 doi:10.20944/preprints202206.0192.v1

Article

Implementation of the eCDC/WHO Recommendation for Molecular Diagnosis of SARS-CoV-2 Omicron Subvariants and Its Challenges

Sin Hang Lee

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

Abstract: Large population passages of the SARS-CoV-2 in the past two and a half years have allowed the circulating virus to accumulate an increasing number of mutations in its genome. The most recently emerging Omicron subvariants have the highest number of mutations in the Spike (S) protein gene and these mutations mainly occur in the receptor-binding domain (RBD) and the N-terminal domain (NTD) of the S gene. The eCDC and the WHO recommend partial Sanger sequencing of the SARS-CoV-2 S gene RBD and NTD on the PCR-positive samples in diagnostic laboratories as a practical means of determining the variants of concern to monitor a possible increased transmissibility, increased virulence, or reduced effectiveness of vaccines against them. The author's diagnostic laboratory has implemented the eCDC/WHO recommendation by sequencing a 398-base segment of the N gene for the definitive detection of SARS-CoV-2 in clinical samples, and sequencing a 445-base segment of the RBD and a 490-509-base segment of the NTD for variant determination. This paper presents 5 selective cases to illustrate the challenges of using Sanger sequencing to diagnose Omicron subvariant when the samples harbor a high level of co-existing minor subvariant sequences with multi-allelic single nucleotide polymorphisms (SNPs) or possible recombinant Omicron subvariants containing a BA.1 NTD and a BA.2 RBD, which can only be detected by using specially designed PCR primers. The current large-scale surveillance programs using next-generation sequencing (NGS) do not face similar problems because NGS focuses on deriving consensus sequence.

Keywords: eCDC; WHO; Sanger sequencing; Omicron variant; minor subvariants; BA.4/BA.5; BA.2; multi-allelic; SNPs; recombinant

1. Introduction

SARS-CoV-2 first emerged in Wuhan, China in December, 2019 and spread to other parts of the world, causing more than 6 million global human deaths [1]. Since SARS-CoV-2 is an RNA virus, nucleotide mutations are expected to occur in its >29,000-base genome due to the well-known high copying error rates of the RNA-dependent RNA polymerase [2]. These enzymatic copying errors invariably generate a large number of nonsynonymous nucleotide substitutions, commonly referred to as mutations, in the virus genome. In any given SARS-CoV-2 infection, there are probably thousands of viral particles each with unique single-nucleotide mutations in the host [3]. Only a small fraction of these intra-host single-nucleotide variants become fixed [4] to be passed to the next generation to infect another host. However, after more than two years of non-stopped transmissions from host to host, a large number of amino acid mutations have accumulated to create the Omicron variant with multiple subvariants [5].

Molecular epidemiological studies on RNA viruses often focus on per-host consensus sequences [6], which tend to summarize each virus population into a single sequence and ignore minor variants. Since the infective dose of the SARS-CoV-2 when administered through aerosols to susceptible human subjects is estimated to be between 1950 and 3000

 doi:10.20944/preprints202206.0192.v1

virions [7,8], some virions of minor variants with multi-allelic single nucleotide polymorphisms (SNPs) [9,10] may also transmit from one host to another [11]. When a new consensus sequence of the S protein gene displays a tendency to become the dominant sequence in the circulating SARS-CoV-2 within a human population, the new consensus sequence is referred to as a new variant of concern (VOC) or a new variant of interest (VOI) in an attempt to correlate the virus variants containing these newly emerging amino acid mutation profiles with a possible increased transmissibility, increased virulence, or reduced effectiveness of vaccines against them [12,13].

It took 10 months for the first VOC to be recognized in the United Kingdom in October 2020, and labeled as the Alpha variant [14]. Subsequently, numerous variants were reported from different countries, including the Alpha, Beta, Gamma, Delta and Omicron variants. From January 2022, the Omicron variant and its subvariants, all characterized by a high number of amino acid mutations in the ACE2 RBD, have largely replaced other VOCs; the original SARS-CoV-2 Wuhan-Hu-1 strains are now rarely detected. The current VOCs are the Omicron BA.1, the Omicron BA.2 and the Delta variant, according to the updated report of the European Centre for Disease Prevention and Control (eCDC) while the Omicron BA.4 and BA.5 are listed as VOIs [15]. The future waves of coronavirus will probably be driven by newer, fitter descendants of the Omicron variant [16].

In the United States, there are no 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 Federal Drug Administration (FDA) to check broadly for the SARS-CoV-2 virus, not for specific variants [17].

After the emergence of the Omicron variant, the eCDC and the World Health Organization (WHO) jointly published the first update of "Methods for the detection and characterisation of SARS-CoV-2 variants" on 20 December 2021, recommending amplicon-based Sanger sequencing of the RBD and the N-terminal domain (NTD) of the S gene to reliably differentiate between the circulating variants in diagnostic laboratories [18]. It is generally believed that RBD mutations are associated with changing infectivity of the virus [19] while NTD mutations/deletions alter the epitope structure and thus affect the immunoreactivity of the spike protein [20]. The eCDC/WHO recommend that when NAAT-based assays are used, confirmatory sequencing of at least a subset of samples should be performed [18]. Due to the high cost, the low sensitivity and the requirement of bioinformatic data analysis, whole genome sequencing (WGS) cannot be implemented in all diagnostic laboratories. In comparison, Sanger sequencing of the S gene can be more feasible and timely than WGS [18]. The eCDC/WHO document also listed a reference that suggested using 7 PCR primers to amplify a 1071bp segment of the NTD and a 1068bp segment of the RBD followed by heminested PCR to generate templates for Sanger sequencing. However, no actual test data have been published to show that such a protocol has been implemented in diagnostic laboratories.

When complex clinical specimens are tested for the presence or absence of a foreign nucleic acid in small quantities, the PCR amplicon of the target DNA or cDNA is usually limited to <500 bp in size. PCR amplification of a 405-bp fragment from the SARS-CoV genome for sequencing and comparing the sequence of the amplicon with reference sequences in the GenBank database was the established method for molecular detection of SARS-CoV during the 2003 SARS outbreak [21,22]. The U.S. CDC's diagnostic protocol for SARS-CoV recommended using three specific primers to perform RT-PCR to amplify a 348-bp genomic cDNA for sequencing "to verify the authenticity of the amplified product" [23]. Attempts to amplify big-sized templates in complex samples often lead to PCR failures [24] although nested PCR may raise the detection sensitivity.

The author of this article followed the recommendations of the eCDC and the WHO [18] and the U.S. CDC's protocol established for SARS-Co-V diagnosis [23] to design an implementable method to sequence a 398-bp N gene amplicon for the definitive detection of SARS-CoV-2 [25,26] followed by sequencing a 445-bp cDNA amplicon of the RBD and a 490-bp cDNA amplicon of the NTD of the S gene to reliably differentiate between the SARS-CoV-2 variants, including the Omicron BA.1 variants in a group of nasopharyngeal

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

swab specimens collected in January, 2022 in the United States [27]. However, since the BA.2 subvariants with their unique LPPA24S mutations in the S gene NTD are becoming more prevalent, the PCR primers designed for the detection of A67V and Δ69-70 may fail to amplify the BA.2 NTD sequence and need to be modified to avoid PCR failures. Although single-nucleotide mutations are more common in the RBD of the S gene, the NTD sequence is more prone to deletions and insertions [28]. Nucleotide deletions and insertions affecting the primer-binding sites invariably lead to PCR failures.

The Omicron variant as a group has many more nucleotide mutations in the consensus or dominant S gene sequence than the earlier variants [15, 29]. Since mutations occur randomly, each clinical sample containing a consensus or a dominant Omicron S gene may also harbor more minor subvariant S gene sequences with multi-allelic SNPs, which are inevitably co-amplified along with the dominant sequence during the PCR amplification process. Co-existence of the PCR products of these minor subvariant sequences with multi-allelic SNPs may cause failures in PCR amplification and in DNA sequencing designed to detect a consensus or a dominant nucleotide sequence. A search of the GenBank database revealed that among the Omicron variant sequences recently deposited into the GenBank, for example, in GenBank Seq ID# OL898842, ON337825 and ON347156, there are major undetermined sequence segments of the S gene RBD and the NTD while the sequence of the concomitant N gene is fully and properly deciphered, indicating an uneven distribution of mutations between different genes in the circulating SARS-CoV-2 and the need to avoid using the S gene sequence as the target for PCR-based diagnostics.

In the United States, since April, 2022 the Omicron BA.2 subvariants have out-competed the BA.1 variant, which had dominated the positive specimens collected in January, 2022 [27]. Compared to the BA.1 variant, all major BA.2 subvariants contain three additional T376A, D405N and R408S mutations, and lack the G446S and G496S mutations in the RBD. Currently, the major worldwide circulating Omicron variants are the BA.2.12.1 with additional L452Q, the BA.2.13 with additional L452M, and the BA.4/BA.5 with additional L452R and F486V in the RBD [30]. The BA.2.12.1 subvariant is given most attention by the American news media [31].

To follow the eCDC/WHO recommendation of partial Sanger sequencing of the RBD and the NTD of PCR-positive samples, this paper reports the need for extending the forward PCR primer further outward to bypass the Omicron BA.2 LPPA24S mutation site so that one set of general PCR primers can be used to amplify a common S gene NTD target of all variants to be used as the template for sequencing. It also presents Sanger sequencing evidence to show that there may be a highly mutated BA.1 S gene NTD as allele sequence in samples infected by an Omicron BA.2 subvariant. Since the NTD and the RBD of the S protein are known to play different roles in the pathogenesis of SARS-CoV-2 infections, such within-host viral population diversity should be brought to the attention of the laboratories, which are performing large-scale WGS, using NGS to derive a consensus sequence on each sample.

2. Materials and Methods

2.1. Patient samples studied

Five (5) selective nasopharyngeal swab specimens collected from non-hospitalized patients with respiratory infection, which were confirmed to be true-positive for SARS-CoV-2 Omicron variant by Sanger sequencing, were further analyzed by bidirectional Sanger sequencing of 3 genomic targets to show the effects of minor subvariant sequences with multi-allelic SNPs on sequencing-based variant diagnosis. Three (3) of these samples belonging to the Omicron BA.2 sub-lineage were collected in April, 2022, one belonging to the Omicron BA.4/BA.5 was collected in June, 2022 and one (1) belonging to the BA.1 sub-lineage was collected in January, 2022. Written consents were obtained from the 4 patients whose samples were positive for BA.2 or BA.4/BA.5 subvariant to allow their samples to be further analyzed for publication. The sample positive for BA.1 subvariant was commercially supplied with independent IRB certification.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

2.2. RNA Extraction from Nasopharyngeal Swab Specimens

As previously reported [25-27], the cellular pellet derived from about 1 mL of the nasopharyngeal swab rinse along with 0.2 mL supernatant after centrifugation was first digested in a buffered solution containing sodium dodecyl sulfate and proteinase K. The digestate was extracted with phenol. The nucleic acid was precipitated by ethanol and redissolved in 50 µL of DEPC-treated water.

2.3. PCR Conditions

The primary and nested RT-PCR conditions were described in detail previously [25-27]. Briefly, to initiate the primary RT-PCR, a total volume of 25 µL mixture was made in a PCR tube containing 20 µL of ready-to-use LoTemp® PCR mix with denaturing chemicals (HiFi DNA Tech, LLC, Trumbull, CT, USA), 1 µL (200 units) of Invitrogen SuperScript III Reverse Transcriptase, 1 µL (40 units) of Ambion™ RNase Inhibitor, 0.1 µL of Invitrogen 1 M DTT (dithiothreitol), 1 µL of 10 µmolar forward primer in TE buffer, 1 µL of 10 µmolar reverse primer in TE buffer and 1 µL of sample RNA 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 mixture was a 25 µL volume of complete PCR mixture containing 20 µL of ready-to-use LoTemp® mix, 1 µL of 10 µmolar forward primer, 1 µL of 10 µmolar reverse primer and 3 µL of molecular grade water.

To initiate the nested PCR, a trace (about 0.2 µL) of the 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.

The crude nested PCR products showing an expected amplicon at agarose gel electrophoresis were subjected to automated Sanger sequencing without further purification.

2.4. DNA Sequencing

About 0.2 µL of the nested PCR products was transferred by a micro-glass rod into a Sanger reaction tube containing 1 µL of 10 µmolar sequencing primer, 1 µL of BigDye® Terminator (v 1.1/Sequencing Standard Kit), 3.5 µL 5× buffer, and 14.5 µL water in a total volume of 20 µL for 20 enzymatic primer extension/termination reaction cycles according to the protocol supplied by the manufacturer (Applied Biosystems, Foster City, CA, USA).

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

2.5. PCR primers and amplicons

Seven (7) sets of primary RT-PCR primers and their corresponding nested PCR primers, which were used to generate nested PCR products to be used as the templates for Sanger sequencing, are listed in Table 1. To maintain detection sensitivity, the maximum size of the primary RT-PCR amplicon was limited to 530 bp for routine diagnostics.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

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

PCR Amplicon	Primer Type	Start	End	Sequence 5'-3'	Size (bp)
N gene Co4/Co3 Nested	Co1 primary F.	28707	28727	ACATTGGCACCCGCAATCCTG	416
	Co8 primary R.	29102	29122	TTGGGTTTGTTCTGGACCACG	
	Co4 nested F.	28720	28740	CAATCCTGCTAACAATGCTGC	398
	Co3 nested R.	29097	29117	TTTGTTCTGGACCACGTCTGC	
S gene RBD S9/S10 Nested	SS1 primary F.	22643	22663	TGTGTTGCTGATTATTCTGTC	460
	SS2 primary R.	23082	23102	AAAGTACTACTACTCTGTATG	
	S9 nested F.	22652	22672	GATTATTCTGTCCTATATAAT	445
	S10 nested R.	23076	23096	CTACTACTCTGTATGGTTGGT	
S gene NTD SB12/SB8 Nested	SB11 primary F.	21594	21614	TCTCTAGTCAGTGTGTTAATC	530
	SB6 primary R.	22103	22123	TTTGAAATTACCCTGTTTTCC	
	SB12 nested F.	21609	22629	TTAATCTTACAACCAGAACTC	509
	SB8 nested R.	22097	22117	ATTACCCTGTTTTCCTTCAAG	
S gene NTD SB7/SB8 Nested	SB5 primary F.	21619	21639	AACCAGAACTCAATTACCCCC	505
	SB6 primary R.	22103	22123	TTTGAAATTACCCTGTTTTCC	
	SB7 nested F.	21628	21648	TCAATTACCCCCTGCATACAC	490
	SB8 nested R.	22097	22117	ATTACCCTGTTTTCCTTCAAG	
S gene RBD NF3/NR4 Nested	PF1 primary F.	22357	22377	TTATGTGGGTTATCTTCAACC	451
	PR2 primary R.	22787	22807	AGTTTGCCCTGGAGCGATTTG	
	NF3 nested F.	22361	22381	GTGGGTTATCTTCAACCTAGG	445
	NR4 nested R.	22785	22805	TTTGCCCTGGAGCGATTTGTC	
S gene NTD/RBD Junction for Omicron BA.1 & BA.2 SB13/NR4 Nested	SB14 primary F.	22107	22127	AACAGGGTAATTTCAAAAATC	707(BA.1)
	PR2 primary R.	22787	22807	AGTTTGCCCTGGAGCGATTTG	701(BA.2)
	SB13 nested F.	22112	22132	GGTAATTTCAAAAATCTTAGG	700 (BA.1)
	NR4 nested R.	22785	22805	TTTGCCCTGGAGCGATTTGTC	694 (BA.2)
S gene NTD/RBD Junction for Omicron BA.1 only. EF2/NR4 Nested	EF1 primary F.	22184	EPE*ins*	ACGCCTATTATAGTGCGTGAG	630 (BA.1)
	PR2 primary R.	22787	22807	AGTTTGCCCTGGAGCGATTTG	
	EF2 nested F.	22189	-EPE*ins*	TATTATAGTGCGTGAGCCAGA	623 (BA.1)
	NR4 nested R.	22785	22805	TTTGCCCTGGAGCGATTTGTC	

In Table 1, the PCR primers were grouped according to the nested PCR amplicons that they collectively generated. The N gene C04/C03 nested PCR amplicon was used as the sequencing template to verify the presence of a SARS-CoV-2 genomic nucleic acid in a given sample. The S gene RBD S9/S10 and S gene NTD SB12/SB8 nested PCR amplicons were used as the templates for Sanger sequencing to detect the key mutations in the RBD and NTD for the diagnosis of variants (or subvariants), including the BA.2 subvariants. The SB7/SB8 nested PCR primers pair was effective in amplifying the NTD segment for all variants except those of the BA.2 lineage. The S gene NF3/NR4 nested PCR amplicon was used as an alternative to detect RBD mutations in case significant mutations involving the S9 and S10 primer-binding sites caused an RBD RT-PCR amplification failure. The last two sets of primers were designed to amplify the junctional region between the RBD and the NTD of the S gene, using the 214EPE*ins* (GAGCCAGAA*ins*) sequence as a discriminator in the 3' end of the forward primers to selectively amplify the BA.1 sequence in case there was a co-existing BA.1 RBD sequence in a sample containing a dominant Omicron BA.2 subvariant sequence.

Since crude nested PCR products were used for DNA sequencing, the sample used for Sanger reaction might contain multiple templates, including a dominant allele sequence and a number of minor subvariant sequences with multi-allelic SNPs.

3. Results

For diagnostic purpose, the Omicron BA.1 lineage is defined by an S gene mutation profile of A67V, Δ69-70, T95I, G142D, and Δ143-145 in the NTD, and S371L, S373P, S375F, K417N, N440K, G446S, S477N, T478K, E484A, Q493R, G496S, Q498R, N501Y and Y505H in the RBD. The Omicron BA.2 lineage is defined by a mutation profile of Δ24-26, A27S and G142D in the NTD, and S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, S477N, T478K, E484A, Q493R, Q498R, N501Y and Y505H in the RBD [15, 28]; and the BA.4/BA.5 subvariant is characterized by additional L452R and F486V with a wildtype Q493 in the RBD [30]. The NTD SB12/SB8 nested RT-PCR amplicon covers the codons

 doi:10.20944/preprints202206.0192.v1

from Q23 to D178, and the RBD S9/S10 nested RT-PCR amplicon covers the codons from S371 to Y505.

Application of this test procedure for the diagnosis of the Omicron BA.1 variant was previously published [27]. Implementation of the test protocol for the diagnosis of Omicron BA.2 and BA4/BA.5 subvariants and the potential effects of minor subvariant sequences with multi-allelic SNPs on sequencing-based diagnostics are illustrated as follows.

3.1. Sanger sequencing-based diagnostic test for Omicron BA.2 and BA.4/BA.5

When the SARS-CoV-2 nucleic acid samples containing little minor subvariant sequences with multi-allelic SNPs are sequenced for variant determination, the diagnostic test is straightforward. For example, the computer-generated bidirectional Sanger sequencing electropherograms of a 445-bp RT-PCR amplicon of the RBD, and a 500-bp and a 494-bp RT-PCR amplicon of the NTD of the S gene are presented in Figures 1-4 as physical evidence for the detection of a dominant sequence of SARS-CoV-2 Omicron BA.2.13 subvariant and that of Omicron BA.4/BA.5 subvariant.

