aaccagctt agagcaaat gtctggtaaa
28621 ggccaaacac aacaaggccca aactgtcaat aagaaatctg ctgtctggc ttctaagaag
28681 cctcggccaaa aacgtactgc cactaaagca tacaatgtaa cacaagctt cgccagaegt
28741 ggtccagaa aaacccaagg aaattttggg gaccaggAAC taatcagaca aggaactgtat

```

Figure 21. This figure 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, caatgtgc (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 [33].

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.

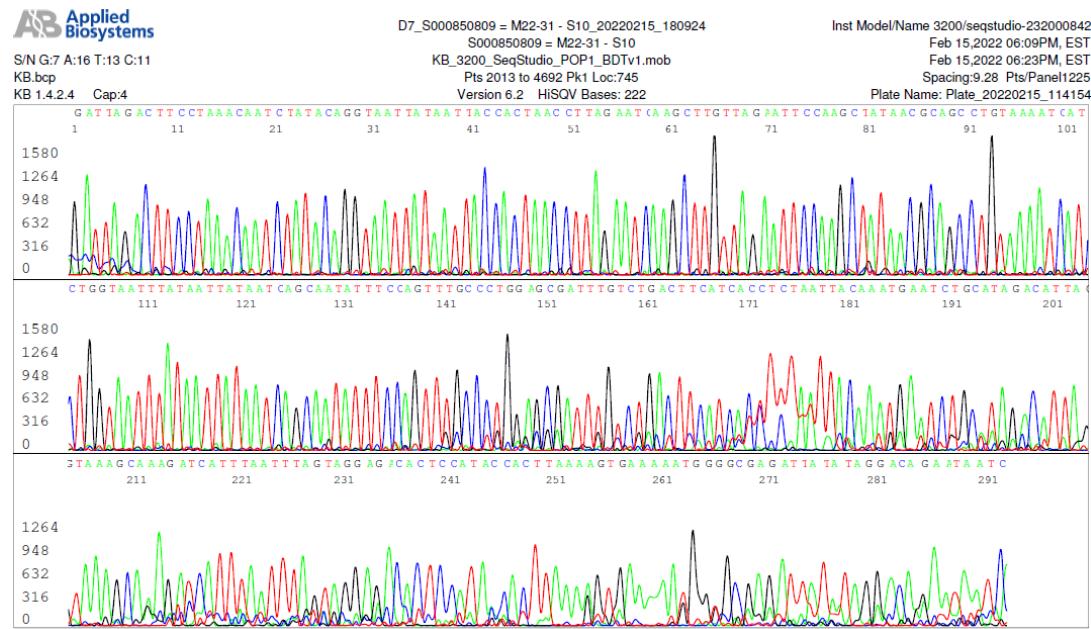


Figure 22. This reverse primer sequencing electropherogram was generated by at least two homeologous gene templates, which shared a 141-base common sequence before the heterogeneous base-calling peaks overlapped. The homologous 141-base sequence reads:

3'-GATTAGACTCTAAACAATCTATACAGGTAAATTATAATTACCACTAACCTAGAA TCAAGCTGTTAGAATTCCAAGCTATAACGGCAGCCTGAAAATCATCTGGTAATTAT AATTATAATCAGCAATATTCCAGTTT-5'.

After the sequence was converted to the 5'-3' format, it reads:

5'-AAACTGGAAATATTGCTGATTATAATTATAATTACCACTAACCTAGATGAT-
TTTACAGGCTGCGTTATAAGCTTGGAAATTCTAACAGCTTGATTCTAAGGTTAG-
TGGTAATTATAATTACCTGTATAGATTGTTAGGAAGTCTAAC; the underlined 138-base sequence encodes amino acids 415-460 of the SARS-CoV-2 S protein TGNI-ADYNYKLPDDFTGCVIAWNSNKLDSKVSGNYNYLYRLFRKSN 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, Δ69-70, T95I, G142D and Δ143-145. One of the sequencing panels showing A67V and Δ69-70 is presented in Figure 23.

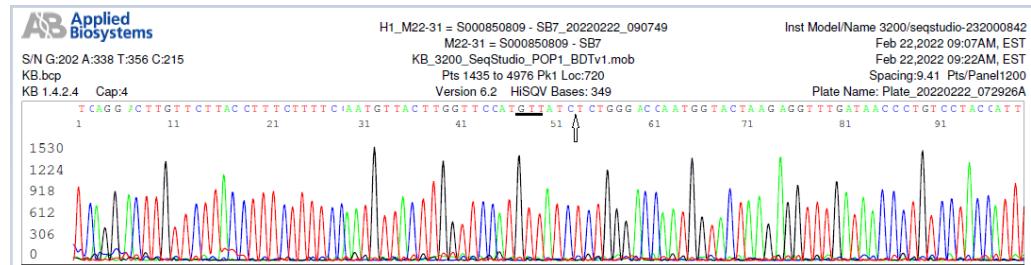


Figure 23. This is an electropherogram showing A67V and Δ69-70, part of the NTD mutations characteristic of an Omicron variant of SARS-CoV-2 in sample M22-31.

Therefore, M22-31 was interpreted as an unusual Omicron BA.1 variant with a GGD deletion in its N gene based on information retrieved from the GenBank.

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.

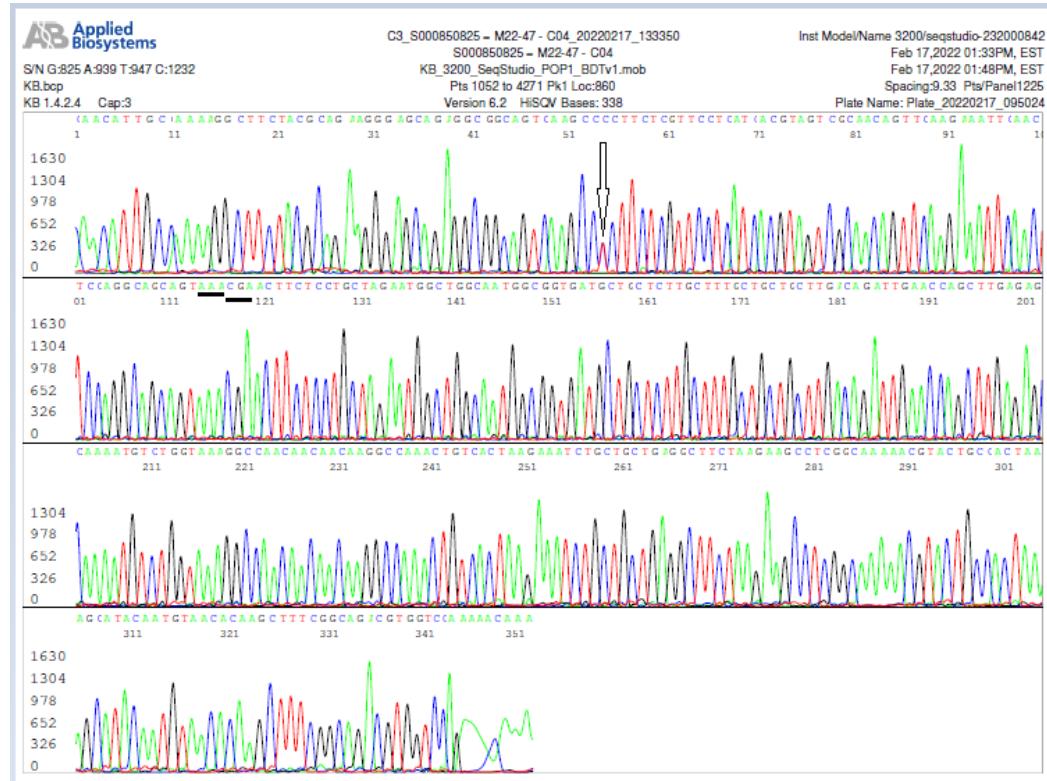


Figure 24. This is a forward N gene sequencing electropherogram on sample M-22 47 generated by two competing templates. One of the 2 templates has a T to C mutation at reference position 28820, indicated by an arrow (the computer read the combined T/C peaks as a “C”). A nucleotide T>C mutation in this 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.