3.1.1. Forward sequencing of the 445-bp target RBD RT-PCR amplicon for diagnosis of Omicron BA.2 and BA.4/BA.5

Without interfering minor subvariant allele sequences, a one-directional sequencing electropherogram revealed more than 10 key amino acid mutations in the RBD based on which an Omicron BA.2 subvariant or a BA.4/BA.5 subvariant can be diagnosed, as illustrated in Figure 1 A and B.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

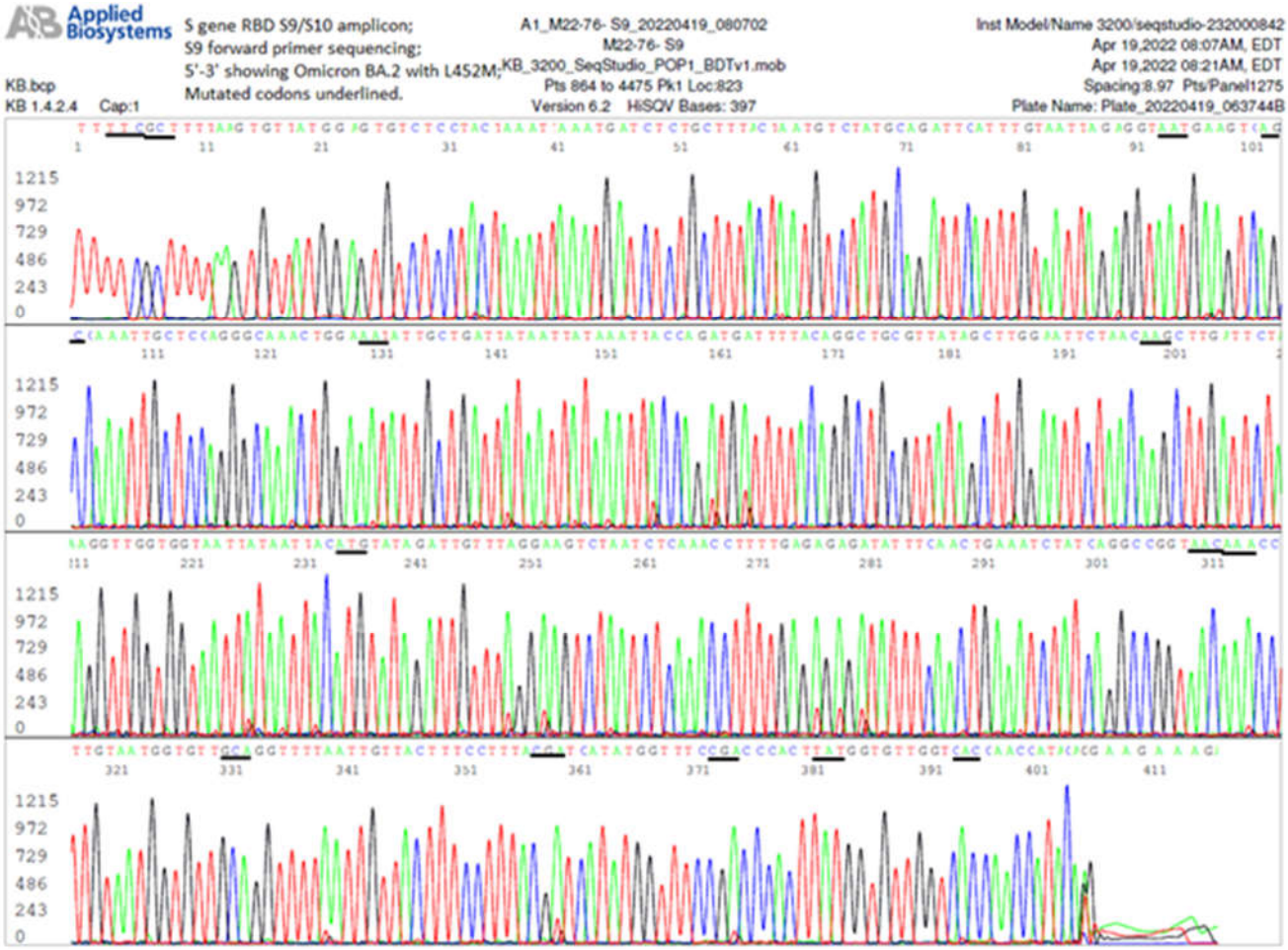


Figure 1A. This is copy of a computer-generated electropherogram showing an S gene RBD sequence generated using S9 forward PCR primer as the sequencing primer. The template was a 445-bp nested RT-PCR product defined by a pair of S9/S10 PCR primers. The underlined codons representing the 14 amino acid mutations in this segment of the S gene, S375F, T376A, D405N, R408S, K417N, N440K, L452M, S477N, T478K, E484A, Q493R, Q498R, N501Y and Y505H, which are characteristic of Omicron BA.2.13.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

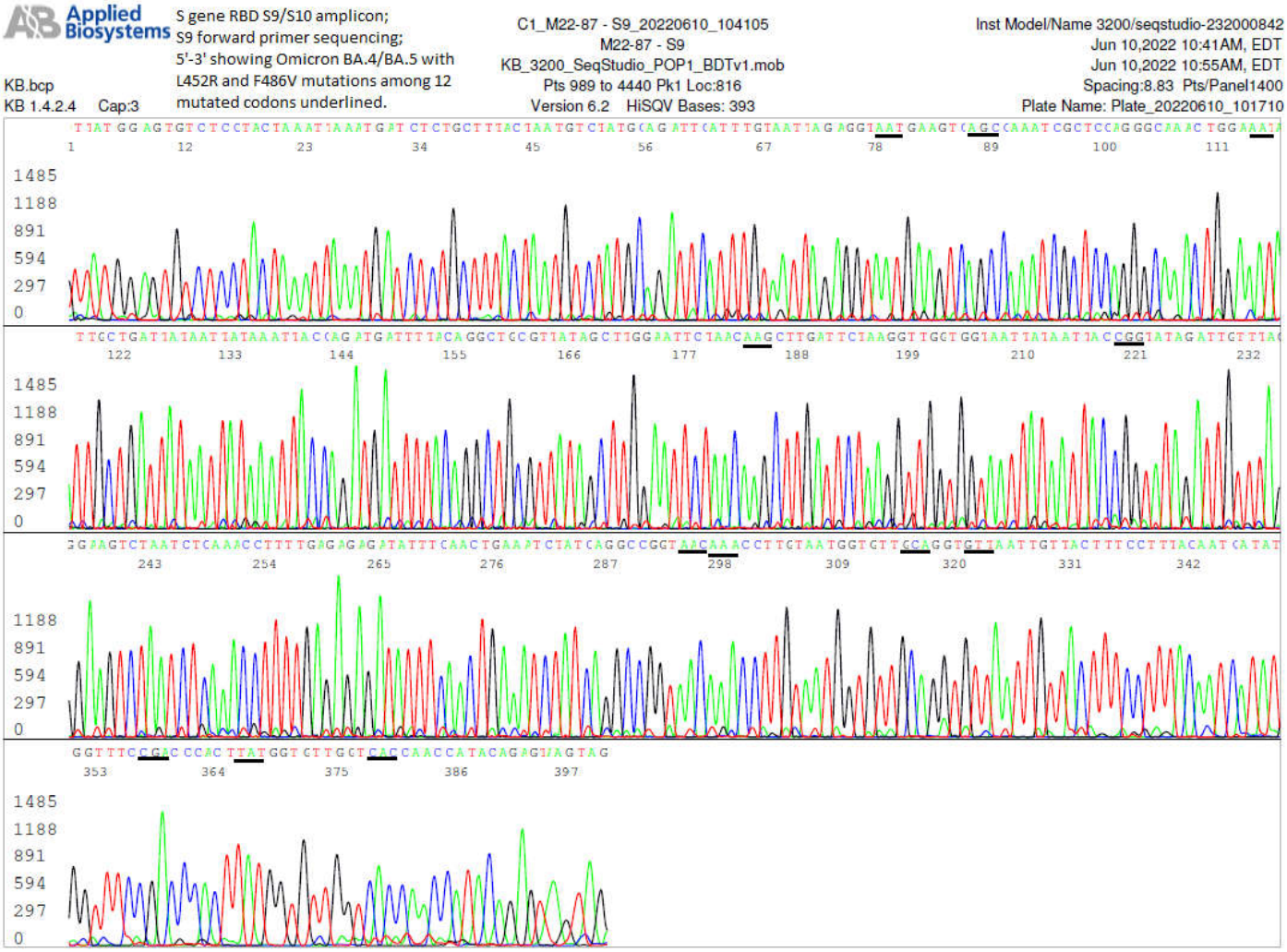


Figure 1B. This is copy of a computer-generated electropherogram showing an S gene RBD sequence generated using S9 forward PCR primer as the sequencing primer. The template was a 445-bp nested RT-PCR product defined by a pair of S9/S10 PCR primers. The underlined codons representing the 12 amino acid mutations in this segment of the S gene, D405N, R408S, K417N, N440K, L452R, S477N, T478K, E484A, F486V, Q498R, N501Y and Y505H, which are characteristic of Omicron BA.4/BA.5.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

3.1.2. Reverse sequencing of the 445-bp target RBD RT-PCR amplicon for diagnosis of Omicron BA.2 and BA.4/BA.5

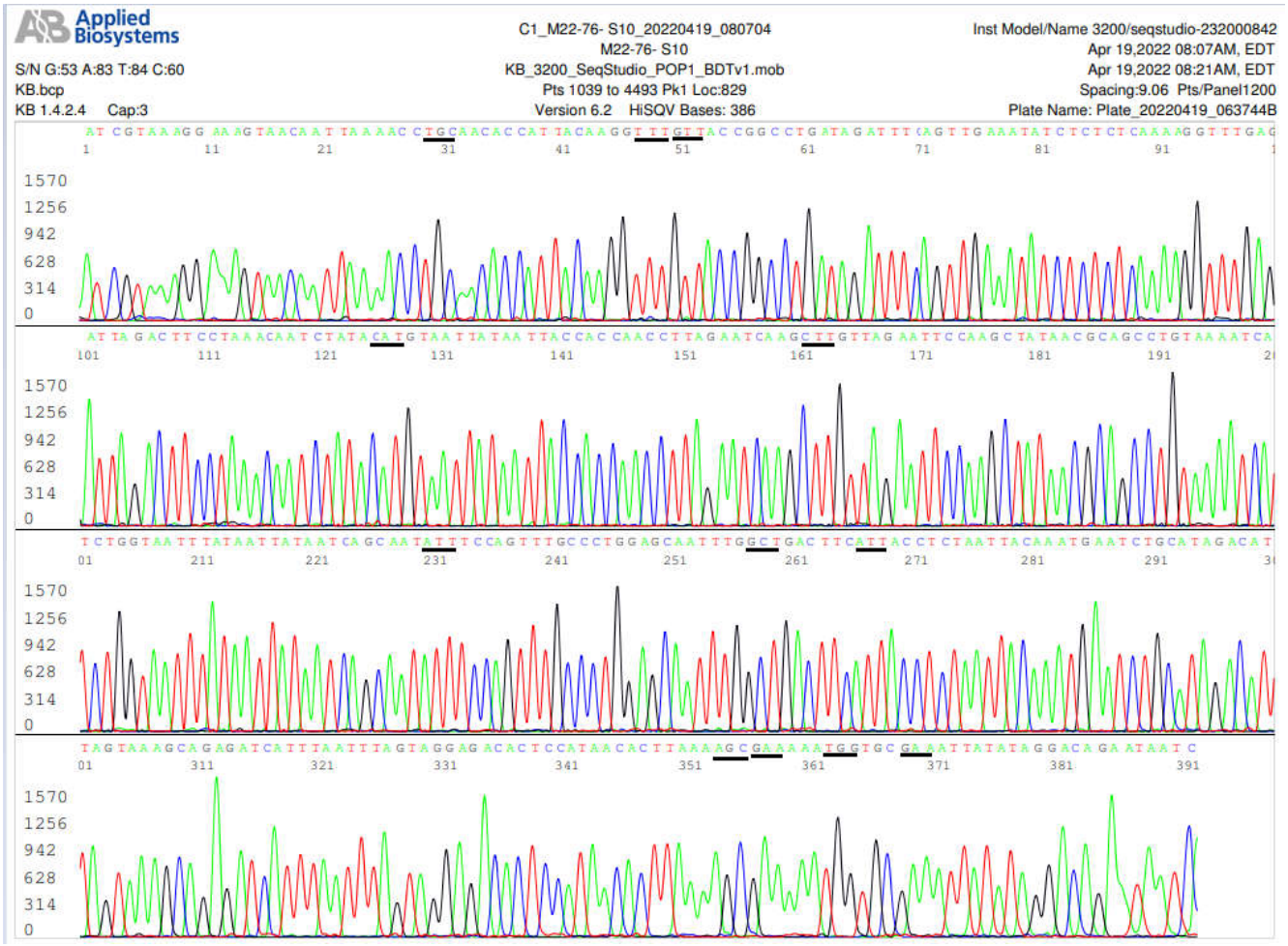


Figure 2A. This is copy of a computer-generated electropherogram showing an S gene RBD sequence generated using S10 reverse PCR primer as the sequencing primer. The template was the same 445-bp nested RT-PCR product used for generating the sequence presented in Figure 1 A. The underlined codons in 3'-5' direction represent the 12 amino acid mutations in this segment of the S gene, E484A, T478K, S477N, L452M, N440K, K417N, R408S, D405N, T376A, S375F, S373P and S371F, which are characteristic of Omicron BA.2.13.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

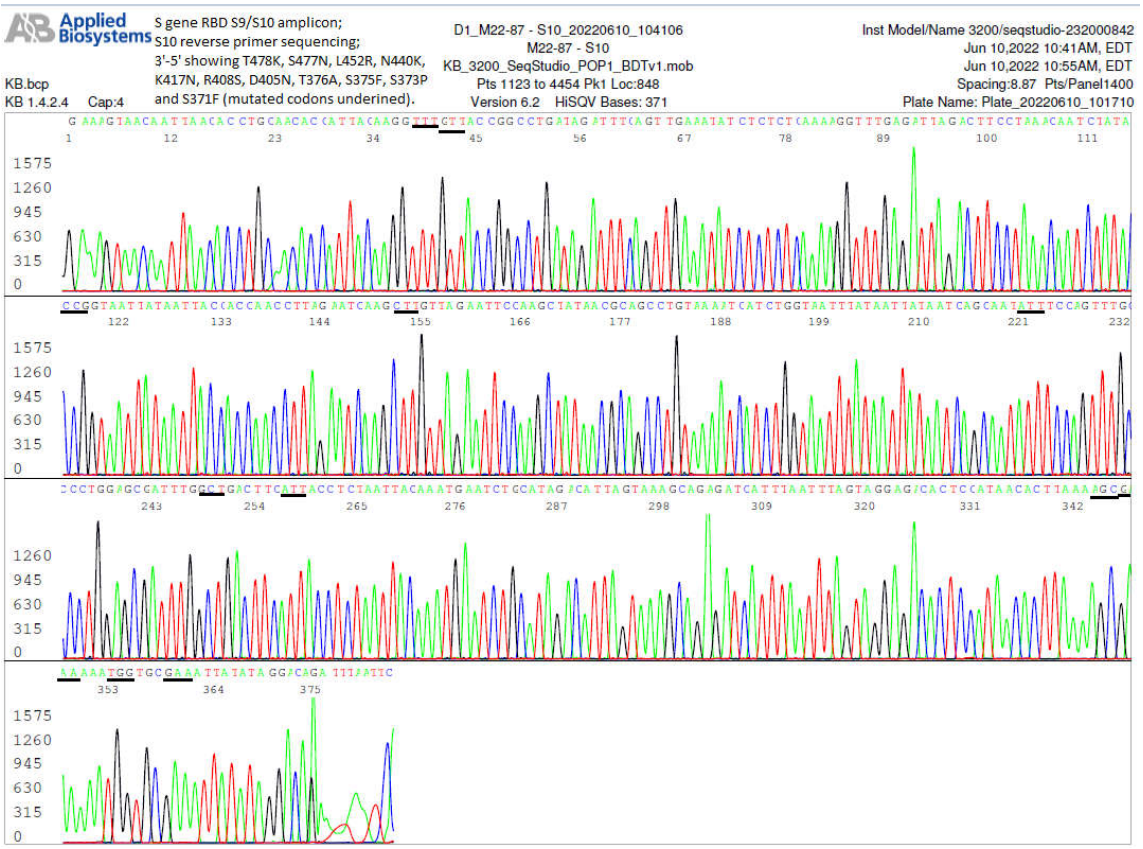


Figure 2B. This is copy of a computer-generated electropherogram showing an S gene RBD sequence generated using S10 reverse PCR primer as the sequencing primer. The template was the same 445-bp nested RT-PCR product used for generating the sequence presented in Figure 1 B. The underlined codons in 3'-5' direction represent the 11 amino acid mutations in this segment of the S gene, T478K, S477N, L452R, N440K, K417N, R408S, D405N, T376A, S375F, S373P and S371F, a mutation profile shared by some Omicron BA.2 subvariants and the Omicron BA.4/BA/5.

3.1.3. Forward sequencing of the target NTD RT-PCR amplicons for Omicron subvariant diagnosis

The S gene NTD is prone to deletions. Although the Omicron BA.2 variant lacks the S gene 69/70 deletion and is associated with the so-called S-gene target failure (SGTF) in some RT-qPCR assays, the BA.2 subvariants do have a 24-26 LPP deletion in the NTD. In order to confirm the presence or absence of these deletions, a set of the general primers is needed to amplify the NTD for Sanger sequencing. This is easily accomplished when there are little interfering minor subvariant sequences with multi-allelic SNPs in the sample, as illustrated in Figures 3 and 4.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

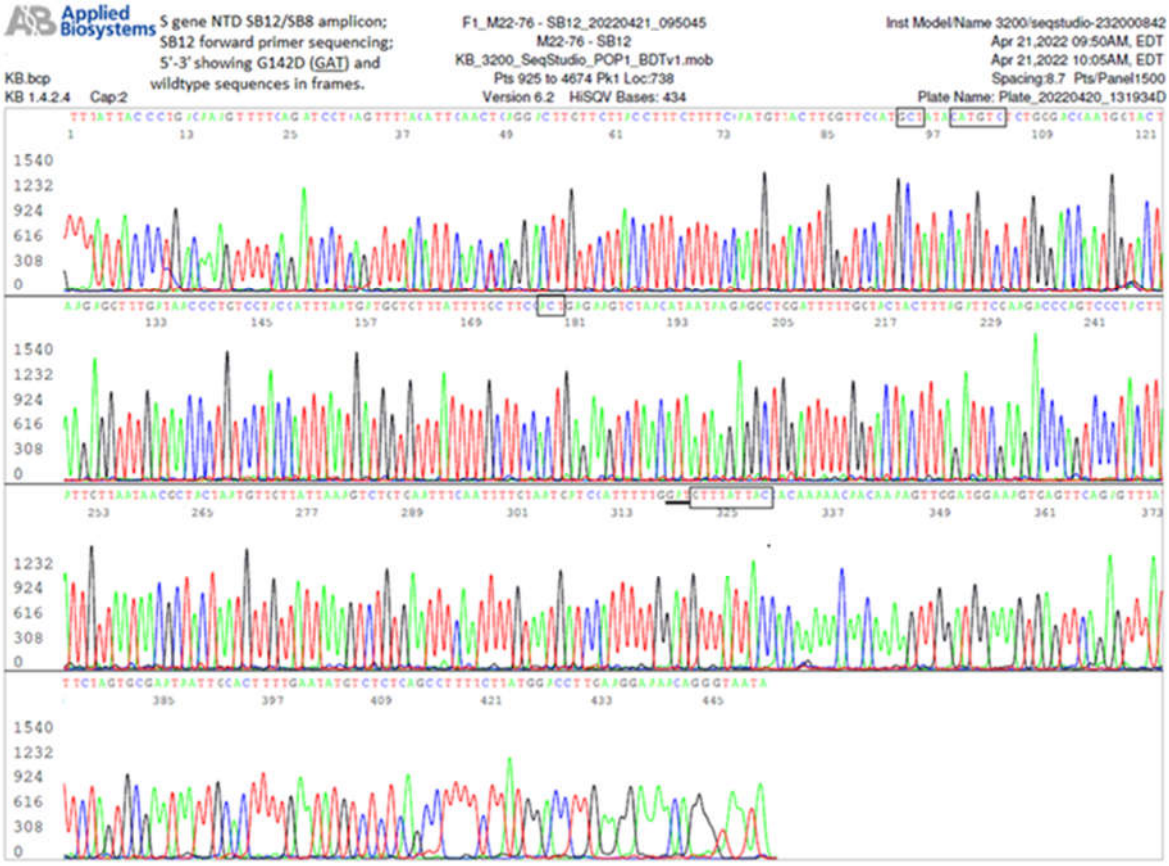


Figure 3A. This is copy of a computer-generated electropherogram showing an S gene NTD sequence generated using SB12 forward PCR primer as the sequencing primer. The template was a 500-bp nested RT-PCR product defined by a pair of SB12/SB8 PCR primers. The single mutated codon G142D in this segment of sequence is underlined. The wildtype codon sequences whose presence distinguishes the Omicron BA.2 lineage from that of Omicron BA.1 are in four rectangular boxes.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

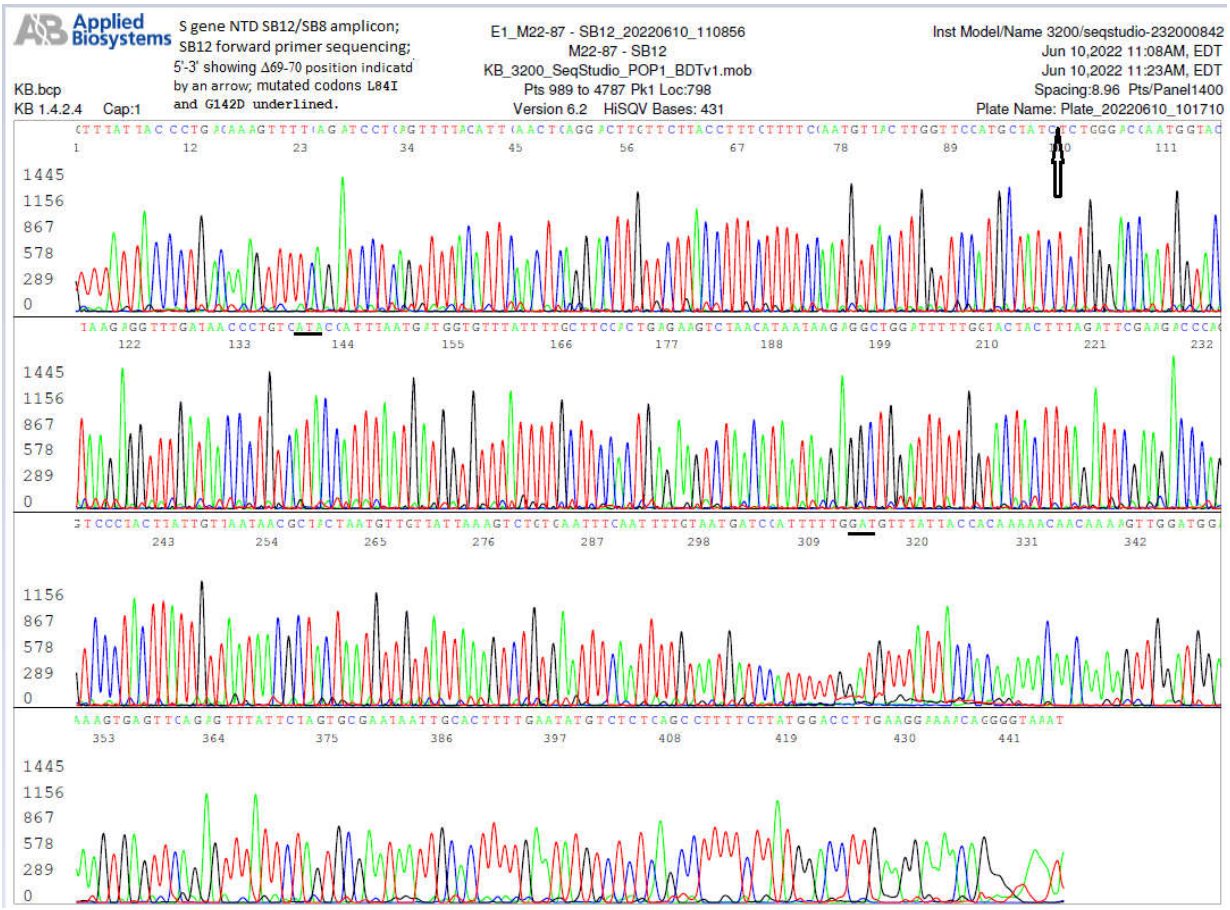


Figure 3B. This is copy of a computer-generated electropherogram showing an S gene NTD sequence generated using SB12 forward PCR primer as the sequencing primer. The template derived from specimen M22-87 was a 494-bp nested RT-PCR product defined by a pair of SB12/SB8 PCR primers. The position of Δ69-70 is pointed by an arrow. The mutated codons of L84I and G142D are underlined. Athough Δ69-70 and G142D are known to be associated with BA.4/BA.5, for example in GenBank Sequence ID: ON691878 among many others, an Omicron BA4./BA.5 with L84I mutation has not been reported in the world literature or annotated in the GenBank database.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

3.1.4. Reverse sequencing of the target NTD RT-PCR amplicons for Omicron subvariant diagnosis

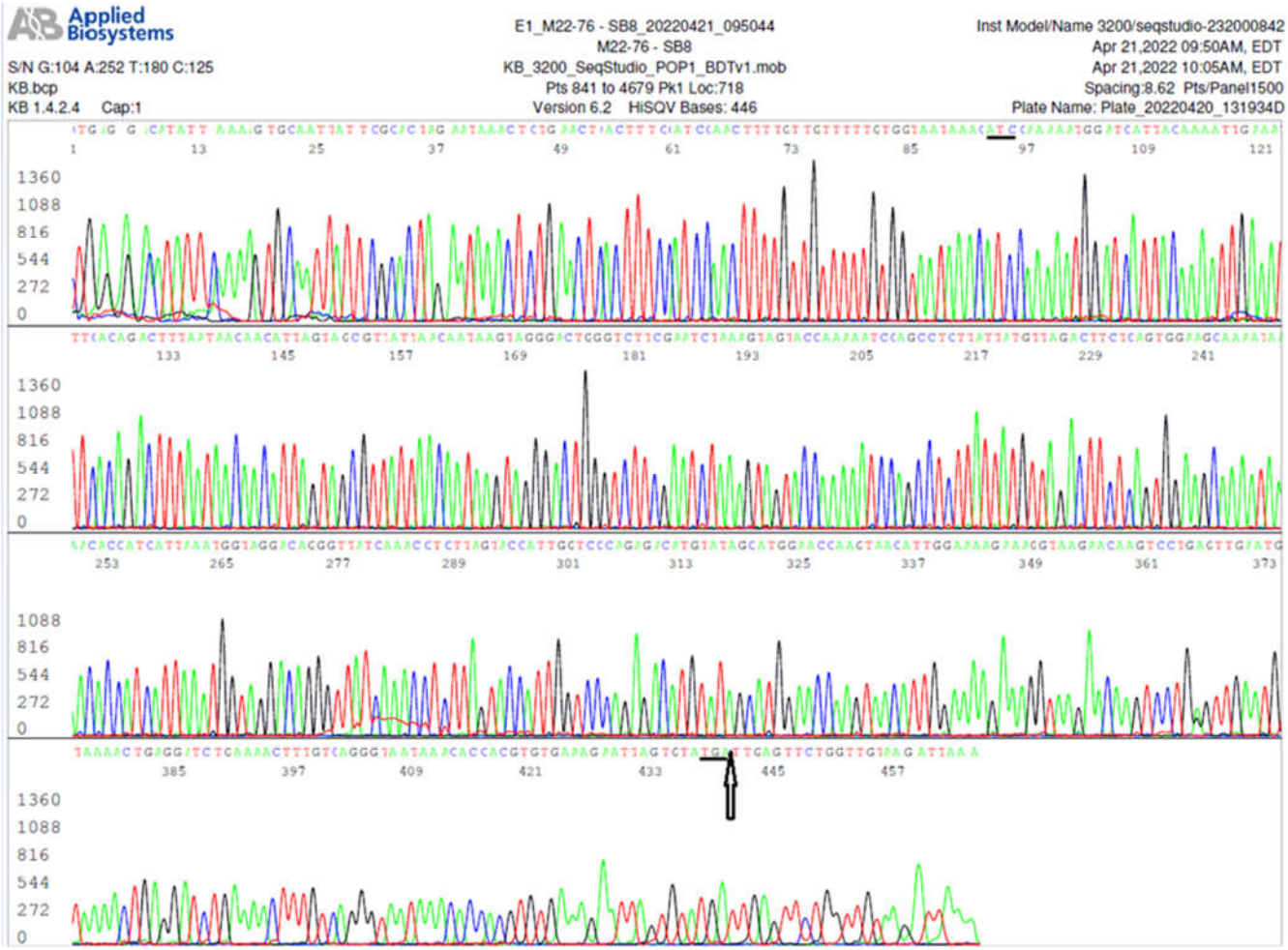


Figure 4A. This is copy of a computer-generated electropherogram showing an S gene NTD sequence generated using SB8 reverse PCR primer as the sequencing primer. The template was the same 500-bp nested RT-PCR product used for generating the sequence presented in Figure 3 A. The two mutated codons, G142D and A27S in 3′-5′ direction, are underlined. The position of Δ24-26 is indicated by an arrow. The NTD sequencing supported the diagnosis of an Omicron BA.2 in this specimen.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

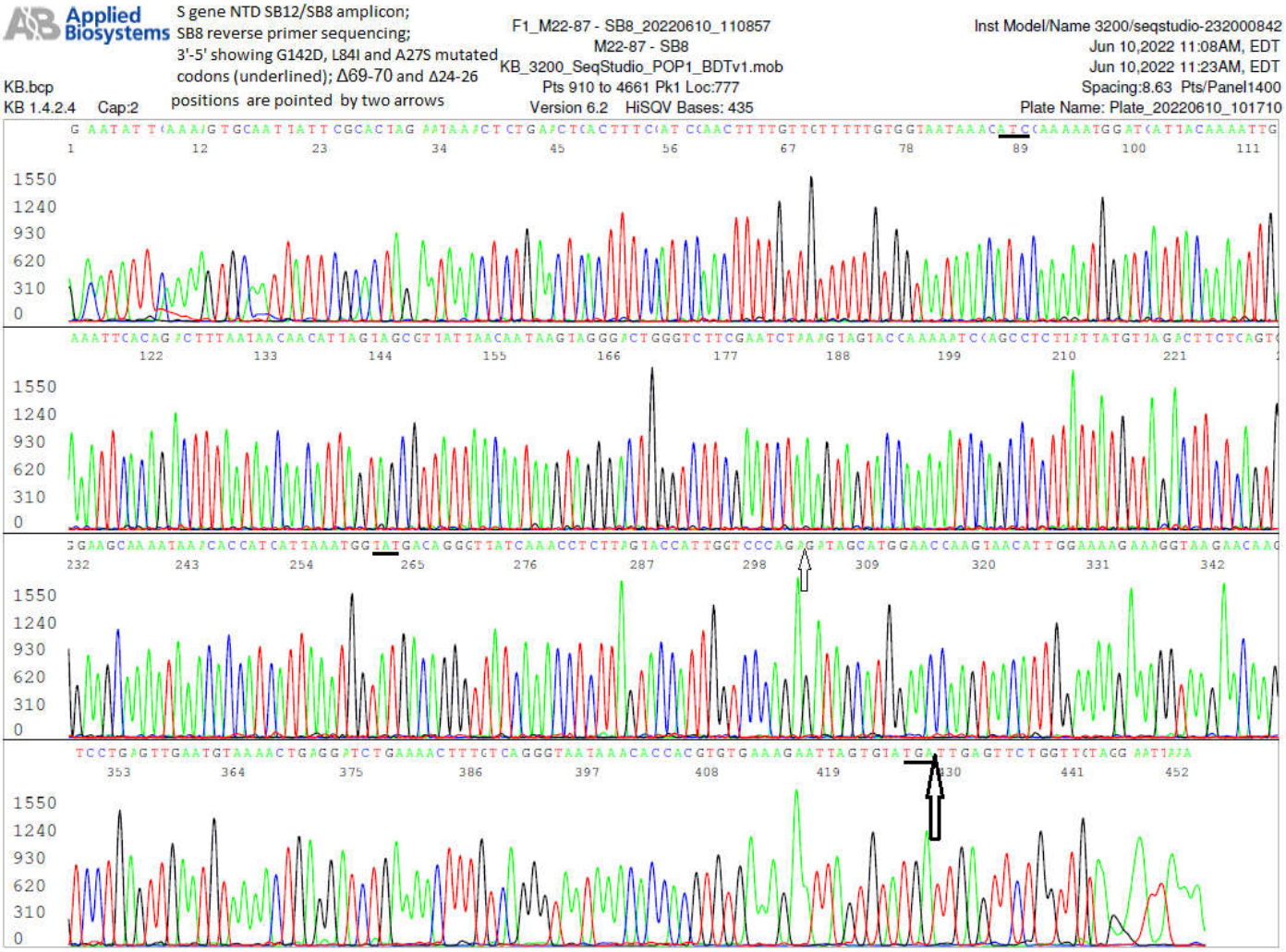


Figure 4B. This is copy of a computer-generated electropherogram showing an S gene NTD sequence generated using SB8 reverse PCR primer as the sequencing primer. The template was the same 494-bp nested RT-PCR product used for generating the sequence presented in Figure 3 B. The position of Δ69-70 is pointed by a small arrow and the position of the Δ24-26 pointed by a big arrow. The mutated codons of G142D, L84I and A27S are underlined. Athough Δ24-26, A27S, Δ69-70 and G142D are known to be associated with BA.4/BA.5, for example in GenBank Sequence ID: ON691878 among many others, an Omicron BA4./BA.5 with L84I mutation has not been reported in the world literature or annotated in the GenBank database.

3.2. Omicron BA.2 in sample containing interfering minor subvariant sequences with multi-allelic SNPs

As demonstrated in Figures 1 and 2, 11-14 key amino mutations in the RBD of the Omicron variant can be detected in one-directional sequencing. A one-directional NTD sequencing can verify the absence of A67V, Δ69-70, T95I, and Δ143-145 mutations (Figure 3 A) or the presence of LPPA24S (Δ24-26, A27S) mutations (Figure 4 A) for further confirmation of a BA.2 sub-lineage. However, some samples positive for an Omicron variant may contain a large number of subvariant sequences with multi-allelic SNPs that may cause uncertain or questionable base calling in Sanger sequencing. One of such examples is illustrated by the sequencing data on a nasopharyngeal swab specimen collected on 25 April 2022 (patient W). The electropherogram of a one-directional forward sequencing of the RBD is presented in Figure 5.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

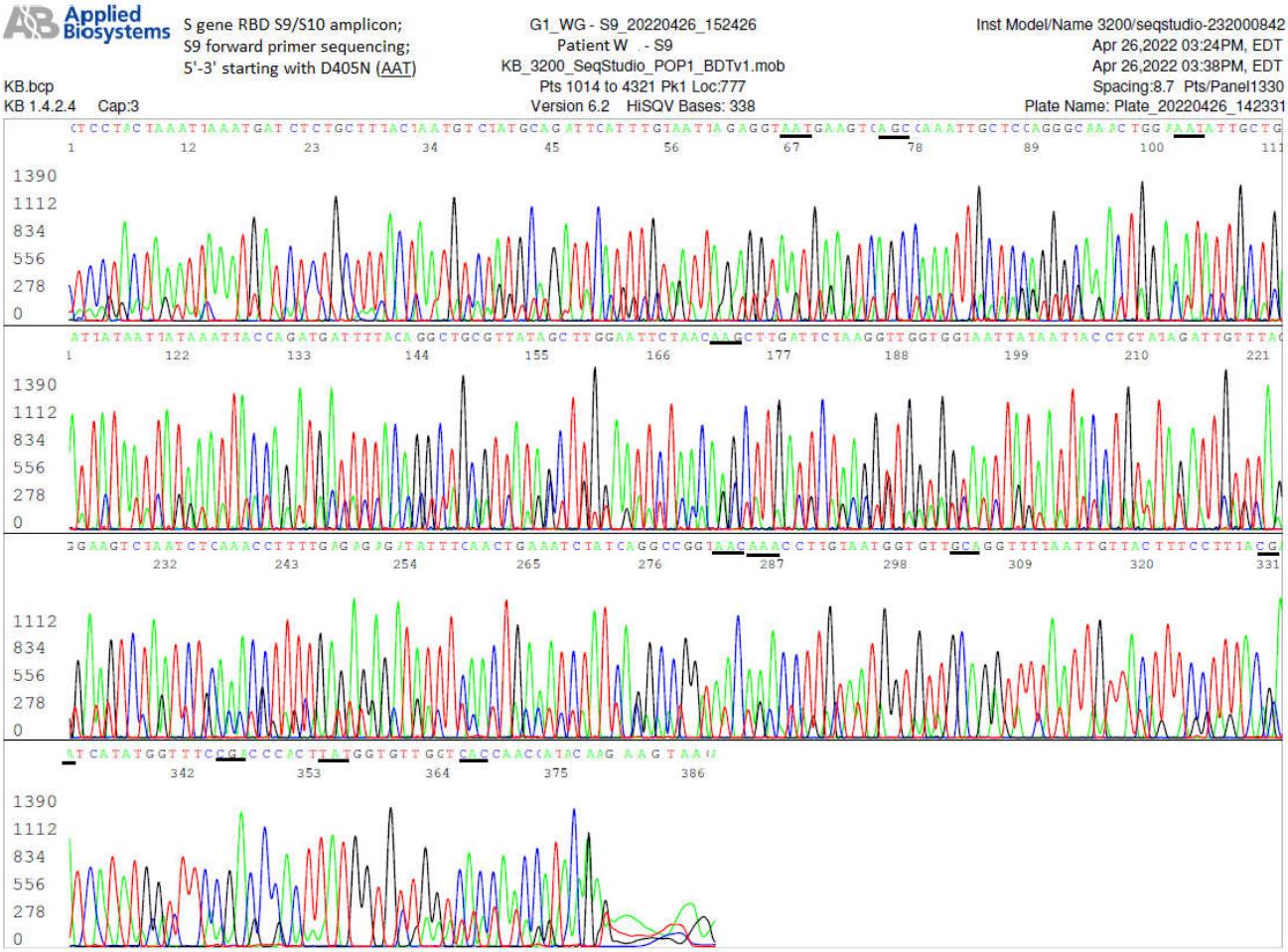


Figure 5. This is copy of a computer-generated electropherogram showing an S gene RBD sequence generated using S9 forward PCR primer as the sequencing primer on Patient W sample. The template was a 445-bp nested RT-PCR product defined by a pair of S9/S10 PCR primers. The dominant sequence shows 11 underlined mutated codons of D405N, R408S, K417N, N440K, S477N, T478K, E484A, Q493R, Q498R, N501Y and Y505H, characteristic of Omicron BA.2. However, there are numerous multicolored low peaks below the high peaks of the dominant sequence, which the computer relies on for base calling.