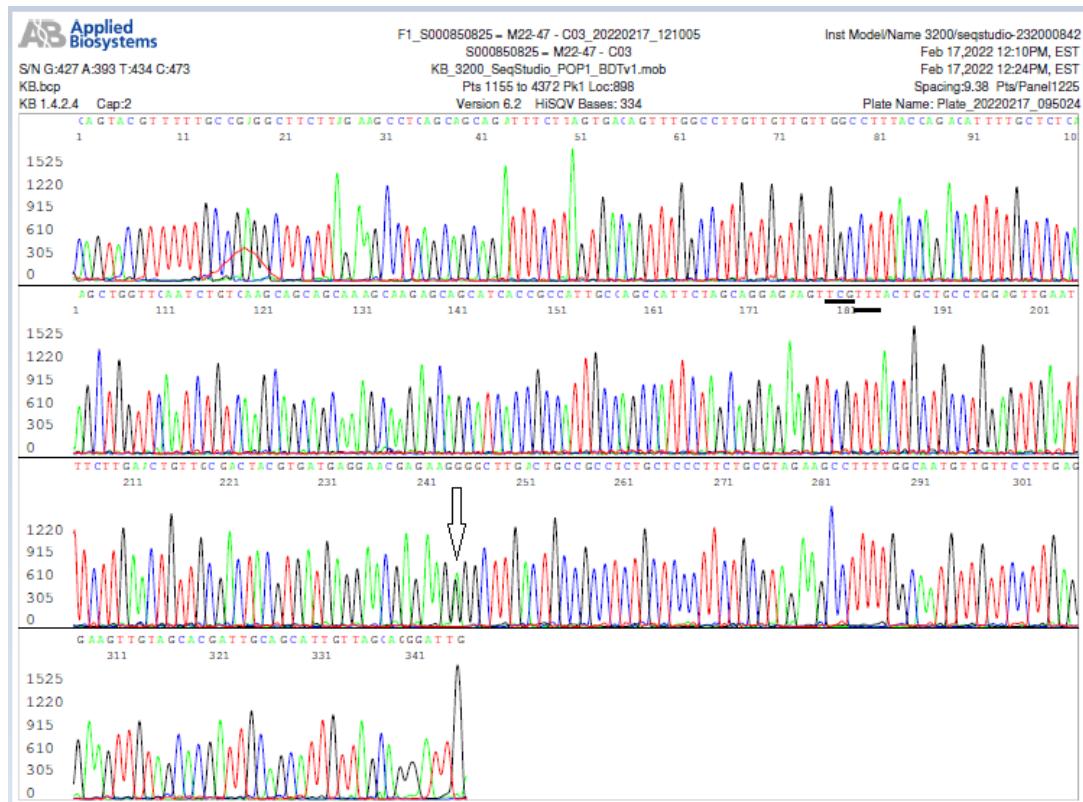


Figure 25. This is an electropherogram of the reverse N gene sequencing of the same nested PCR product that was used to generate the electropherogram presented in Figure 24. The mutated nucleotide G peak in the competing template is superimposed on the "A" peak of the parental sequence, pointed by an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

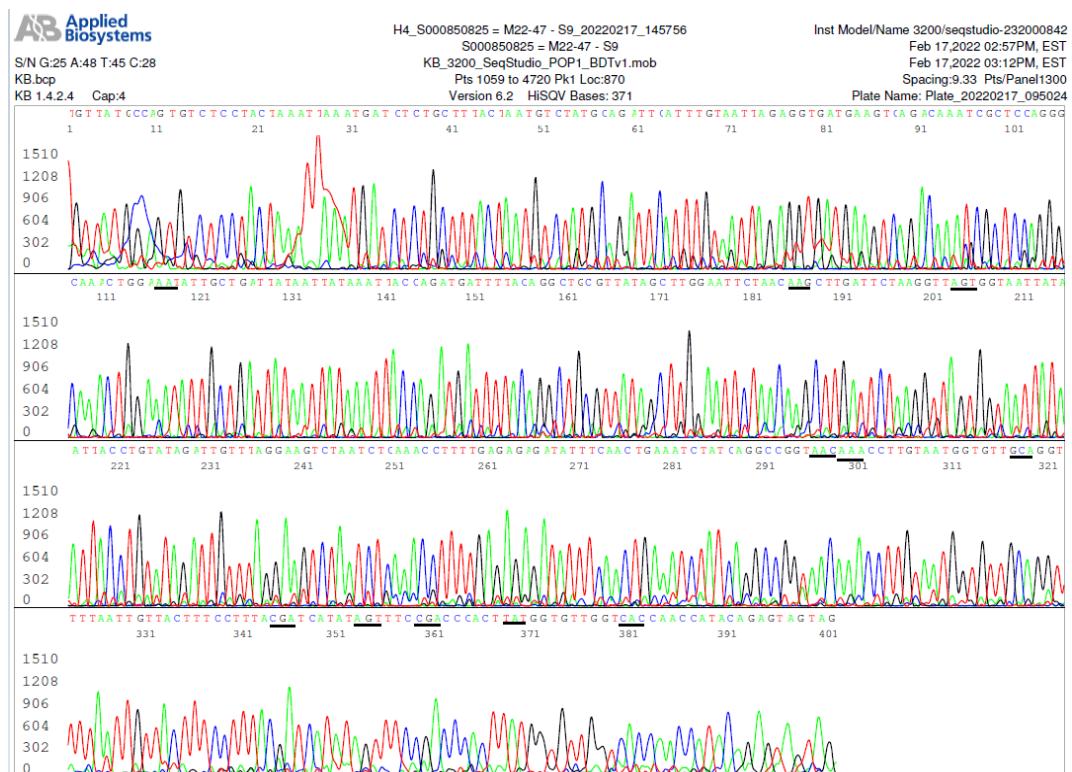


Figure 26. This is an electropherogram of the forward primer sequencing of the S gene RBD nested PCR products of sample M22-47 (Figure 3, panel C, lane 29). It shows K417N, N440K, G446S, S477N,

T478K, E484A, Q493R, G496S, Q498R, N501Y and Y505H mutations in the dominant sequence, which is diagnostic of an Omicron variant BA.1.