In order to prove that the multicolored low peaks in the electropherogram presented in Figure 5 were reproducible in their specific positions of the sequence and not random sequencing noise artefacts, aliquots of the nucleic acid extract used to generate the electropherogram of Figure 5 were reamplified by 3 sets of nested RT-PCR for re-sequencing the N gene, the RBD and the NTD of the S gene in one single run to reduce possible sequencing artefacts introduced by between-run technical and reagent variations. The electropherograms of the bidirectional Sanger sequencing of the 3 target gene segments are presented in Figures 6-11 for comparison.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

3.2.1. Minor subvariant sequences with multiallelic SNPs in S gene RBD were reproducible

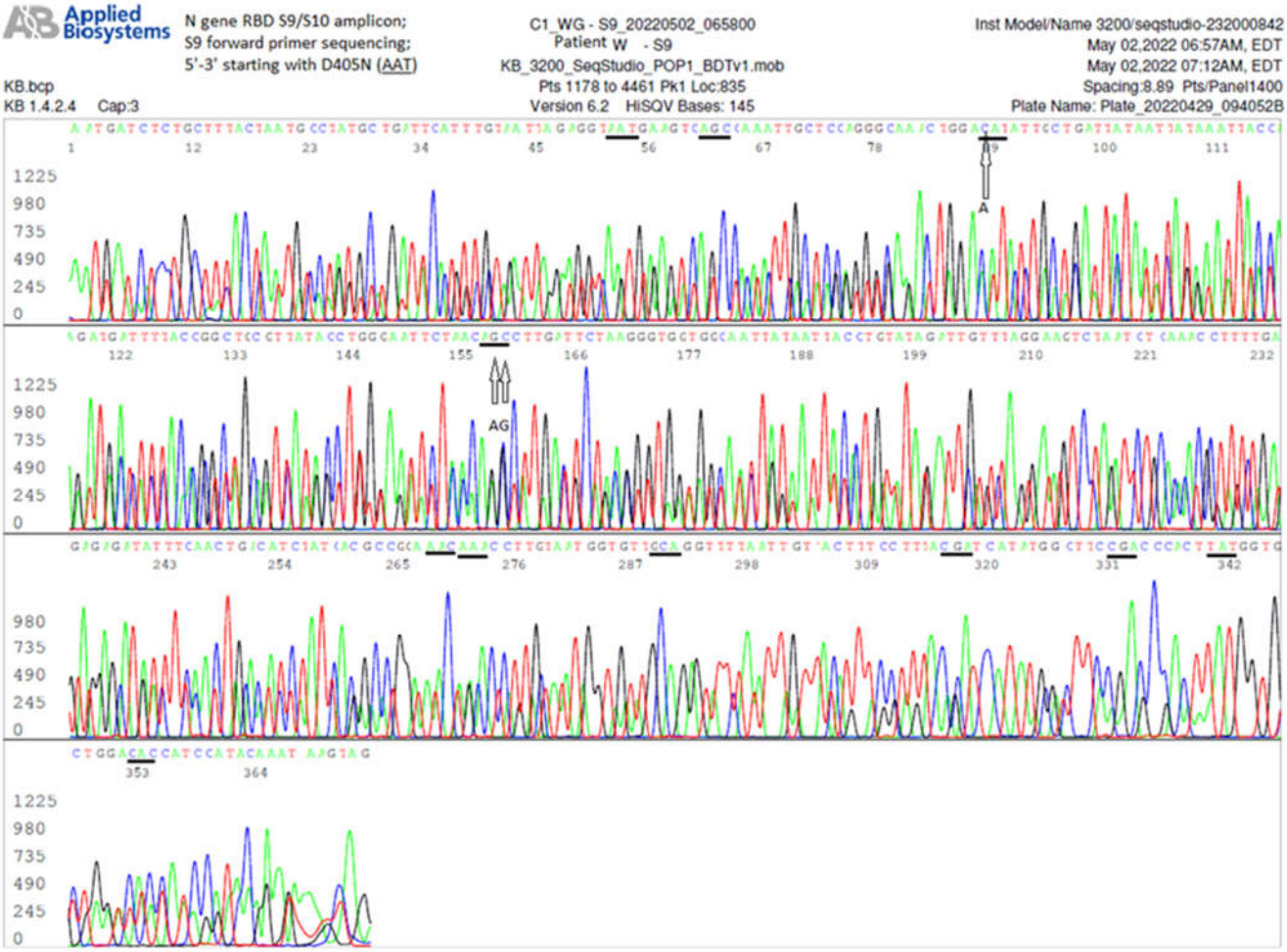


Figure 6. This electropherogram of a repeated forward sequencing of the RBD in the Patient W sample shows that the positions of the multicolored low peaks illustrated in Figure 5 were reproduced and the repeated sequence confirmed the 11 mutated codons shown in Figure 5. However, some multicolored low peaks shown in Figure 5 have now increased in height to become the dominant peaks (indicated by arrows), affecting the computer's base-calling accuracy. The letters, A, G and A under the 3 arrows represent the correct bases whose sequencing peaks were overshadowed by the allelic base peaks. The fact that the sequence of the multicolored low peaks observed in Figure 5 was reproducible in repeated sequencing (in Figure 6) confirms that these secondary low peaks are not technical artefacts but represent true minor subvariant sequences with multi-allelic SNPs, as observed and reported by others using WGS [9-11].

3.2.2. Only the dominant RBD sequence was analyzed in Sanger sequencing

In the sequencing procedure, about 0.2μL of unpurified nested PCR products, often after further dilution in water by the operator depending on the fluorescence density observed in gel electrophoresis, was used as the sample material to initiate a Sanger reaction. Therefore, the nested PCR products of minor subvariant sequences with multi-allelic SNPs transferred into the Sanger reaction mixture were greatly reduced. In addition, subvariant sequences with multi-allelic SNPs may not have a fully matched primer-binding site for the sequencing primer. Under such conditions, the dominant sequence equipped with a primer-binding site fully matching the sequence of the sequencing primer is the preferred template in the repeated enzymatic primer extension/termination cycles during Sanger reaction. As the result, one of the two bidirectional sequencing electropherograms may show better base-calling data, which actually represent the dominant sequence in a mixture of diverse PCR products. One of such examples is shown in Figure 7, representing

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

a sequence reverse-complementary to that shown in Figure 6 by excluding the interfering minor subvariant sequences.

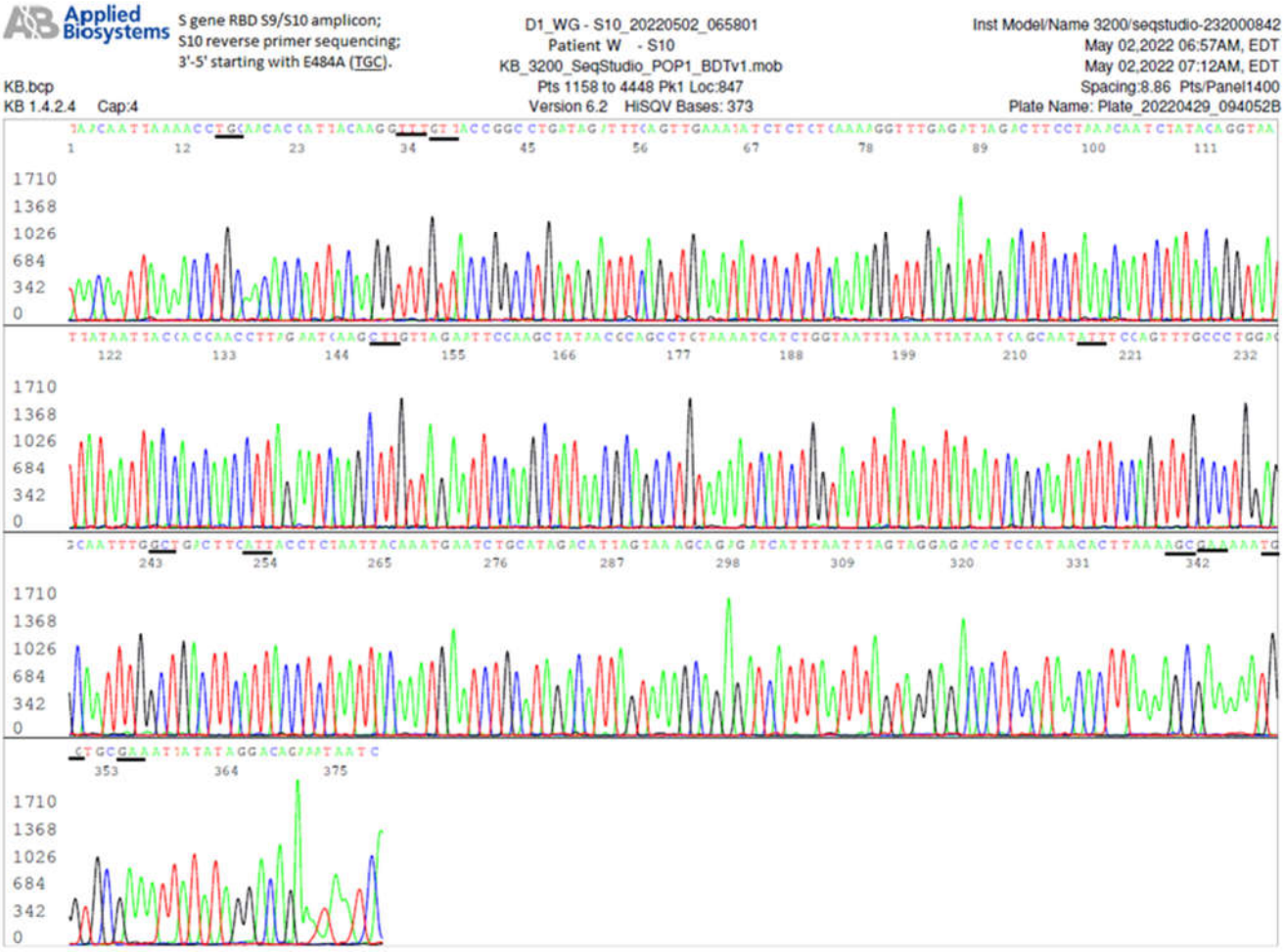


Figure 7. This is an electropherogram of the S10 reverse primer sequencing of the same nested RT-PCR products used to generate the sequence of Figure 6, showing 11 underlined mutated codons E484A, T478K, S477N, N440K, K417N, R408S, D405N, T376A, S375F, S373P and S371F in one dominant sequence. There were no interfering minor variant sequences with multi-allelic SNPs as shown in Figure 6 although the same nested PCR products were used as the template(s) for the bidirectional sequencing in one single run.

3.2.3. Minor subvariant sequences with multi-allelic SNPs in S gene NTD are common findings

Minor subvariant sequences with multi-allelic SNPs may be co-amplified with the dominant target NTD sequence of the S gene, as demonstrated in Figure 8.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

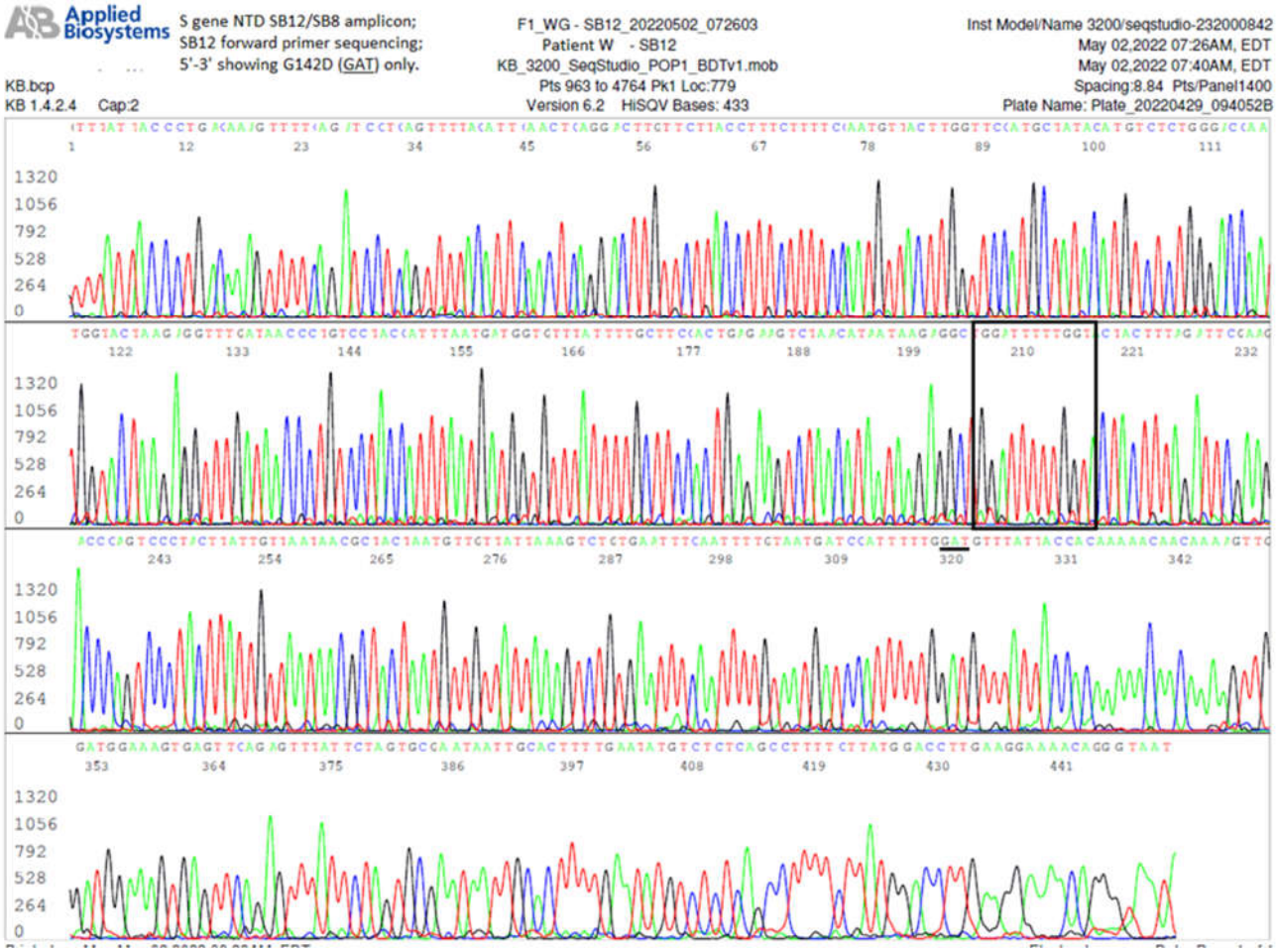


Figure 8. This is copy of a computer-generated electropherogram showing an S gene NTD sequence generated using SB12 forward PCR primer as the sequencing primer on the Patient W sample. Similar to the sequence illustrated in Figure 3, this sequence has one solitary G142D mutation (codon underlined), indicative of an Omicron BA.2 lineage. However, there are stretches of colored low peaks in the sequence. These colored low peaks may be part of the nonspecific background sequencing noise artefacts resulting from uncontrollable technical or reagent deviations, or may indicate the presence of subvariant sequences with multi-allelic SNPs. If the sequence of these colored low peaks can be reproduced by bidirectional sequencing of the same PCR products in a reverse-complementary direction, they must be part of a subvariant sequence with multi-allelic SNPs. To evaluate the latter hypothesis, a short sequence with 10 colored low peaks is framed in a rectangular box for comparison with the reverse sequence in Figure 9 below.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

3.2.4. Reverse primer sequencing confirmed minor subvariant sequences with multi-allelic SNPs in the S gene NTD

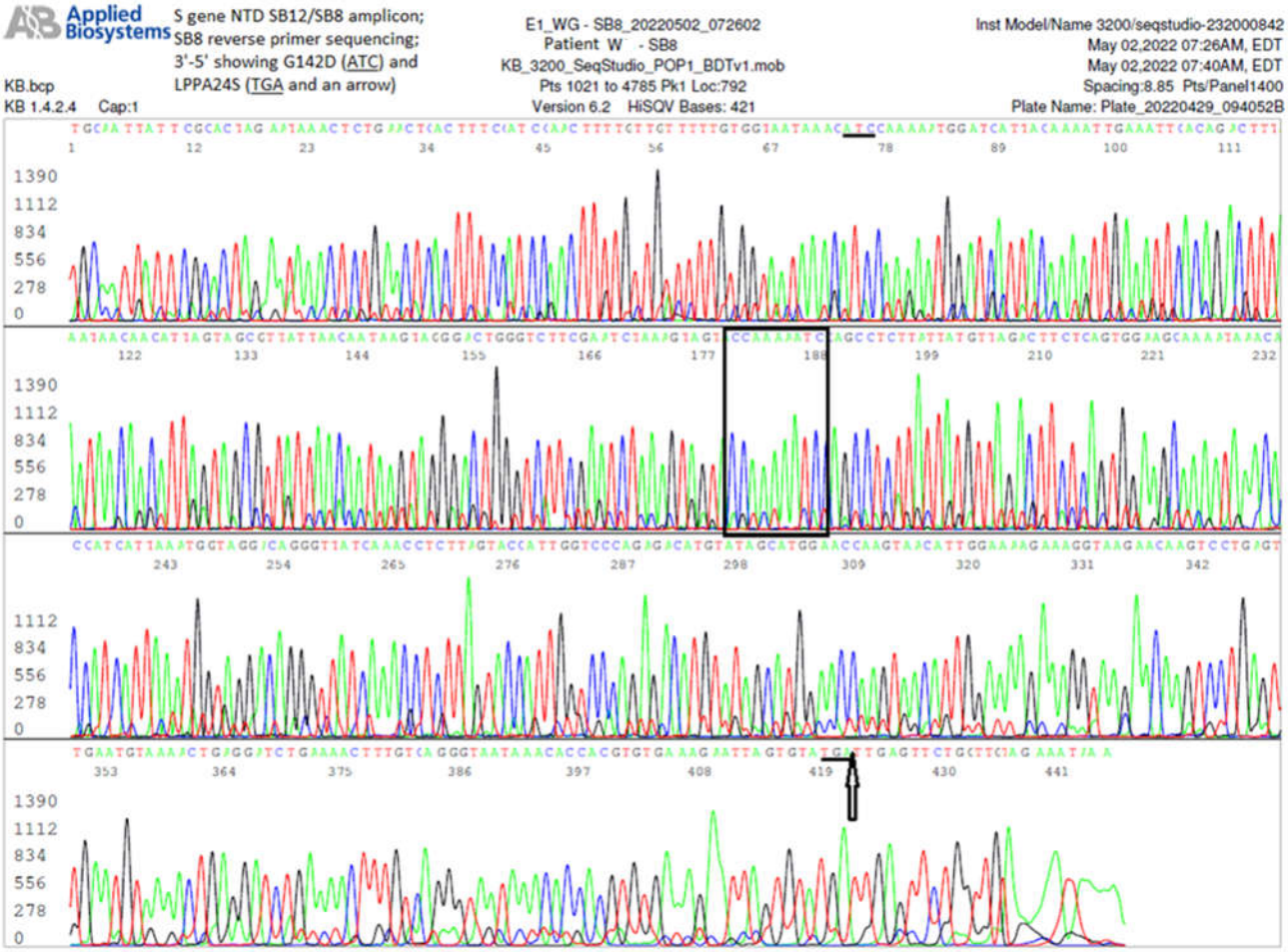
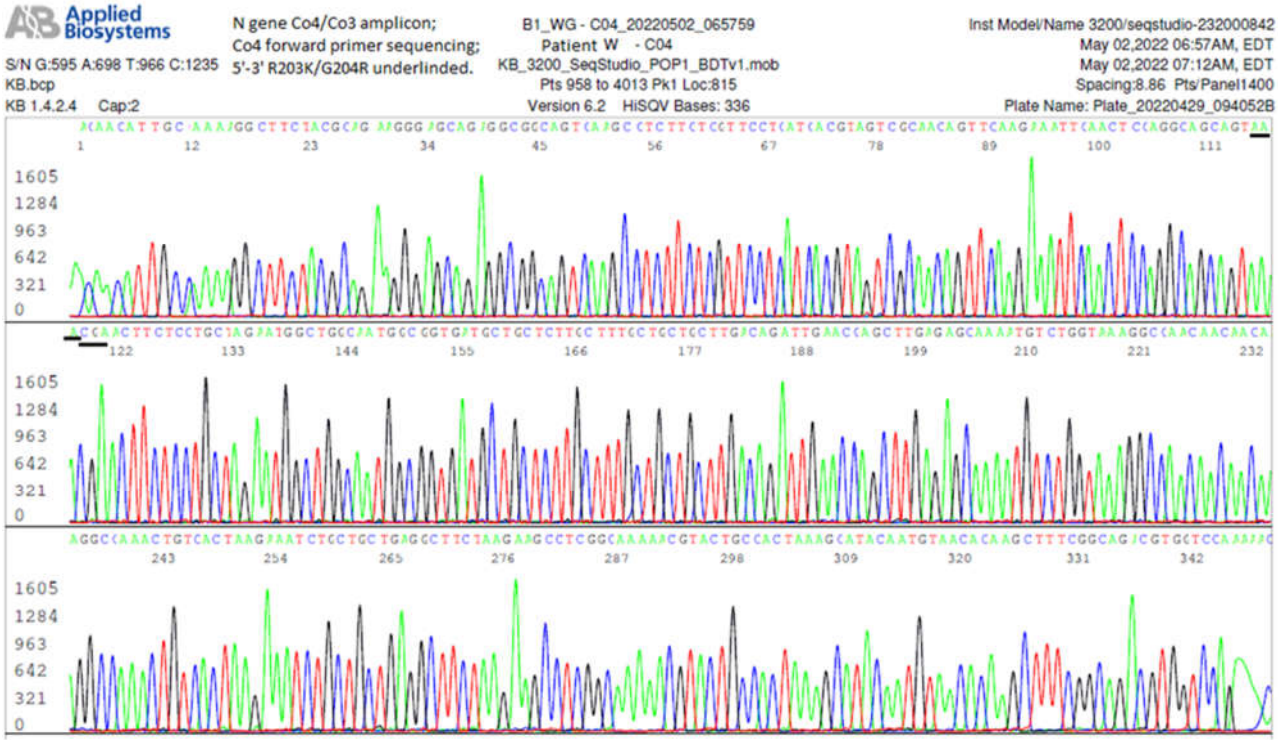


Figure 9. This is copy of a computer-generated electropherogram showing an S gene NTD sequence generated using SB8 reverse PCR primer as the sequencing primer. The template was the same nested PCR products used to generate the sequence in Figure 8. The two mutated codons, G142D and A27S in 3'-5' direction, are underlined, and the position of Δ24-26 is indicated by an arrow. In addition, the sequence of the colored low peaks in the rectangular box reads 3'-AGTCTAACAT-5', which is the reverse complement of the colored low peak sequence of 5'- ATGTTAGACT-3' in the rectangular box in Figure 8. Similar reverse-complementary colored low peak sequences can be found on both sides of the rectangular boxes. Since these low-peak sequences are reproducible in bidirectional sequencing, they are considered to be true PCR products of minor Omicron subvariant sequences with multi-allelic SNPs. However, it is not clear if these colored low peaks indicate multiple mutated nucleotides in one subvariant sequence or numerous subvariant sequences, each with a unique single-nucleotide mutation.

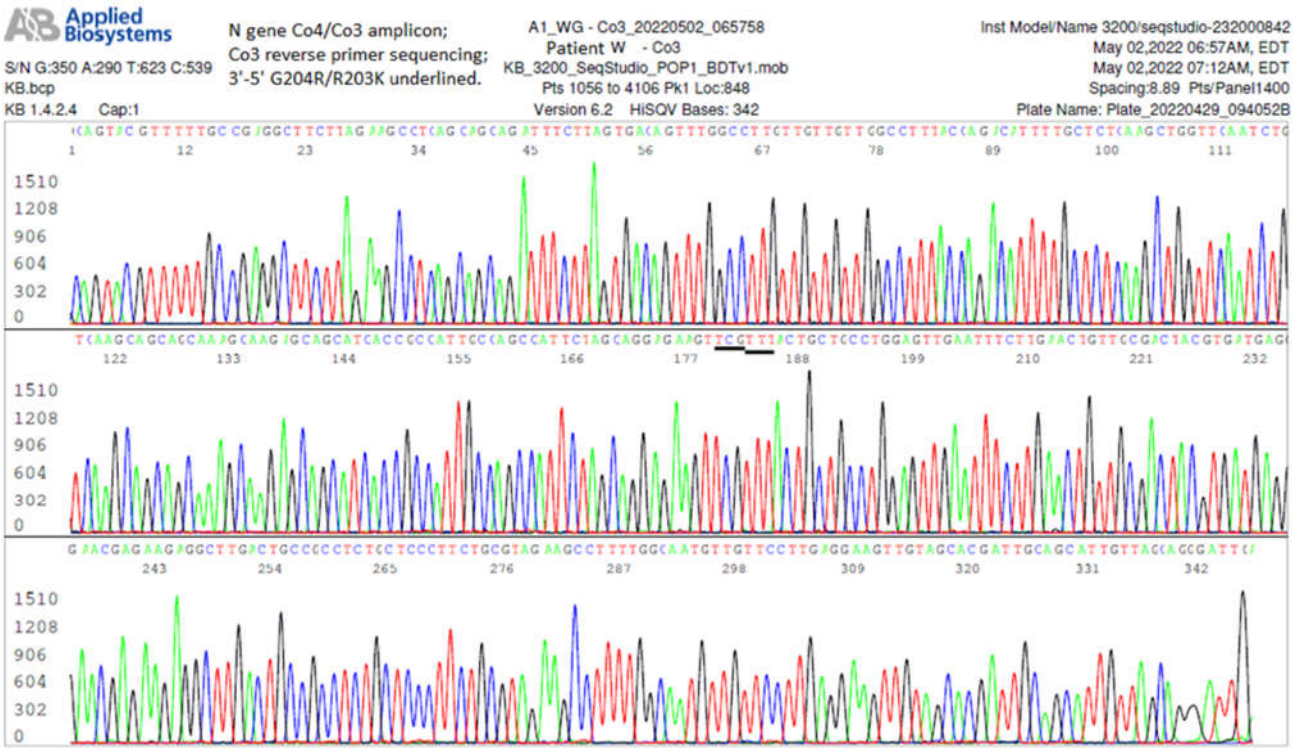
3.2.5. Uneven distribution of multi-allelic SNPs in Omicron variants

As previously reported, the 398-base N gene target selected for molecular diagnosis of SARS-CoV-2 is more conserved than the S gene RBD and NTD in the circulating Omicron BA.1 variant [27]. This finding is also confirmed in sequencing Omicron BA.2 subvariants and demonstrated in Figure 10 A and B.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1



(A)



(B)

Figure 10. A and B. These two bidirectional sequencing electropherograms show the N gene R203K and G204R mutations (codons underlined), which are present in all Omicron variants, using Co4 primer as the forward sequencing primer (A) and Co3 primer as the reverse sequencing primer (B) on the Patient W sample. Compared to the sequences of the S gene presented in Figures 6-9, these two bidirectional sequences derived from the same sample and generated in the same run display no secondary colored peaks to suggest intra-host N gene diversity.

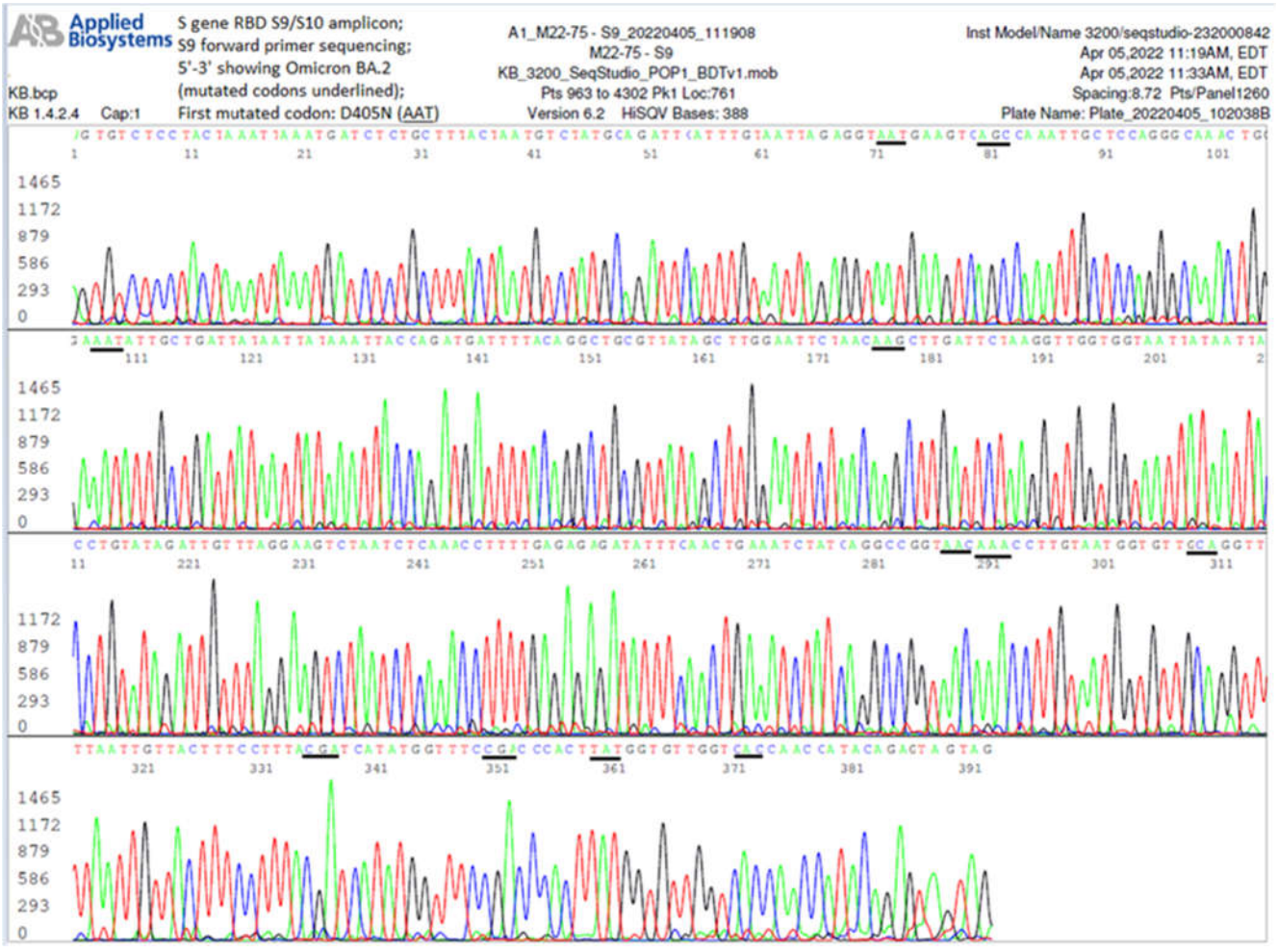
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

3.3. Omicron BA.2 variant containing both BA.1 NTD and BA.2 NTD

In a nasopharyngeal swab specimen collected on 3 April 2022 from a symptomatic adult patient in Connecticut, U.S.A., identified as sample M22-75, bidirectional sequencing of the N gene nested RT-PCR products demonstrated a 398-base SARS-CoV-2 N gene sequence with R203K and G204R mutations. However, bidirectional Sanger sequencing of the S gene RBD and the NTD nested RT-PCR products revealed that the sample actually contained an Omicron BA.2 variant with two different NTD sequences. One of them belongs to the BA.2 sub-lineage and the other to the BA.1 sub-lineage. Since an Omicron BA.2 subvariant containing a co-existent BA.1 S gene NTD mutation profile has not been reported in the world literature, the relevant sequencing findings are presented as follows.

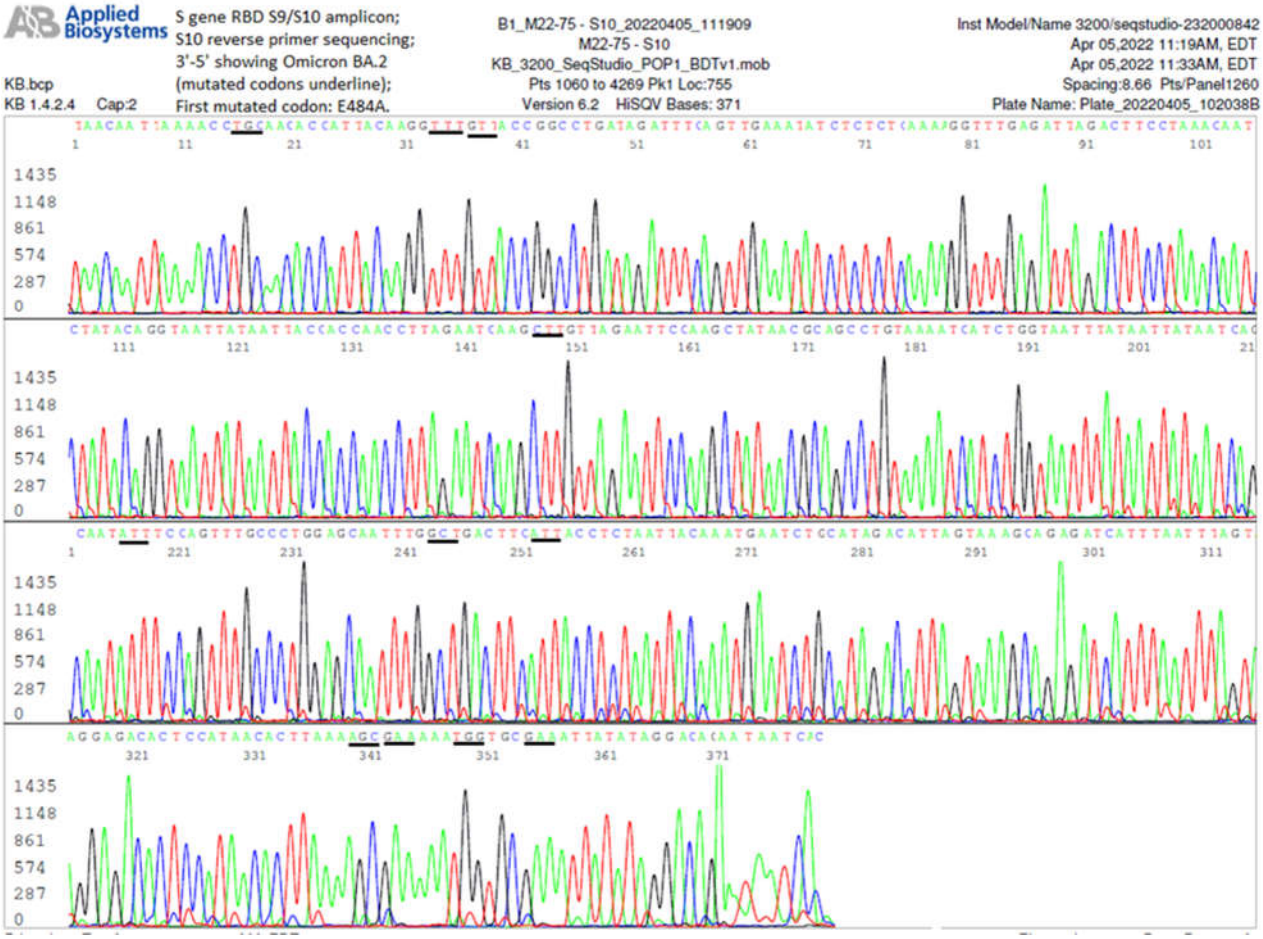
3.3.1. Verifying the RBD mutations of Omicron BA.2

Bidirectional sequencing of a 445-bp RBD nested RT-PCR amplicon confirmed that the dominant sequence of the SARS-CoV-2 in sample M22-75 was that of the Omicron BA.2 subvariant (Figure 11).



(A)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1



(B)

Figure 11. A and B. These two electropherograms show the bidirectional Sanger sequencing results of the 445-bp RBD nested RT-PCR products on sample M22-75. The 11 mutated codons, D405N, R408S, K417N, N440K, S477N, T478K, E484A, Q493R, Q498R, N501Y and Y505H in 5′-3′ direction, are underlined in (A); the 11 mutated codons, E484A, T478K, S477N, N440K, K417N, R408S, D405N, T376A, S375F, S373P and S371F in 3′-5′ direction, are also underlined in (B). In combination, these 15 amino acid mutations in the RBD support the diagnosis of a SARS-CoV-2 Omicron BA.2 subvariant in this sample.