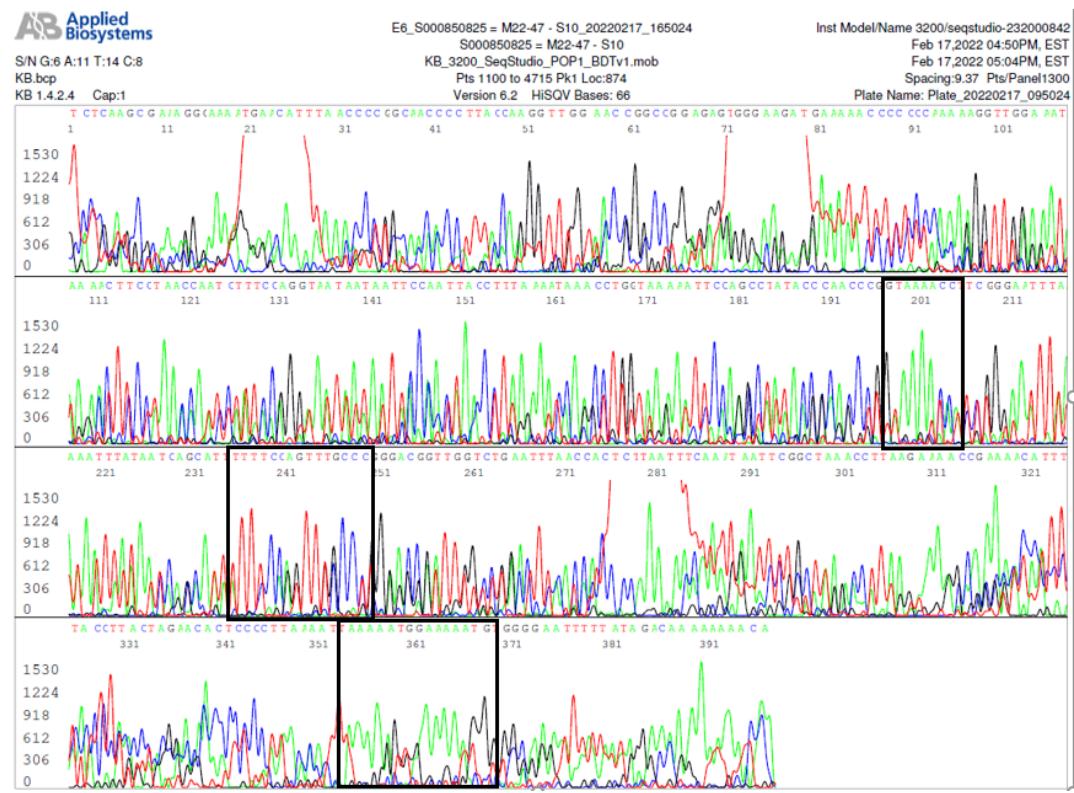


Figure 27. This is an electropherogram of the S gene RBD reverse sequencing of the same nested PCR product that was used to generate the electropherogram presented in Figure 26. Accurate base calling was not possible due to multiple overlapping sequences. But the electropherogram showed at least 3 short 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 8 A).

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.

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.

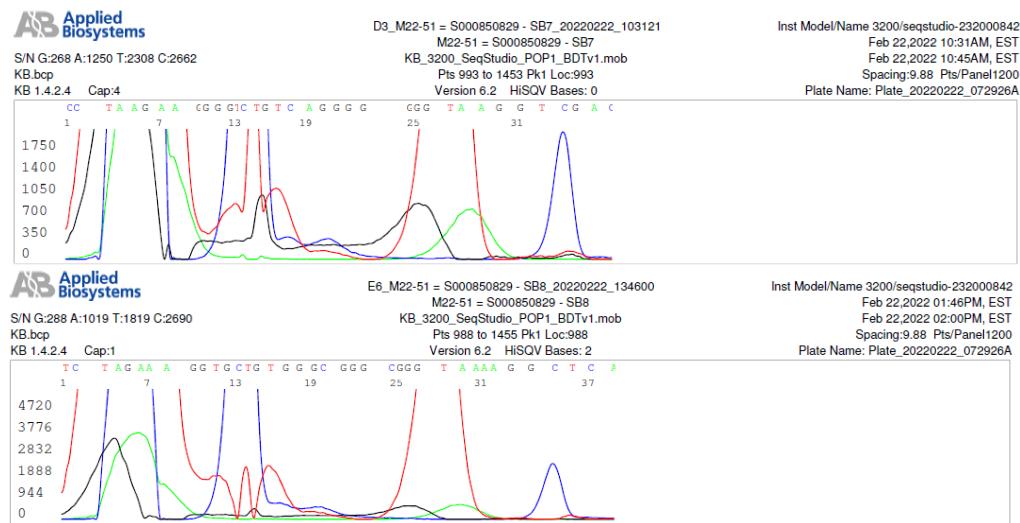


Figure 28. These two bidirectional sequencing electropherograms confirmed that there was no NTD SB7/SB8 nested PCR amplicon on sample M22-51 as shown in Figure 3, Panel G, Lane 51.

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'-TTATGTGGTTATCTTCAACC; 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'-TTT-GCCCTGGAGCGATTGTC. 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 29. This 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).

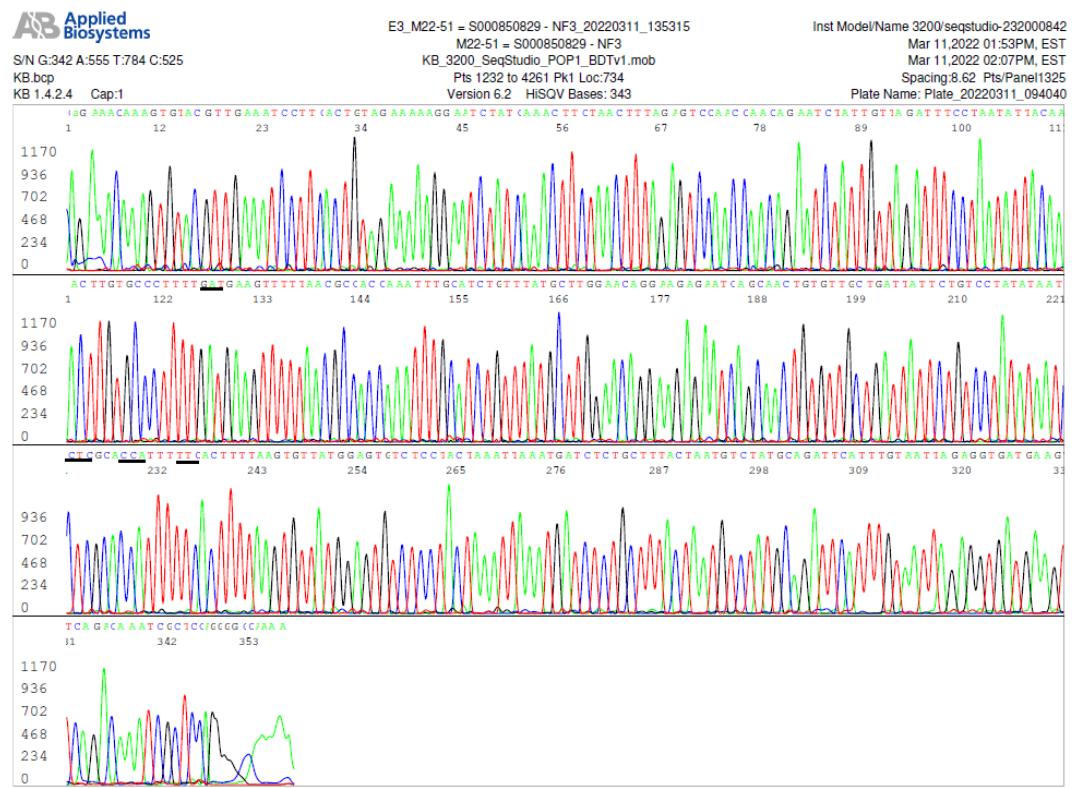


Figure 30. This is an electropherogram of the forward sequencing of the sample M22-51 nested RT-PCR amplicon illustrated in Figure 29, using the forward nested PCR NF3 primer as the sequencing primer. It shows G339D, S371L, S373P and S375F mutations (codons underlined), which are characteristic of an Omicron variant.