3.3.2. Atypical Omicron BA.1 NTD in sample containing Omicron BA.2 subvariant

Sample M22-75 was the first Omicron BA.2 subvariant encountered in this laboratory. For the diagnosis of SARS-CoV-2 variant Omicron BA.1, which was most prevalent in January 2022 in the U.S., the nested PCR amplicon generated by a pair of SB7/SB8 primers was routinely used at the time as the template for Sanger sequencing for the detection of A67V, Δ69-70, T95I, G142D, and Δ143-145 mutations in the NTD of the S gene [27]. When this pair of primers was used to perform nested RT-PCR on sample M22-75 and as the sequencing primers, two complementary paired electropherograms were generated, which are presented in Figures 12 and 13. Submission of the 474-base 5′-3′ composite sequence derived from these two sequences to the GenBank for BLAST analysis induced a highly unusual report, which was copied and pasted as Figure 14.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

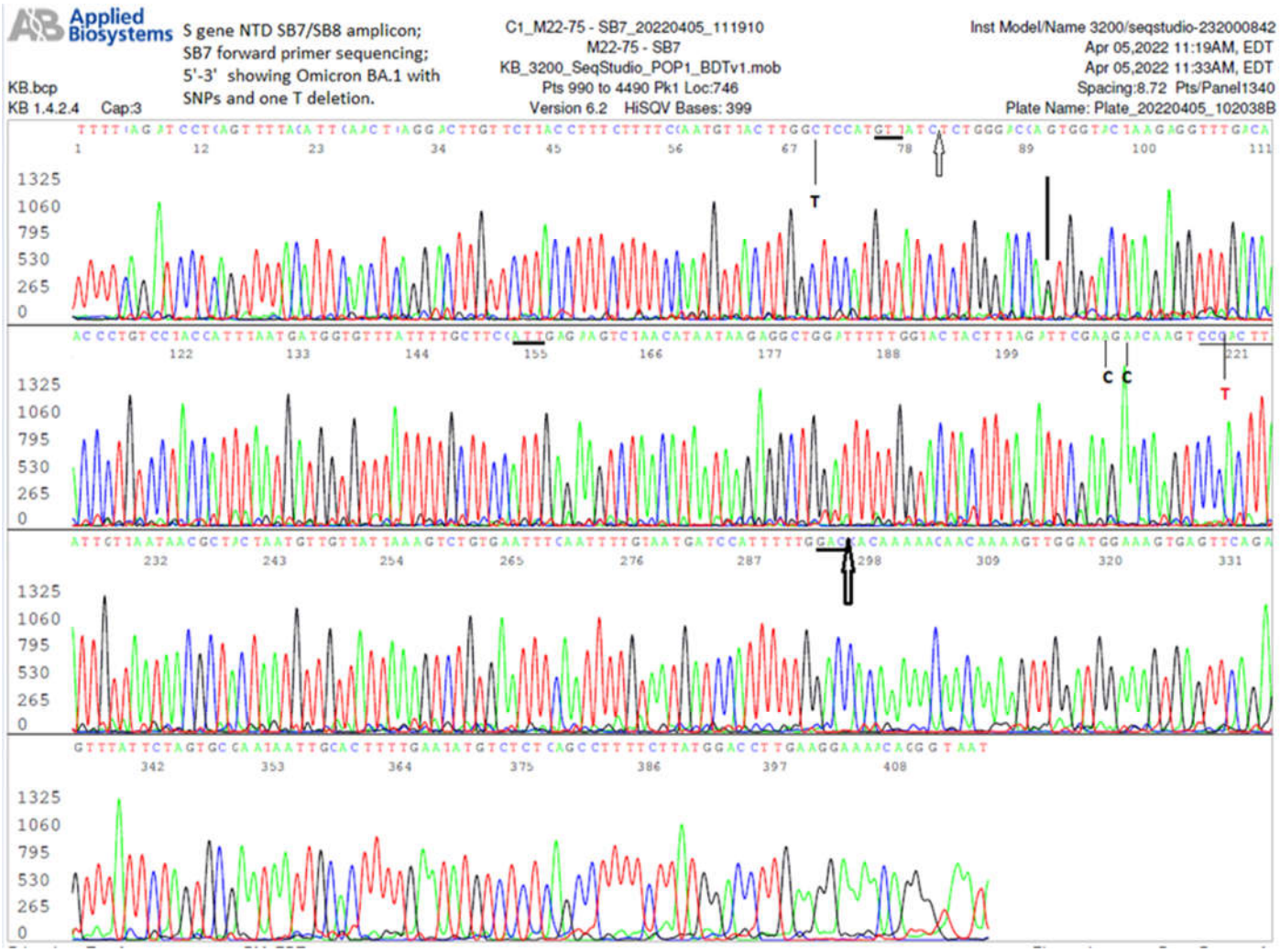


Figure 12. This is copy of an electropherogram showing the dominant S gene NTD sequence with A67V, Δ69-70, T95I, G142D and Δ143-145 mutations commonly used to define an Omicron BA.1 subvariant in sample M22-75. The template was a 474-bp nested PCR amplicon defined by the SB7/SB8 primer pair. The sequence was generated using SB7 PCR primer as the forward sequencing primer. The mutated codons are underlined and the positions of Δ69-70 and Δ143-145 are indicated by a small arrow and a big arrow, respectively. The black letters, T, C and C, each under a thin vertical line, indicate 1 "T>C" and 2 "C>A" base mutations when compared with a GenBank reference sequence (Figure 14). There is also a "T" base deletion at position 220 in this sequence tracing, indicated by a thin line above a letter T typed in red color. In addition, at position 92 indicated by a heavy vertical line there are two superimposing competing peaks, composed of a green "A" peak and a black "G" peak, which the computer chose base "G" to call as the dominant nucleotide at this position. Converting the wildtype base "A" to "G" at this position changes the amino acid codon of asparagine to that of arginine, creating a new nonsynonymous N74R mutation. This very Sanger sequencing electropherogram indicates that there are at least two competing unclassified Omicron BA.1 S gene NTD subvariant allele sequences in this sample although the RBD is that of a BA.2 sublineage.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

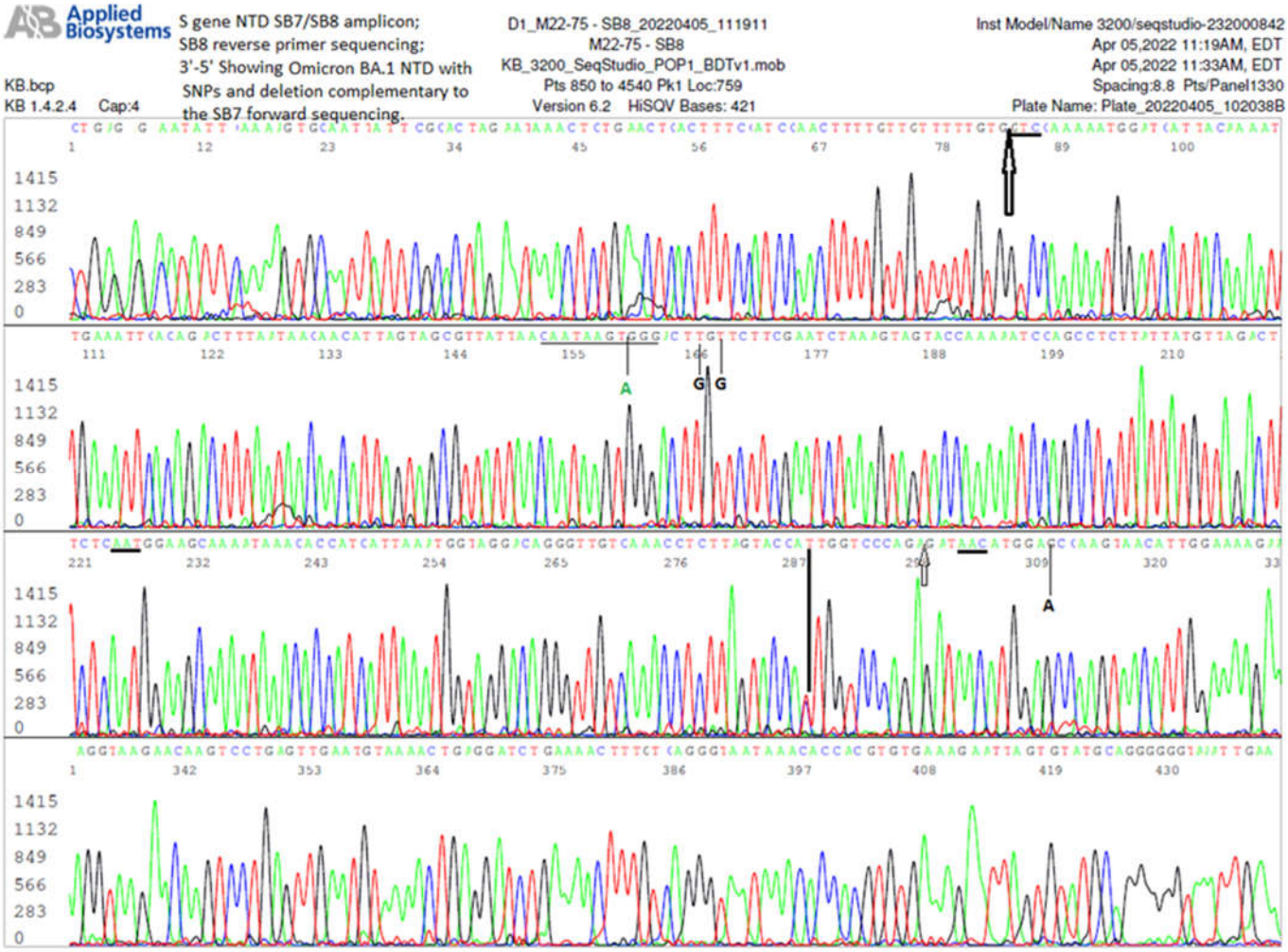


Figure 13. This is an electropherogram showing a 3′-5′ sequence derived from same PCR products used to generate the sequence presented in Figure 12, using SB8 PCR primer as the reverse sequencing primer. This reverse complementary sequence confirms that all nucleotide mutations and deletions shown in Figure 12 are valid findings. The ambiguous A/G double peak read as a mutated base "G" at position 92 in Figure 12 is now read as base "T" by the computer in Figure 13, indicating the presence of two competing allele sequences, both being used as the templates for generating a segment of dominant Omicron BA.1 S gene NTD sequence, which has not been recorded in the SARS-CoV-2 genomes annotated in the GenBank database, as shown in Figure 14.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

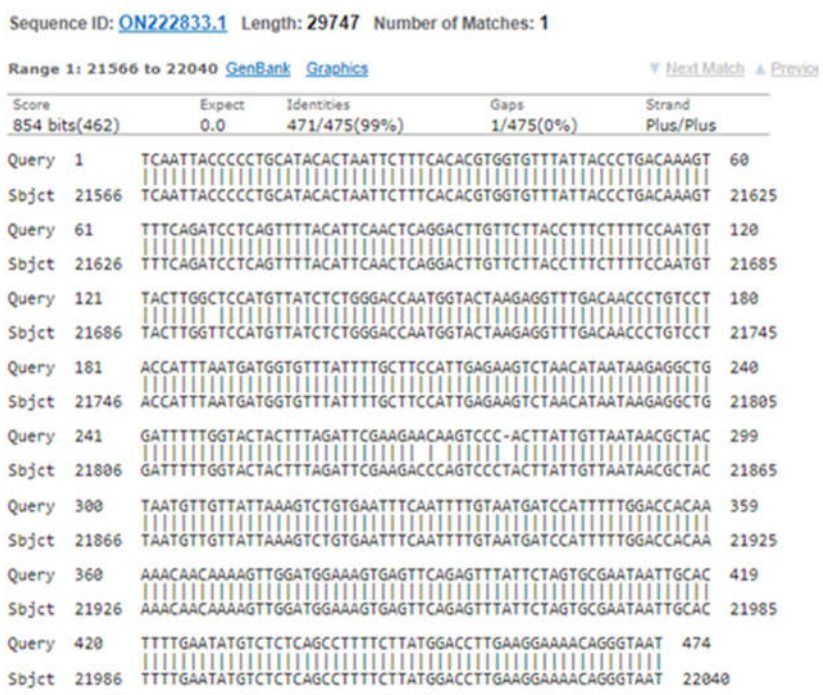


Figure 14. This is copy of the returned BLAST alignment report from the GenBank after submission of the 474-base composite sequence derived from the bidirectional sequencing data displayed in Figures 12 and 13. This report with a title "Severe acute respiratory syndrome coronavirus 2 isolate SARS-CoV-2/human/USA/FL-CDC-ASC210848875/2022, complete genome" shows that the submitted sequence has 3 nucleotide mutations and 1 nucleotide deletion when aligned against the best matching SARS-CoV-2 reference sequence in the GenBank database.

3.3.3. Next-generation sequencing showed Omicron BA.2 NTD only

Because of the unusual finding of a BA.1 NTD associated with a BA.2 RBD sequence in one clinical sample by target Sanger sequencing, an aliquot of the M22-75 nasopharyngeal swab sample was submitted to the Connecticut Department of Public Health Katherine A. Kelley State Public Health Laboratory to be sequenced with the Clear Labs next-generation DNA sequencing (NGS) instrument. According to information received from the State Public Health Laboratory, the NGS instrument generated a typical Omicron BA.2 NTD mutations, along with a typical BA.2 RBD sequence identical to those listed in Figure 11. The S gene NTD sequence of the NGS FASTA file was copied and pasted in Figure 15.

```
NNNNNNNNNNNNNATGTTTGTTTTTCTTGTTTTATTGCCACTAGTCTCTAGTCAGTGTGT
TAATCTTATAACCAGAACTCAATCATACACTAATTCTTTCACACGTGGTGTTTATTACCC
TGACAAAGTTTTCAGATCCTCAGTTTTACATTCAACTCAGGACTTGTTCTTACCTTTCTT
TTCCAATGTTACTTGGTTCCATGCTATACATGTCTCTGGGACCAATGGTACTAAGAGGTT
TGATAACCCTGTCCTACCATTTAATGATGGTGTTTATTTTGCTTCCACTGAGAAGTCTAA
CATAATAAGAGGCTGGATTTTTGGTACTACTTTAGATTCGAAGACCCAGTCCCTACTTAT
TGTTAATAACGCTACTAATGTTGTTATTAAAGTCTGTGAATTTCAATTTTGTAATGATCC
ATTTTTGGATGTTTATTACCACAAAAACAACAAAAGTTGGATGGAAAGTGAGTTCAGAGT
TTATTCTAGTGCGAATAATTGCACTTTTGAATATGTCTCTCAGCCTTTTCTTATGGACCT
TGAAGGAAAACAGGGTAATTTCAAAAATCTTAGGGAATTTGTGTTTAAGAATATTGATGG
TTATTTTAAAATATATTCTAAGCACACGCCTATTAATTTAGGGCGTGATCTCCCTCAGGG
```

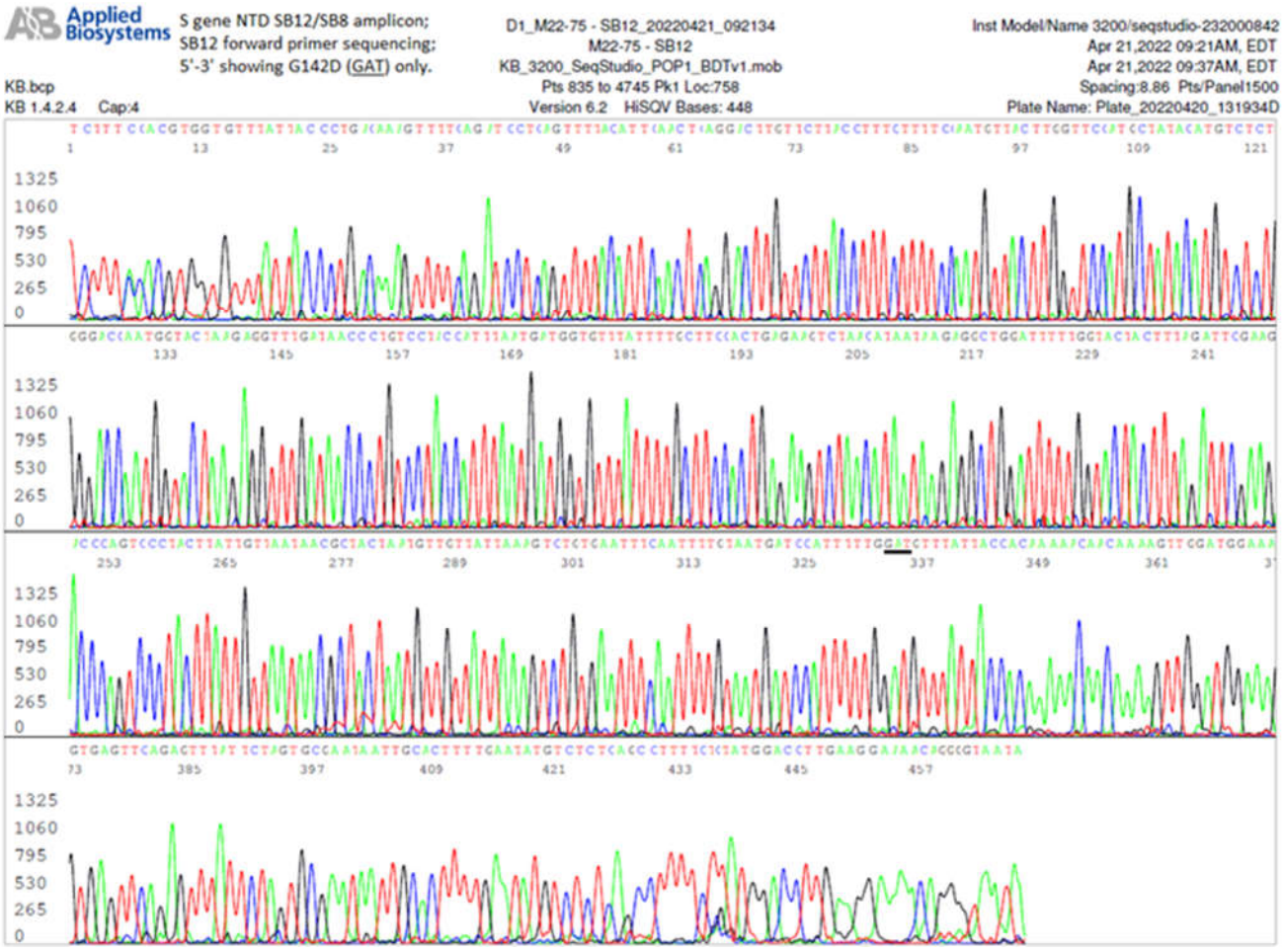
Figure 15. This is copy of the S gene NTD sequence excised from the FASTA file generated by NGS on sample M22-75. The key amino acid codons, which are involved in A67V, Δ69-70, T95I, G142D and Δ143-145 mutations, for distinguishing between the BA.1 and the BA.2 sub-lineages are framed in 4 rectangular boxes. Comparison of this sequence with that illustrated in Figure 12 shows that NGS did not detect A67V, Δ69-70, T95I and Δ143-145. The codon of G142D mutation, which is shared by both BA.1 and BA.2, was detected as "GAT" instead of "GAC" (Figure 12). These sequence discrepancies indicate that the NGS and Sanger sequencing technologies were using different allele templates in one sample to generate their respective NTD sequences.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

3.3.4. Selective RT-PCR amplification of S gene allele sequences by different primer sets

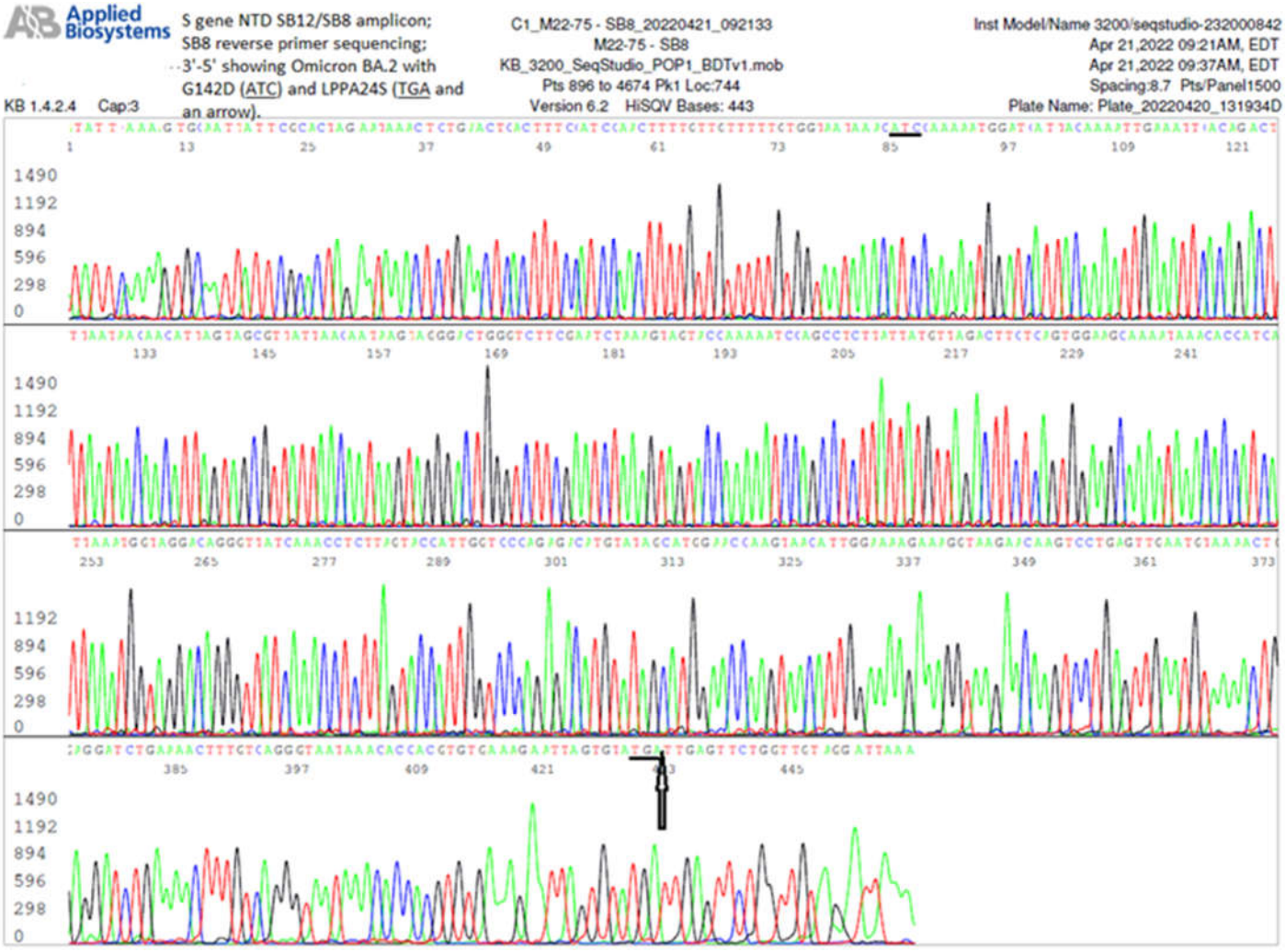
In order to verify by Sanger sequencing that the nucleic acid extract of sample M22-75 in fact contained both BA.1 and BA.2 NTD sequences, aliquots of the nucleic acid extract were re-amplified by two slightly different sets of nested RT-PCR primers, one for all known SARS-CoV-2 strains prior to the emergence of Omicron BA.2 and the other for all SARS-CoV-2 strains, including the BA.2 subvariants.

Although the BA.2 subvariants do not have A67V, Δ69-70, T95I, G142D and Δ143-145 mutations, they all have Δ24-26 and A27S mutations in the S gene NTD. Deletion of the three LPP24-26 codons (TTACCCCCT) in the BA.2 NTD rendered the SB5 and SB7 forward PCR primers nonfunctional for amplification of the BA. 2 subvariant NTD sequences. A new set of primary forward (SB11) and nested forward (SB12) PCR primers was designed to bypass the site of Δ24-26 for amplification of all Omicron S gene NTD sequences, including those of BA.2 sub-lineage for Sanger reaction. Bidirectional sequencing of the SB12/SB8 PCR products of sample M22-75 showed a typical BA.2 S gene NTD (Figure 16 A and B) with a sequence identical to that generated by NGS (see Figure 16).



(A)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

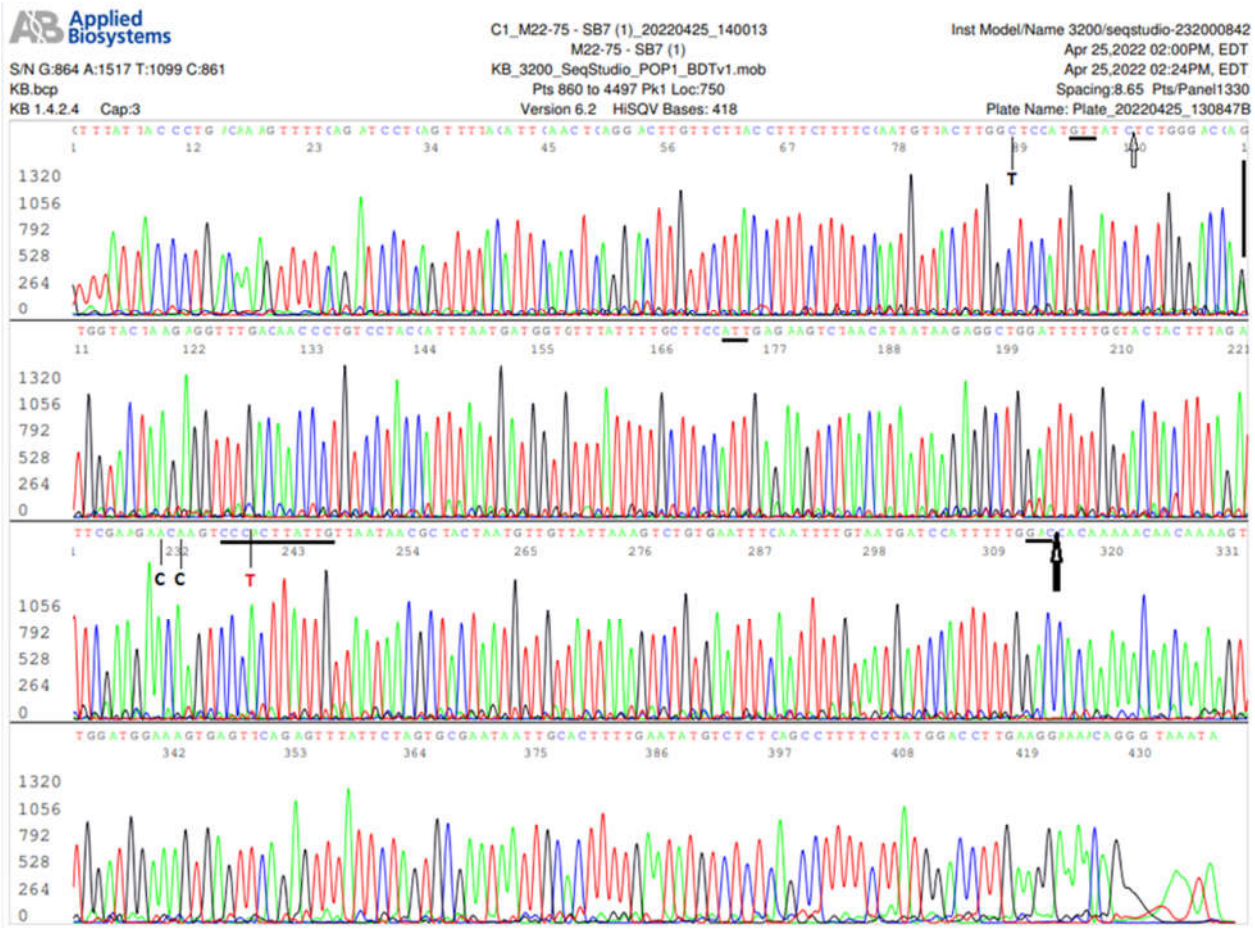


(B)

Figure 16. A and B. These are two electropherograms showing bidirectional sequencing of an SB12/SB8 nested PCR amplicon of the S gene NTD on sample M22-75, using SB12 primer as the forward sequencing primer (A) and SB8 primer as the reverse sequencing primer (B). The sample was first amplified by a pair of SB11/SB6 primary RT-PCR primers and the primary RT-PCR products were reamplified by a pair of SB12/SB8 nested PCR primers to generate the template for Sanger sequencing. The Δ24-26, A27S and G142D deletion/mutations characteristic of the BA. 2 subvariants are indicated by two underlined mutated codons and one arrow as for Figures 3 A and 4 A. The co-existent BA.1 NTD sequence in the sample was not amplified. In theory, both the SB11/SB6 primary RT-PCR primer pair and the SB12/SB8 nested PCR primer pair are capable of amplifying the BA.2 and BA.1 NTD sequences. However, no BA.1 NTD sequence was identified in Figure 16.

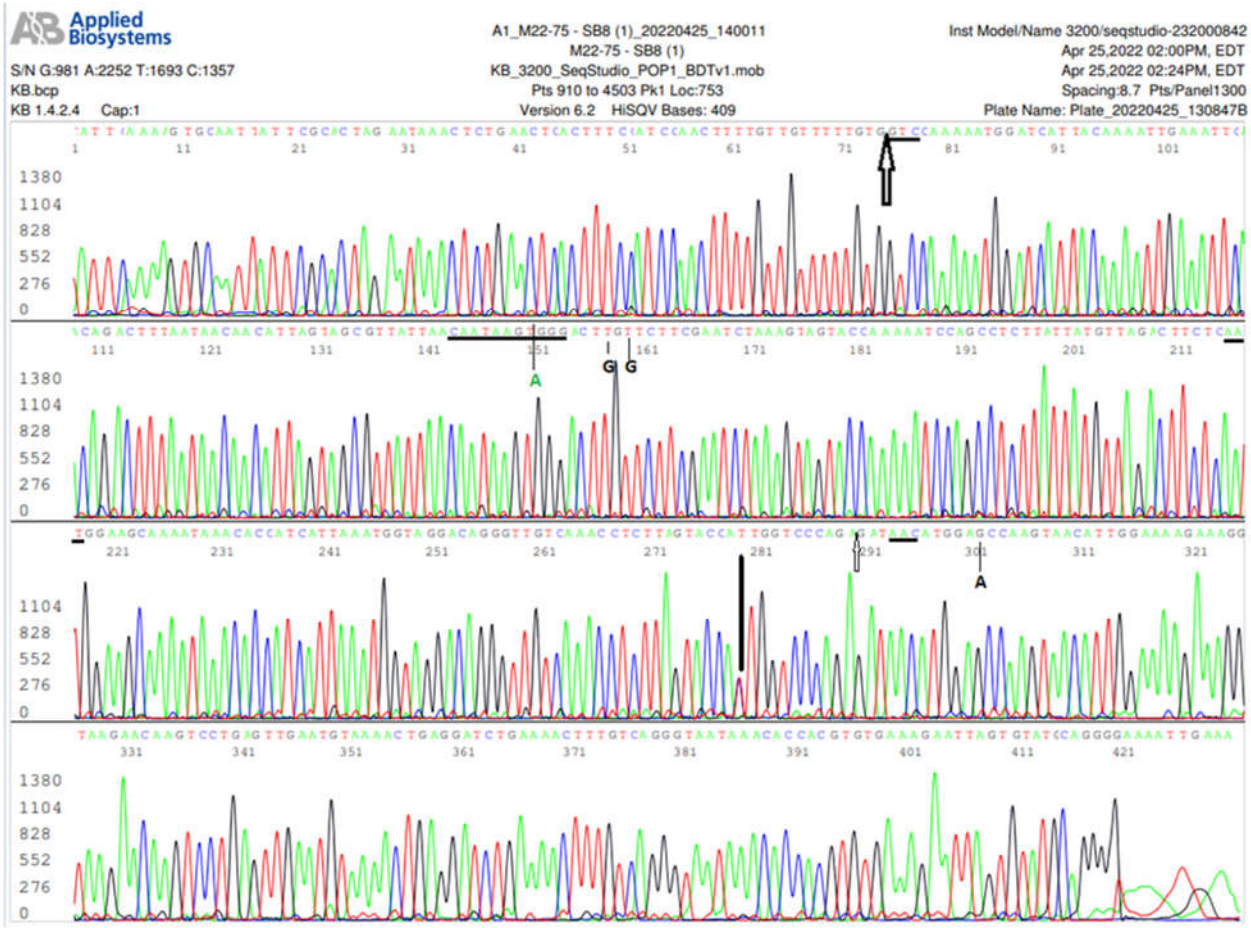
In order to provide irrefutable evidence that the atypical BA.1 NTD sequence illustrated in Figures 12 and 13 was not the result of technical errors or deviations, an aliquot of the nucleic acid extract, which was used to generate the sequences presented in Figures 12, 13 and 16 was re-amplified by SB5/SB6 primary RT-PCR followed by amplification of the primary PCR products with a pair of SB7/SB8 nested PCR primers. Bidirectional sequencing of the SB7/SB8 nested PCR products reproduced all the original results observed and presented in Figures 12 and 13. The newly reproduced electropherograms are presented in Figure 17 A and B.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1



(A)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1



(B)

Figure 17. A and B. These two electropherograms show the results of a repeated bidirectional sequencing of the S gene NTD nested RT-PCR product amplified by a pair of SB7/SB8 nested PCR primers on an aliquot of the same nucleic acid extract used to generate the sequences presented in Figures 12, 13 and 16. As described for Figures 12 and 13, the SB7 primer was used as the forward sequencing primer (A) and the SB8 primer as the reverse sequencing primer (B). All the mutations and deletions, including the superimposing competing A/G peaks, which were present in Figures 12 and 13, were fully reproduced as illustrated in Figure 17 A and B. Such a high degree of reproducibility in Sanger sequencing indicates that these mutations and deletion are true SNPs in the S gene NTD of an Omicron variant although these SNPs have not been reported in the world literature and annotated in the GenBank database, as shown in Figure 14.

3.3.5. Questionable recombined BA.1 NTD and BA.2 RBD in the Omicron S gene

Attempts were made to generate a >1200 bp long RT-PCR amplicon from sample M22-75, including the mutations of the NTD shown in Figure 17 and the key mutations of the RBD shown in Figure 11, to be used as one sequencing template but failed. Two new sets of PCR primers were designed to amplify a 600-700 bp segment of the S gene at the junction between the RBD and the NTD to investigate if the sample M22-75 contained a BA.1 RBD in addition to the BA.2 RBD sequence shown in Figure 11. The results are illustrated in Figure 18.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

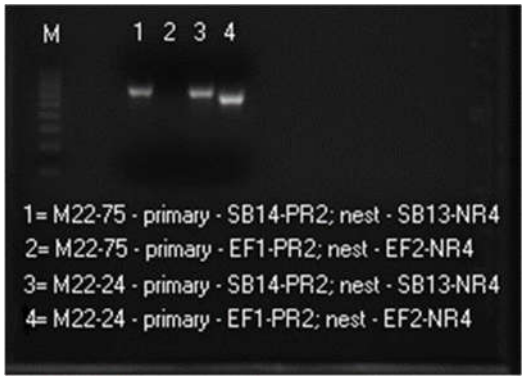
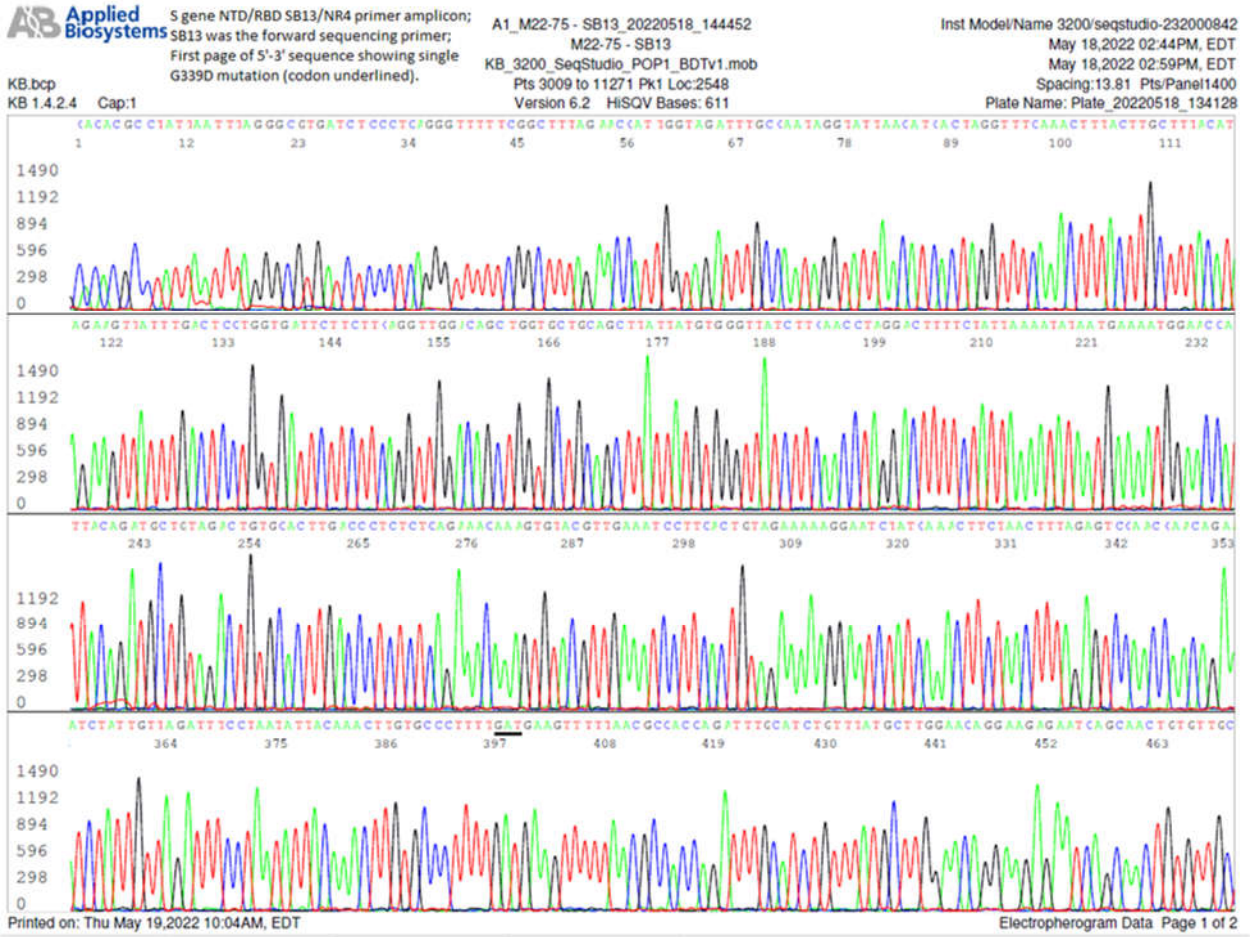