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).

```

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

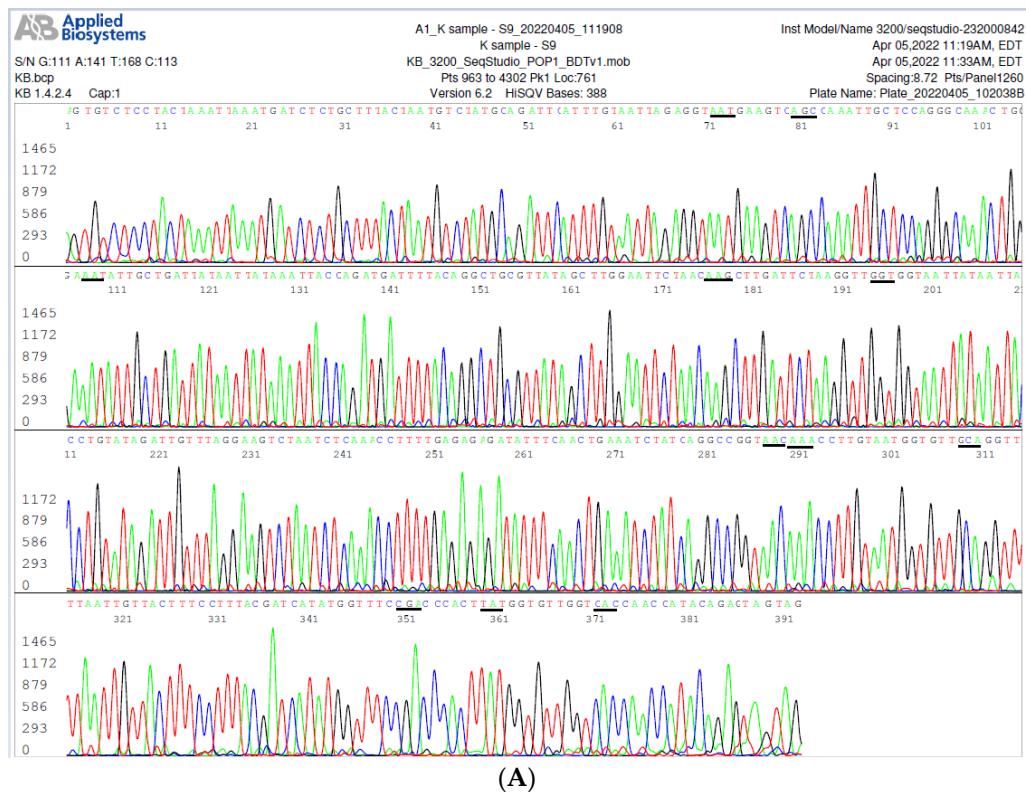
/translation="MFVFLVLLPLVSSQCVNLTRTQLPPAYTNSFTRGVYYPDKVFR
SSVLHSTQDLFLPFFSNVTWFHVISGTNGTKRFDFNPVLPFNDGVYFAIEKSNIIRGW
IFGTTLDSKTQSLLIVNNXXXXVIKCEFQFCNDPFLGVXYHKNNSWMEXXXRVXXX
XXXXXXFEYVSQPFXXDLEGKQGNFKNLREFVFKNIDGYFKIYSKHTPIIVEPEDLPQ
GFSALEPLVDPPIGINITRFQTLALHRSYLTGPDSSSGWTAGAAAYVGYLQPRFL
LKYNENGTITDAVDCALDPLSETKCTLKSFTVEKGIYQTSNFRVQPTESIVRFPNITN
LXXXXXXXXXXXXXXXXXXXXNRKRISNXXXXXSVLYNSASFSTFKCXXXXXXKLNDLCF
XNVYADSFVXXXXXXVRQIAPGQTGKXXDYNYKLPXDFTGXXXXXXXXXXXXKVGGXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXGVEGXXCYFPLQSYGXXXXXXXXXXXX
XXVVLSFXXXXXXXXXXXXXXNLVKNKCVNFNFNXXTGTGVLTESNKKFLPFQQFG
RDIADTTDAVRDPQTLEILDITPCSFGGVSITPGTNTSNQAVLYQGVNCTEVPAI
HADQLPTWRVYSTGSNVFQTRAGCLIGAEYVNNSYECIDIPIGAGICASYQTQTKSHR
RARSVASQSIIAYTXXLGAENSVAYSNNXXXXNFTIXXTTEILPVXXXXTSVXXXX

```

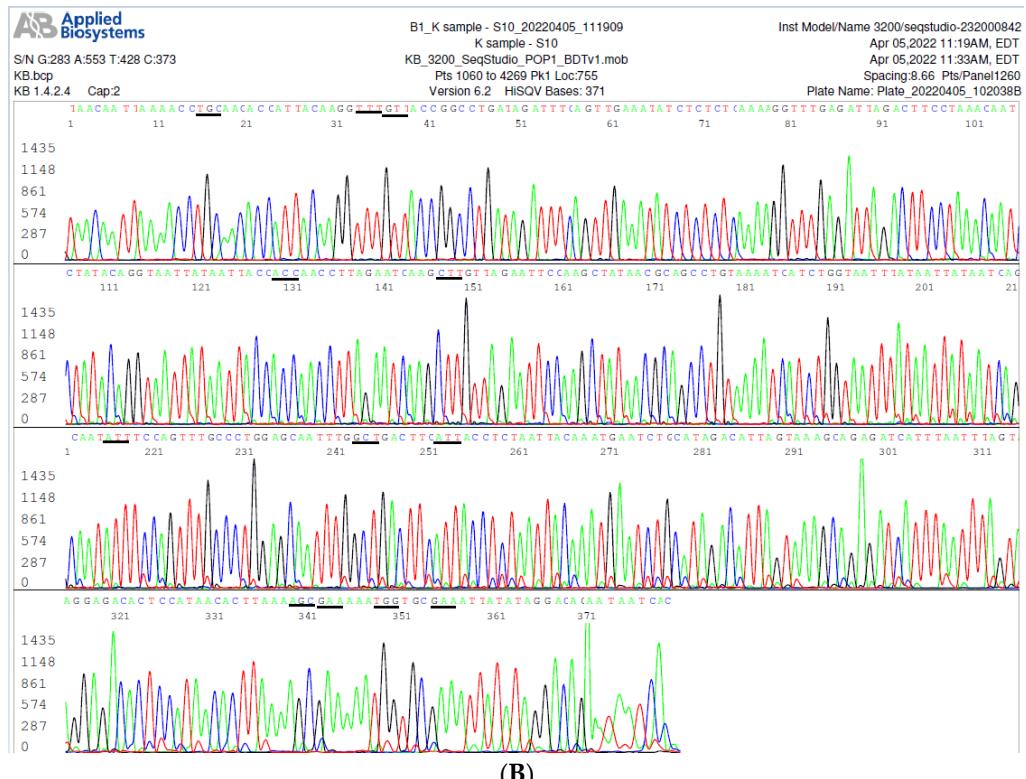
Figure 31. This is an S protein NTD/RBD amino acid sequence retrieved from GenBank Sequence ID# OL898842. The underlined bold letters “VIS”, “I”, “II” and “EPE” marked the sites of mutations “A67V”, “Δ69-70”, “T95I”, “Δ211, L212I”, and “ins214EPE”, respectively. In the GenBank database, the letter X (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.

3.10. Recombined BA.1 NTD and a BA.2 RBD sequence in the Omicron S gene

During the course of the study the nucleic acid of a positive nasopharyngeal swab specimen, which was collected on 3 April, 2022 from an adult patient presenting with sore throat and fatigue, was sequenced. The N gene sequencing showed the R203K and G204R mutations commonly shared by all Omicron subvariants. The S gene RBD sequence showed a profile of the Omicron BA.2 subvariant amino acid mutations (Figure 32 A and B). However, the S gene NTD sequence exhibited A67V, Δ69-70, T95I, G142D and Δ143-145 mutations that are characteristic of a BA.1 Omicron, along with several SNPs, including one base deletion. A competing template in the NTD with an A-to-G mutation indicates the existence of at least two viruses infecting the same host (See Figure 33 A, B and C. Patient identity on all electropherograms has been blinded and labeled as “K Sample”).



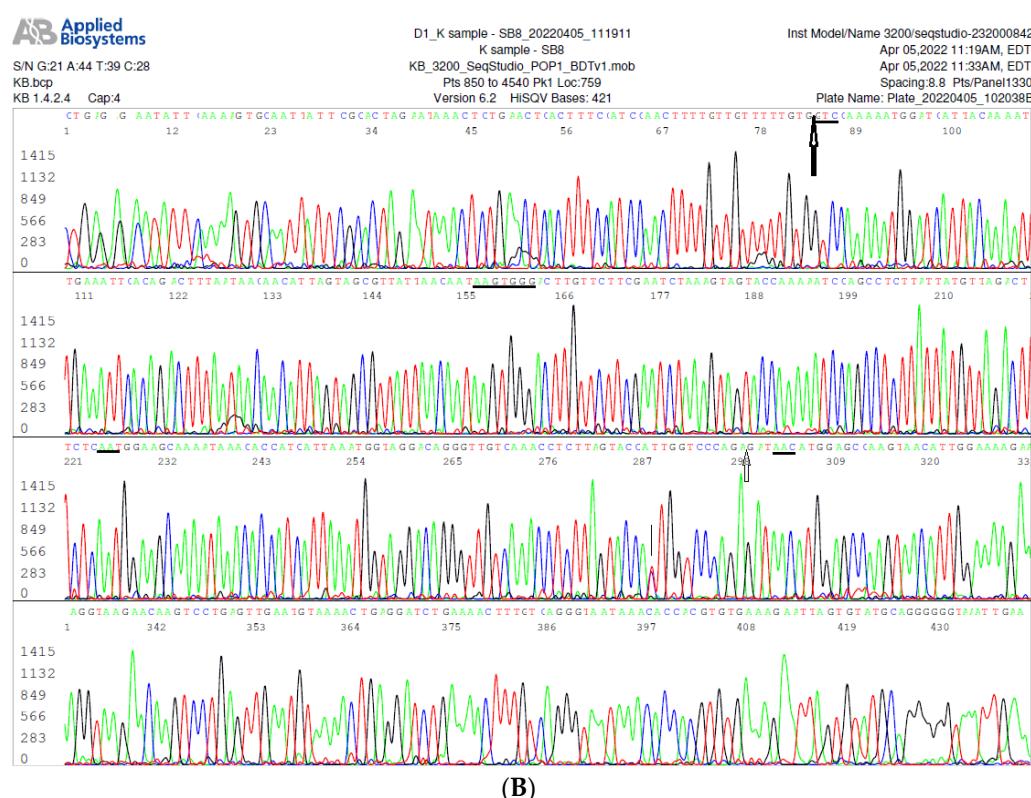
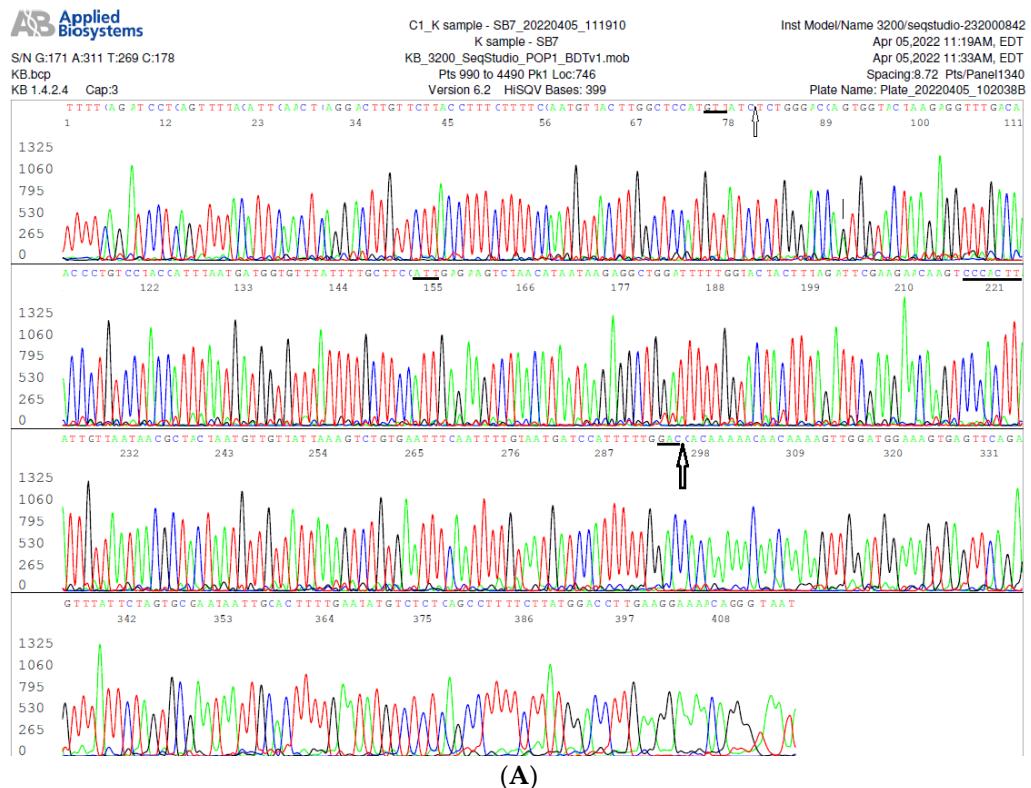
(A)



(B)

Figure 32. A and B. These are 2 electropherograms showing a forward and a reverse Sanger sequencing of a 445-bp segment of the S gene RBD with a mutation profile of S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, G446, S477N, T478K, E484A, Q493R, G496, Q498R, N501Y and Y505H, characteristic of that for Omicron BA.2.

The S gene NTD bidirectional sequencing of the same K sample showed A67V, Δ69-70, T95I, G142D and Δ143-145 mutations, a profile most commonly associated with the Omicron BA.1, not with the BA.2 subvariant (Figure 33 A and B).



Severe acute respiratory syndrome coronavirus 2 genome assembly, chromosome: 1

Sequence ID: [OW180100.1](#) Length: 29873 Number of Matches: 1Range 1: 21616 to 22090 [GenBank](#) [Graphics](#) [Next Match](#) [Previous Match](#)

Score	Expect	Identities	Gaps
854 bits(462)	0.0	471/475(99%)	1/475(0%)
Query_1	TCAATTACCCCTGCATAACATAATTCTTCACACGTGGTGTATTACCCCTGACAAAGT	60	
Subject_21616	TCAATTACCCCTGCATAACATAATTCTTCACACGTGGTGTATTACCCCTGACAAAGT	21675	
Query_61	TTTCAGATCCTCAGTTTACATTCACACTCAGGACTTGTCTTACCTTCTTTCCAAATGT	120	
Subject_21676	TTTCAGATCCTCAGTTTACATTCACACTCAGGACTTGTCTTACCTTCTTTCCAAATGT	21735	
Query_121	TACTTGGCTCCATGTTATCTCTGGGACCAATGGTACTAAGAGGTTTGACAACCCGTCT	180	
Subject_21736	TACTTGGCTCCATGTTATCTCTGGGACCAATGGTACTAAGAGGTTTGACAACCCGTCT	21795	
Query_181	ACCATTAAATGATGGTGTATTCTGCTTCAATGAGAAGTCTAACATATAAGAGGCTG	240	
Subject_21796	ACCATTAAATGATGGTGTATTCTGCTTCAATGAGAAGTCTAACATATAAGAGGCTG	21855	
Query_241	GATTTTGGTACTACTTAGATTCGAAGAACGACCC-ACCTATTGTTAACGCTAC	299	
Subject_21856	GATTTTGGTACTACTTAGATTCGAAGAACGACCCAGTCCCTACTTATTGTTAACGCTAC	21915	
Query_300	TAATGTTGTATTAAAGTGTGAATTTCATTTGTAATGATCCATTGGACCAACAA	359	
Subject_21916	TAATGTTGTATTAAAGTGTGAATTTCATTTGTAATGATCCATTGGACCAACAA	21975	
Query_360	AAACAAACAAAGTGGATGGAAAGTGTGAGTTCAAGAGTTTATTCTAGTGCAGAAATTGCAC	419	
Subject_21976	AAACAAACAAAGTGGATGGAAAGTGTGAGTTCAAGAGTTTATTCTAGTGCAGAAATTGCAC	22035	
Query_420	TTTGAAATATGTCCTCAGCCCTTTCTTATGGACCTTGAAAGGAAACAGGGTAAT	474	
Subject_22036	TTTGAAATATGTCCTCAGCCCTTTCTTATGGACCTTGAAAGGAAACAGGGTAAT	22090	

(C)

Figure 33 A, B and C. Figure 33 A and B are electropherograms showing the S gene NTD bidirectional sequencing results with the A67V, Δ69-70, T95I, G142D and Δ143-145 mutations that are characteristic of the Omicron BA.1 variant (see Figure 9A and Figure 10A), along with several SNPs, including one base deletion, which have not been published in the GenBank database as illustrated in a GenBank BLAST report in Figure 33 C. The single nucleotide deletion of "T" (Figure 33 C) shown by the underlined sequence 5'- CCCACTT in Figure 33 A is confirmed by the underlined sequence 3'-AAGTGGG in Figure 33 B. A competing virus with an A-to-G mutation is indicated by a vertical thin line pointing to a base read as G (peak position 91) by the computer in Figure 33 A, but as T (peak position 288) in Figure 33 B.

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 [34]. 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 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 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 specific RT-PCR primers in early 2020

PCR is a primer-initiated template-directed exponential enzymatic polymerization of 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 non-specificity of PCR amplification by using the CDC-recommended SARS-CoV specific RT-PCR primers and diagnostic protocol [15] for accurate detection of SARS-CoV-2 at the early stage of the COVID-19 outbreak to avoid or to curtail a pandemic. 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 specific RT-PCR primers [15] 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 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 [26].

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 [35-38]. 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 the average. Mismatch impact rapidly declined at positions further away from the 3'-terminal position although there were exceptions [38].

The Sanger sequencing results presented in this paper confirm that the CDC-recommended SARS-CoV Cor-p-R1 (-) reverse PCR primer is able to amplify a corresponding 348-bp target cDNA of the SARS-CoV-2 gene for diagnostic purpose even when there were 3 mismatches in a primer, one of them located at the 3'-terminal position (Figure 1 B). 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 a "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 (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, 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 [39]. 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 [26], 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 [21].

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 [33].

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, Δ31-33, 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, Δ69-70, T95I, G142D and Δ143-145 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 [40] 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 [41]. However, only a small fraction of these intra-host single-nucleotide variants become fixed [42] 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 [43], transmission risk [44], and treatment outcome [45] in other RNA viruses. The existence of these multi-allelic SNPs involving the RBD of the 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 demon-

strated by Sanger sequencing (Figure 13). As previously reported, there were no demonstrable multi-allelic SNPs in the N gene [26] or in the S gene RBD and NTD [28] 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 [26].

4.5. *Continued SARS-CoV-2 mutations need routine sequencing diagnostics*

As the SARS-CoV-2 variants continue circulating in the populations, mixed variant infections and viral recombination may occur, which may generate ambiguous gene sequencing data as demonstrated in a recent case presented in Figures 32 and 33. Routine Sanger or next-generation sequencing at least the S gene RBD and NTD [46, 47] of the PCR-positive samples is needed to monitor the potential impacts of these Omicron sub-variant and their recombined variant infections on critical COVID-19 countermeasures, including vaccines, therapeutics, and diagnostics. One group of investigators suggested using RT-PCR to generate a 733-bp amplicon of the RBD sequence and send the PCR amplicon to a commercial laboratory for Sanger sequencing to determine the variants [48]. Another group developed specific primers and probes for RT-qPCR to detect mutations in the S gene for variant determination [49]. Their Sanger sequencing results [48] and RT-qPCR probe-based test results [49] were reported to be fully comparable to those generated by whole-genome sequencing. While whole-genome sequencing relying on the NGS technology is widely applied, varying error rates in NGS have been observed [50]. The first genomic sequences of SARS-CoV-2 isolates from patient specimens in China [51] and in the United States [52] were verified by Sanger sequencing to avoid base-calling errors generated by NGS.

5. Conclusions

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 and for variant determination. Routine partial S gene sequencing can timely discover multi-allelic SNPs and potential viral recombination in the circulating variants for monitoring their potential impacts on vaccine efficacies, therapeutics and diagnostics.

Funding: This research was partially funded by The Institute for Pure and Applied Knowledge and The Energetic Health Institute.

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.

Conflicts of Interest: Sin Hang Lee is Director of the Milford Molecular Diagnostics Laboratory specialized in developing DNA sequencing-based diagnostic tests implementable in community hospital laboratories.

1. Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., Chen, J., ... Tan, W. (2020). Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet (London, England)*, 395(10224), 565–574. [https://doi.org/10.1016/S0140-6736\(20\)30251-8](https://doi.org/10.1016/S0140-6736(20)30251-8)
2. Zhou, P., Yang, X. L., Wang, X. G., Hu, B., Zhang, L., Zhang, W., Si, H. R., Zhu, Y., Li, B., Huang, C. L., Chen, H. D., Chen, J., Luo, Y., Guo, H., Jiang, R. D., Liu, M. Q., Chen, Y., Shen, X. R., Wang, X., Zheng, X. S., ... Shi, Z. L. (2020). A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, 579(7798), 270–273. <https://doi.org/10.1038/s41586-020-2012-7>
3. Shang, J., Wan, Y., Luo, C., Ye, G., Geng, Q., Auerbach, A., & Li, F. (2020). Cell entry mechanisms of SARS-CoV-2. *Proceedings of the National Academy of Sciences of the United States of America*, 117(21), 11727–11734. <https://doi.org/10.1073/pnas.2003138117>
4. Coronavirus Updates. Worldometer. Available online: <https://www.worldometers.info/coronavirus/> (accessed on 4 April 2022).
5. CDC. Severe Acute Respiratory Syndrome (SARS). Available online: <https://www.cdc.gov/sars/about/faq.html> (accessed on 4 March 2022).
6. Hu, Z., Song, C., Xu, C., Jin, G., Chen, Y., Xu, X., Ma, H., Chen, W., Lin, Y., Zheng, Y., Wang, J., Hu, Z., Yi, Y., & Shen, H. (2020). Clinical characteristics of 24 asymptomatic infections with COVID-19 screened among close contacts in Nanjing, China. *Science China. Life sciences*, 63(5), 706–711. <https://doi.org/10.1007/s11427-020-1661-4>
7. Liu, W. Y., Chuang, Y. C., Liu, T. J., Chien, C. W., & Tung, T. H. (2021). Insights from the comparisons of SARS-CoV and COVID-19 outbreaks: The evidence-based experience of epidemic prevention in China. *Medicine*, 100(6), e24650.
8. Zeng, C., Evans, J. P., King, T., Zheng, Y. M., Oltz, E. M., Whelan, S., Saif, L. J., Peebles, M. E., & Liu, S. L. (2022). SARS-CoV-2 spreads through cell-to-cell transmission. *Proceedings of the National Academy of Sciences of the United States of America*, 119(1), e2111400119. <https://doi.org/10.1073/pnas.2111400119>
9. Zhang, X., Wu, S., Wu, B., Yang, Q., Chen, A., Li, Y., Zhang, Y., Pan, T., Zhang, H., & He, X. (2021). SARS-CoV-2 Omicron strain exhibits potent capabilities for immune evasion and viral entrance. *Signal transduction and targeted therapy*, 6(1), 430. <https://doi.org/10.1038/s41392-021-00852-5>
10. Lambrou, A. S., Shirk, P., Steele, M. K., Paul, P., Paden, C. R., Cadwell, B., Reese, H. E., Aoki, Y., Hassell, N., Zheng, X. Y., Talarico, S., Chen, J. C., Oberste, M. S., Batra, D., McMullan, L. K., Halpin, A. L., Galloway, S. E., MacCannell, D. R., Kondor, R., Barnes, J., ... Strain Surveillance and Emerging Variants NS3 Working Group (2022). Genomic Surveillance for SARS-CoV-2 Variants: Predominance of the Delta (B.1.617.2) and Omicron (B.1.1.529) Variants - United States, June 2021–January 2022. *MMWR. Morbidity and mortality weekly report*, 71(6), 206–211. <https://doi.org/10.15585/mmwr.mm7106a4>
11. Godlee, F. Caution, vaccines, testing: the only way forward. *BMJ* 2021;374:n1781 <http://dx.doi.org/10.1136/bmj.n1781>
12. FDA to CDC. Letter Dated 1 December 2020. Available online: <https://www.fda.gov/media/134919/download> (accessed on 4 March 2022)
13. Zhong, N. S., Zheng, B. J., Li, Y. M., Poon, Xie, Z. H., Chan, K. H., Li, P. H., Tan, S. Y., Chang, Q., Xie, J. P., Liu, X. Q., Xu, J., Li, D. X., Yuen, K. Y., Peiris, & Guan, Y. (2003). Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet (London, England)*, 362(9393), 1353–1358. [https://doi.org/10.1016/S0140-6736\(03\)14630-2](https://doi.org/10.1016/S0140-6736(03)14630-2)
14. Drosten, C.; Preiser, W.; Günther, S.; Schmitz, H.; Doerr, H.W. Severe acute respiratory syndrome: Identification of the etiological agent. *Trends Mol. Med.* 2003, 9, 325–327.
15. 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).
16. CDC. Severe Acute Respiratory Syndrome (SARS). Available online: <https://www.cdc.gov/sars/about/faq.html> (accessed on 4 March 2022).
17. Taleghani, N., & Taghipour, F. (2021). Diagnosis of COVID-19 for controlling the pandemic: A review of the state-of-the-art. *Biosensors & bioelectronics*, 174, 112830. <https://doi.org/10.1016/j.bios.2020.112830>
18. CDC. Research Use Only 2019-Novel Coronavirus (2019-nCoV) Real-time RT-PCR Primers and Probes Updated June 6, 2020 <https://www.cdc.gov/coronavirus/2019-ncov/lab/rt-pcr-panel-primer-probes.html> (accessed on 4 March 2022).
19. eCDC. Implications of the emergence and spread of the SARS CoV 2 B.1.1.529 variant of concern (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> (accessed on 4 March 2022).

20. FDA. Tests Expected to Fail to Detect the SARS-CoV-2 Omicron Variant (As of 12/27/2021) https://www.fda.gov/medical-devices/coronavirus-covid-19-and-medical-devices/sars-cov-2-viral-mutations-impact-covid-19-tests?utm_medium=email&utm_source=govdelivery#omicron-reduced (accessed on 4 March 2022).

21. 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-diagnosticseuas> (accessed on 19 January 2021)

22. CDC. Update on the SARS-CoV-2 Omicron Variant. 12/03/2021: Lab Alert: https://www.cdc.gov/csels/dls/locs/2021/12-03-2021-lab-alert-CDC_Update_SARS-CoV-2_Omicron_Variant.html (accessed on 4 March 2022).

23. Walker, A., Houwaart, T., Wienemann, T., Vasconcelos, M. K., Strelow, D., Senff, T., Hülse, L., Adams, O., Andree, M., Hauka, S., Feldt, T., Jensen, B. E., Keitel, V., Kindgen-Milles, D., Timm, J., Pfeffer, K., & Dilthey, A. T. (2020). Genetic structure of SARS-CoV-2 reflects clonal superspreading and multiple independent introduction events, North-Rhine Westphalia, Germany, February and March 2020. *Euro surveillance : bulletin European sur les maladies transmissibles = European communicable disease bulletin*, 25(22), 2000746. <https://doi.org/10.2807/1560-7917.ES.2020.25.22.2000746>

24. Gupta, K., Toelzer, C., Williamson, M. K., Shoemark, D. K., Oliveira, A., Matthews, D. A., Almuqrin, A., Staufer, O., Yadav, S., Borucu, U., Garzoni, F., Fitzgerald, D., Spatz, J., Mulholland, A. J., Davidson, A. D., Schaffitzel, C., & Berger, I. (2022). Structural insights in cell-type specific evolution of intra-host diversity by SARS-CoV-2. *Nature communications*, 13(1), 222. <https://doi.org/10.1038/s41467-021-27881-6>

25. Wang, Y., Wang, D., Zhang, L., Sun, W., Zhang, Z., Chen, W., Zhu, A., Huang, Y., Xiao, F., Yao, J., Gan, M., Li, F., Luo, L., Huang, X., Zhang, Y., Wong, S. S., Cheng, X., Ji, J., Ou, Z., Xiao, M., ... Zhao, J. (2021). Intra-host variation and evolutionary dynamics of SARS-CoV-2 populations in COVID-19 patients. *Genome medicine*, 13(1), 30. <https://doi.org/10.1186/s13073-021-00847-5>

26. Lee, S.H. 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 Stud. Opin.* 2021, 2, 230–278.

27. Lee, S.H. Testing for SARS-CoV-2 in cellular components by routine nested RT-PCR followed by DNA sequencing. *Int. J. Geriatr. Rehabil.* 2020, 2, 69–96.

28. Lee, S.H. 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>

29. CDC. SARS-CoV-2 Variant Classifications and Definitions. Available online: <https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html>

30. 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>

31. CoV-lineages Gibhub. Proposal to split B.1.1.529 to incorporate a newly characterised sibling lineage #361 <https://github.com/cov-lineages/pango-designation/issues/361>

32. Lee S. H. (2016). Lyme disease caused by *Borrelia burgdorferi* with two homeologous 16S rRNA genes: a case report. *International medical case reports journal*, 9, 101–106. <https://doi.org/10.2147/IMCRJ.S99936>

33. Shagin, D. A., Lukyanov, K. A., Vagner, L. L., & Matz, M. V. (1999). Regulation of average length of complex PCR product. *Nucleic acids research*, 27(18), e23. <https://doi.org/10.1093/nar/27.18.e23>

34. Appenzeller, T. Democratizing the DNA sequence. *Science*. 1990, 247, 1030–1032.

35. Kwok, S.; Kellogg, D.E.; McKinney, N.; Spasic, D.; Goda, L.; L.evenson, C.; Sninsky, J.J. Effects of primer-template mismatches on the polymerase chain reaction: human immunodeficiency virus type 1 model studies. *Nucleic Acids Res.* 1990, 18, 999–1005. doi: 10.1093/nar/18.4.999. PMID: 2179874; PMCID: PMC330356.

36. O'Dell, S.D.; Humphries, S.E.; Day, I.N. PCR induction of a TaqI restriction site at any CpG dinucleotide using two mismatched primers (CpG-PCR). *Genome Res.* 1996, 6, 558–568. doi: 10.1101/gr.6.6.558. PMID: 8828045.

37. Ruiz-Villalba, A.; van Pelt-Verkuil, E.; Gunst, Q.D.; Ruijter, J.M.; van den Hoff, M.J. Amplification of nonspecific products in quantitative polymerase chain reactions (qPCR). *Biomol. Detect. Quantif.* 2017, 14, 7–18. doi: 10.1016/j.bdq.2017.10.001. PMID: 29255685

38. Stadhouders, R., Pas, S. D., Anber, J., Voermans, J., Mes, T. H., & Schutten, M. (2010). The effect of primer-template mismatches on the detection and quantification of nucleic acids using the 5' nuclease assay. *The Journal of molecular diagnostics: JMD*, 12(1), 109–117. <https://doi.org/10.2353/jmoldx.2010.090035>

39. Ryu, K.H.; Choi, S.H.; Lee, J.S. Restriction primers as short as 6-mers for PCR amplification of bacterial and plant genomic DNA and plant viral RNA. *Mol Biotechnol.* 2000, 14, 1–3.

40. Narayanan, S., Ritchey, J. C., Patil, G., Narasaraju, T., More, S., Malayer, J., Saliki, J., Kaul, A., Agarwal, P. K., & Ramachandran, A. (2021). SARS-CoV-2 Genomes From Oklahoma, United States. *Frontiers in genetics*, 11, 612571. <https://doi.org/10.3389/fgene.2020.612571>

41. Callaway E. (2021). Beyond Omicron: what's next for COVID's viral evolution. *Nature*, 600(7888), 204–207. <https://doi.org/10.1038/d41586-021-03619-8>

42. Li, J., Du, P., Yang, L., Zhang, J., Song, C., Chen, D., Song, Y., Ding, N., Hua, M., Han, K., Song, R., Xie, W., Chen, Z., Wang, X., Liu, J., Xu, Y., Gao, G., Wang, Q., Pu, L., Di, L., ... Chen, C. (2022). Two-step fitness selection for intra-host variations in SARS-CoV-2. *Cell reports*, 38(2), 110205. <https://doi.org/10.1016/j.celrep.2021.110205>

43. Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E., & Andino, R. (2006). Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature*, 439(7074), 344–348. <https://doi.org/10.1038/nature04388>

44. Poon, L. L., Song, T., Rosenfeld, R., Lin, X., Rogers, M. B., Zhou, B., Sebra, R., Halpin, R. A., Guan, Y., Twaddle, A., DePasse, J. V., Stockwell, T. B., Wentworth, D. E., Holmes, E. C., Greenbaum, B., Peiris, J. S., Cowling, B. J., & Ghedin, E. (2016). Quantifying influenza virus diversity and transmission in humans. *Nature genetics*, 48(2), 195–200.

45. To, K. K., Chan, J. F., Chen, H., Li, L., & Yuen, K. Y. (2013). The emergence of influenza A H7N9 in human beings 16 years after influenza A H5N1: a tale of two cities. *The Lancet. Infectious diseases*, 13(9), 809–821. [https://doi.org/10.1016/S1473-3099\(13\)70167-1](https://doi.org/10.1016/S1473-3099(13)70167-1)

46. ECDC and WHO Regional Office for Europe. Methods for the detection and characterisation of SARS-CoV-2 variants – first update. 20 December 2021 <https://www.ecdc.europa.eu/en/publications-data/methods-detection-and-characterisation-sars-cov-2-variants-first-update>

47. Peronace, C.; Tallerico, R.; Colosimo, M.; De Fazio, M.; Pasceri, F.; Talotta, I.; Panduri, G.; Pintomalli, L.; Oteri, R.; Calantoni, V.; Fiorillo, M.T.; Gallelli, L.; Cione, E.; Minchella, P. BA.1 Omicron Variant of SARS-CoV-2: First Case Reported in Calabria Region, Italy. *COVID 2022*, 2, 211-215. <https://doi.org/10.3390/covid2030016>

48. Bloemen, M., Rector, A., Swinnen, J., Ranst, M. V., Maes, P., Vanmechelen, B., & Wollants, E. (2022). Fast detection of SARS-CoV-2 variants including Omicron using one-step RT-PCR and Sanger sequencing. *Journal of virological methods*, 304, 114512. <https://doi.org/10.1016/j.jviromet.2022.114512>

49. Chung, H. Y., Jian, M. J., Chang, C. K., Lin, J. C., Yeh, K. M., Chen, C. W., Hsieh, S. S., Hung, K. S., Tang, S. H., Perng, C. L., Chang, F. Y., Wang, C. H., & Shang, H. S. (2022). Emergency SARS-CoV-2 Variants of Concern: Novel Multiplex Real-Time RT-PCR Assay for Rapid Detection and Surveillance. *Microbiology spectrum*, 10(1), e0251321. <https://doi.org/10.1128/specrum.02513-21>

50. Pfeiffer, F.; Gröber, C.; Blank, M.; Händler, K.; Beyer, M.; Schultze, J.L.; Mayer, G. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Sci. Rep.* 2018, 8, 10950. [CrossRef]

51. Ren, L.-L.; Wang, Y.-M.; Wu, Z.-Q.; Xiang, Z.-C.; Guo, L.; Xu, T.; Jiang, Y.-Z.; Xiong, Y.; Li, Y.-J.; Li, X.-W.; et al. Identification of a novel coronavirus causing severe pneumonia in human: A descriptive study. *Chin. Med. J.* 2020, 133, 1015–1024. [CrossRef]

52. Harcourt, J.; Tamin, A.; Lu, X.; Kamili, S.; Sakthivel, S.K.; Murray, J.; Queen, K.; Tao, Y.; Paden, C.R.; Zhang, J.; et al. Severe Acute Respiratory Syndrome Coronavirus 2 from Patient with Coronavirus Disease, United States. *Emerg. Infect. Dis.* 2020, 26, 1266–1273. [CrossRef]