Figure 18. This is an image of agarose gel electrophoresis of the nested PCR products amplified by the SB13/NR4 primer pair for BA.1 and BA.2, and by the EF2/NR4 primer pair for BA.1 only, on sample M22-75 (lanes 1 and 2) and on sample M22-24 (lanes 3 and 4) as a known Omicron BA.1 control. The sequences of the primary and nested PCR primers are listed in Table 1. This gel image shows that the SB13/NR4 primer pair generated a nested PCR product band on both samples (lanes 1 and 3). But the EF2/NR4 primer pair generated only a nested PCR product band on M22-24 (lane 4), but not on M22-75 (lane 2). Since the EF1 and EF2 forward primers have the 214 EPE insert sequence in the 3′ end, the EF2/NR4 primers selectively amplified the BA.1 RBD. The absence of a PCR band in lane 2 indicates that the M22-75 sample did not contain an amplifiable BA.1 RBD sequence; a sequence of BA.2 RBD in the lane 1 PCR product is verified by the electropherogram presented in Figure 19.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

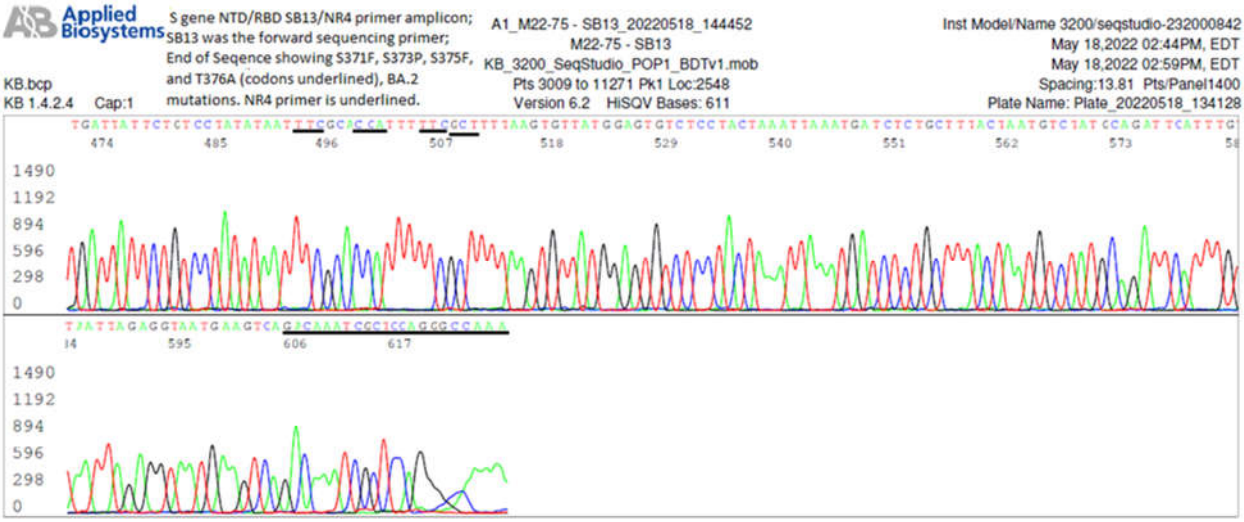


Figure 19. This two-paged electropherogram shows a 5′-3′ forward sequencing of the 694-bp nested PCR product illustrated in Lane 1 in Figure 18. The mutated codons G339D, S371F, S373P, S375F, and T376A indicative of a BA.2 RBD and the 21-base NR4 primer site in the end of the sequence are underlined.

3.4. Two mechanisms for the failure of Omicron S gene target RT-PCR amplification

Unexpected sequence mutation in primer-binding sites is a well-known cause of PCR failure in nucleic acid-based test for SARS-CoV-2 [32]. As previously reported, among 29 nasopharyngeal swab samples collected in January 2022 that were confirmed to contain a segment of SARS-CoV-2 N gene by Sanger sequencing, routine nested RT-PCR failed to generate a template from the S gene RBD or NTD for Sanger sequencing in 2 (7%) samples. Moving the RBD target upstream by using a new set of PCR primers successfully generated a 445-bp cDNA amplicon for sequencing in 1 of these two samples, identified as M22-51 [27]. Further investigation revealed that the mechanism leading to the initial failure of the NTD amplification was different from that responsible for the RBD RT-PCR failure in this sample.

3.4.1. Mutation of primer-binding site caused the RBD RT-PCR failure

As reported previously, the nasopharyngeal swab of sample M22-51 was positive for SARS-CoV-2 N gene with R203K and G204R mutations verified by Sanger sequencing. But routine nested RT-PCR amplification of the S gene RBD and NTD segments failed to generate a band on agarose gel electrophoresis; bidirectional Sanger sequencing confirmed the absence of PCR products [27]. Moving the RBD RT-PCR primers 291 bases upstream toward the NTD was able to generate a 445-bp nested PCR amplicon to be used as the sequencing template, thus providing sequencing evidence for the diagnosis of an Omicron BA.1 subvariant, as shown in Figures 20 and 21.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

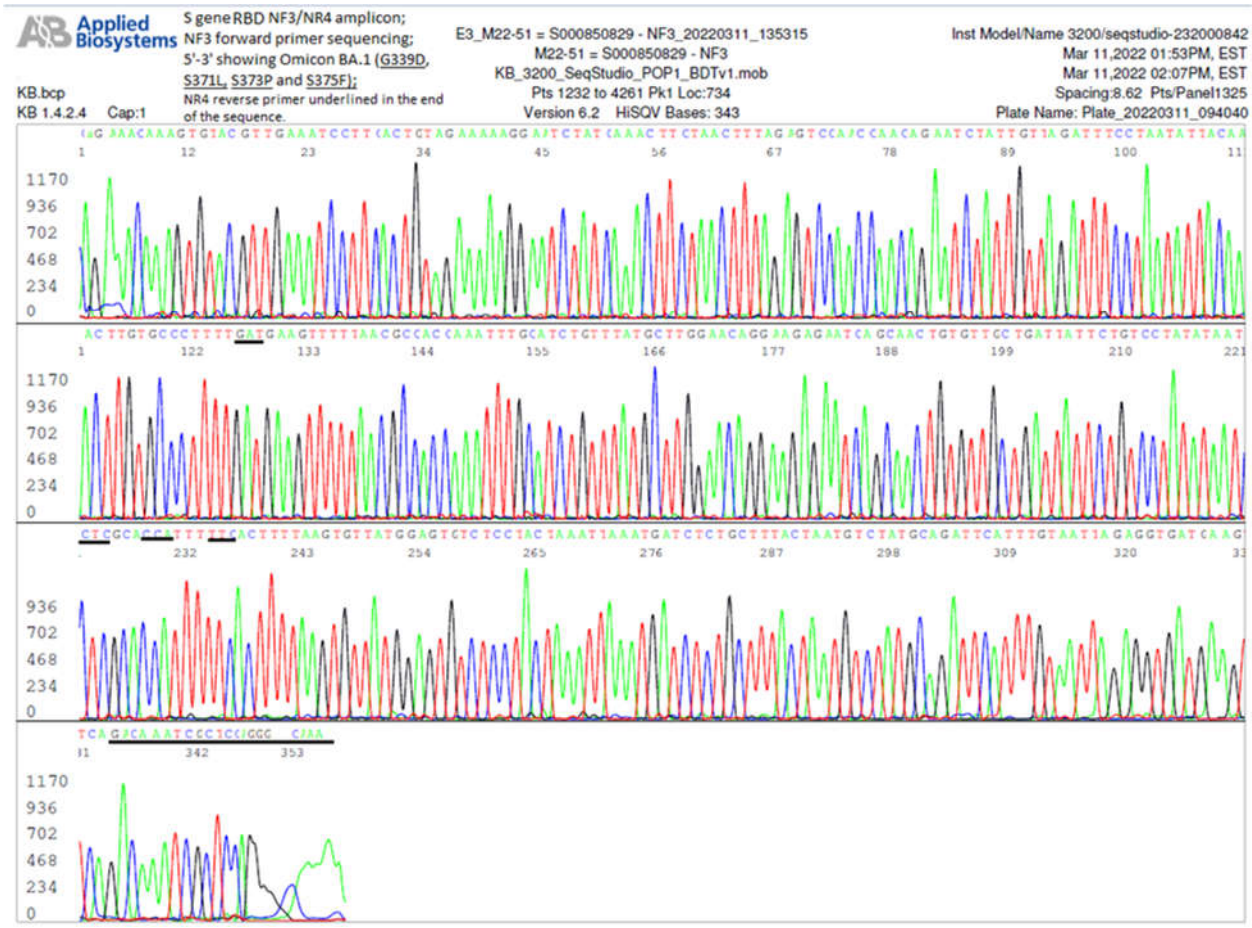


Figure 20. This is an electropherogram showing part of an S gene RBD sequence generated using NF3 nested PCR primer as the forward sequencing primer on the nasopharyngeal swab sample No. M22-51 collected in January, 2022. The routine RT-PCR primers designed for amplification of the S gene RBD and NTD of the SARS-CoV-2 Omicron BA.1 variant could not generate a nested PCR amplicon to be used as the template for Sanger sequencing on this sample [27]. A new set of nested PCR primers, NF3: 5′-GTGGGTTATCTTCAACCTAGG (22361-22381) and NR4: 5′-TTTGCCCTG-GAGCGATTTGTC (22785-22805), was used to amplify a 445-bp upstream segment of the RBD for sequencing, which confirmed the presence of G339D, S371L, S373P and S375F mutations (mutated codons and NR4 primer site underlined), consistent with an Omicron BA.1 subvariant. However, sequencing this alternative target segment for evaluation misses most of the key amino acid mutations in the RBD, spanning from D405 to Y505.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

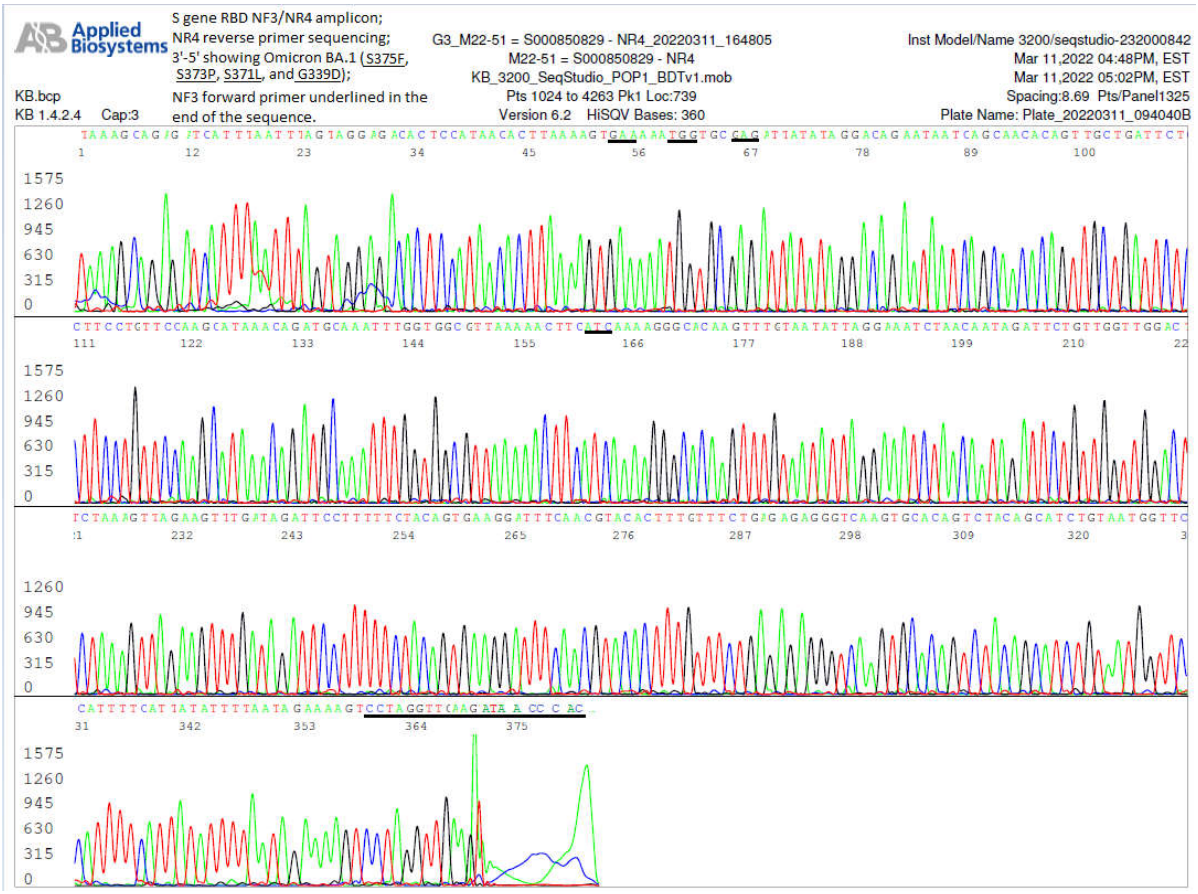


Figure 21. This is a reverse-complementary sequence generated from the same nested PCR product used to produce the sequence presented in Figure 20. The NR4 reverse nested PCR primer was the sequencing primer. The mutated codons for the S375F, S373P, S371L and G339D and NF3 primer site in 3′-5′ direction are underlined.

Both Figures 20 and 21 show that there were little minor subvariant sequences with multi-allelic SNPs in the RBD in the bidirectional sequencing electropherograms. The initial routine RBD RT-PCR failure was due to mutations affecting the PCR primer-binding site(s).

3.4.2. Minor subvariant sequences with multi-allelic SNPs caused S gene RT-PCR amplification failure

In routine diagnostic tests for SARS-CoV-2 variants, the primary and nested forward RT-PCR primers, labeled SB5 and SB7, respectively, were placed in a location of the S gene NTD to cover the A67V, Δ69-70, T95I, G142D and Δ143-145 mutations commonly used to help define variants of concern. The sequence of the SB5 primary forward PCR primer is 5′-AACCAGAACTCAATTACCCCC-3′ and that of the SB7 nested forward PCR primer is 5′-TCAATTACCCCCTGCATACAC-3′. These two sequences are highly conserved among all early SARS-CoV-2 variants, including the initial Omicron. The amplicon used for Sanger sequencing is 490 bp in size for the wildtype Wuhan strains. However, as stated in Section 3.3.4., the recently emerging Omicron subvariants, such as those of the BA.2 sublineage, have a Δ24-26 LPP (TTACCCCCT) deletion, which renders these forward primers nonfunctional. Two new primary and nested forward PCR primers, labeled SB11 and SB12, respectively, were used to replace the SB5 and SB7 forward primers in order to bypass the Δ24-26 site. The success of using the SB11 and SB12 forward primers to amplify the BA.2 NTD was illustrated in the sequences presented in Figures 4, 9 and 16B, which all ended in the SB12 forward nested PCR primer sequence “GAGTTCTGGTTGTAAGATTAA” (reverse complement to SB12).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1

But surprisingly, while the SB5/SB7 PCR forward primers, which were used successfully to amplify the NTD of other Omicron BA.1 subvariants for Sanger sequencing, failed to amplify the S gene NTD segment in sample M22-51 [27], the SB11/SB12 forward primers specifically designed to bypass the Δ24-26 to cover the BA.2 sub-lineage was capable of amplifying a SB12/SB8 primer-defined BA.1 NTD nested RT-PCR amplicon for Sanger sequencing. Yet, the computer was able to perform base-calling in the forward sequencing only, but failed to call the bases in the reverse sequencing direction due to interference by minor subvariant sequences with multi-allelic SNPs. The computer-generated bidirectional sequencing electropherograms are presented in Figure 22 and Figure 23 for illustration.

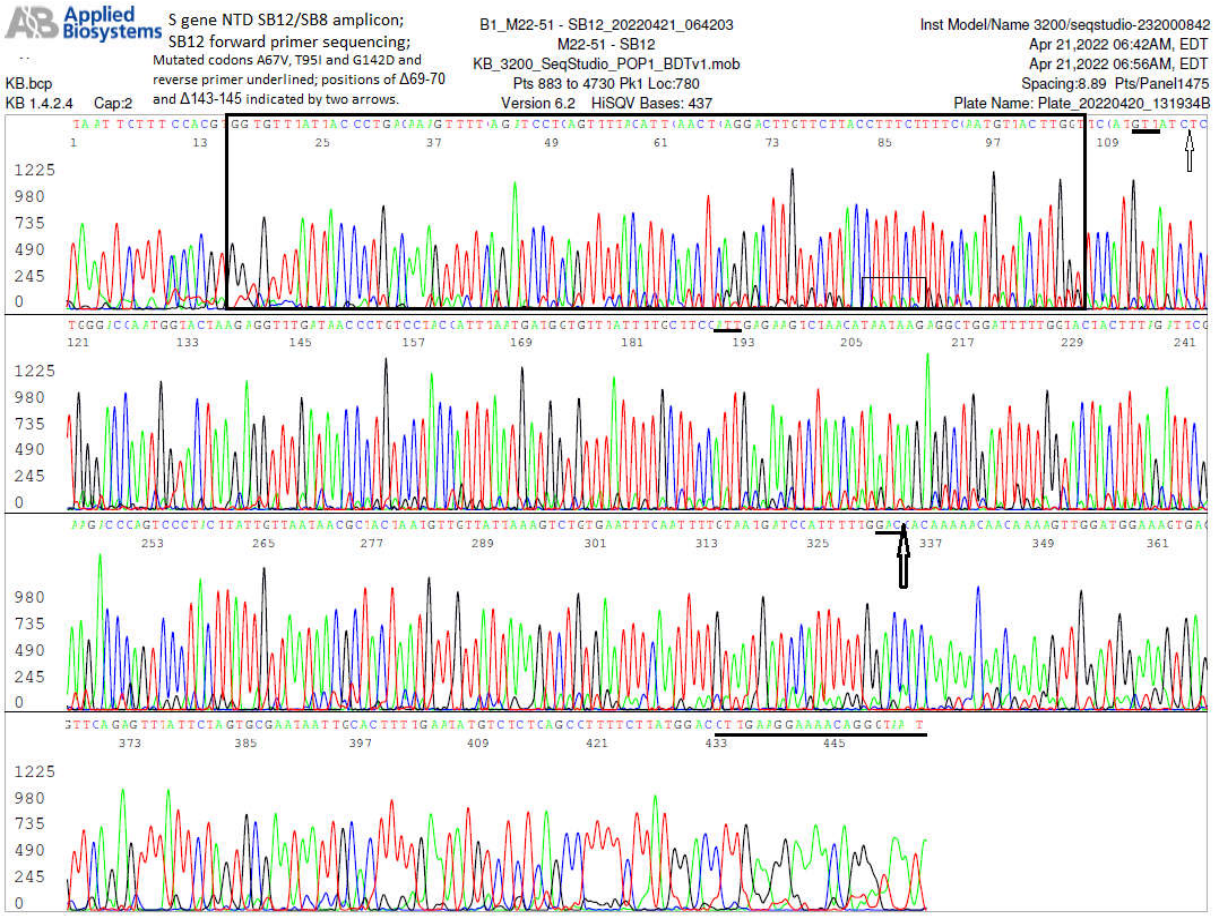
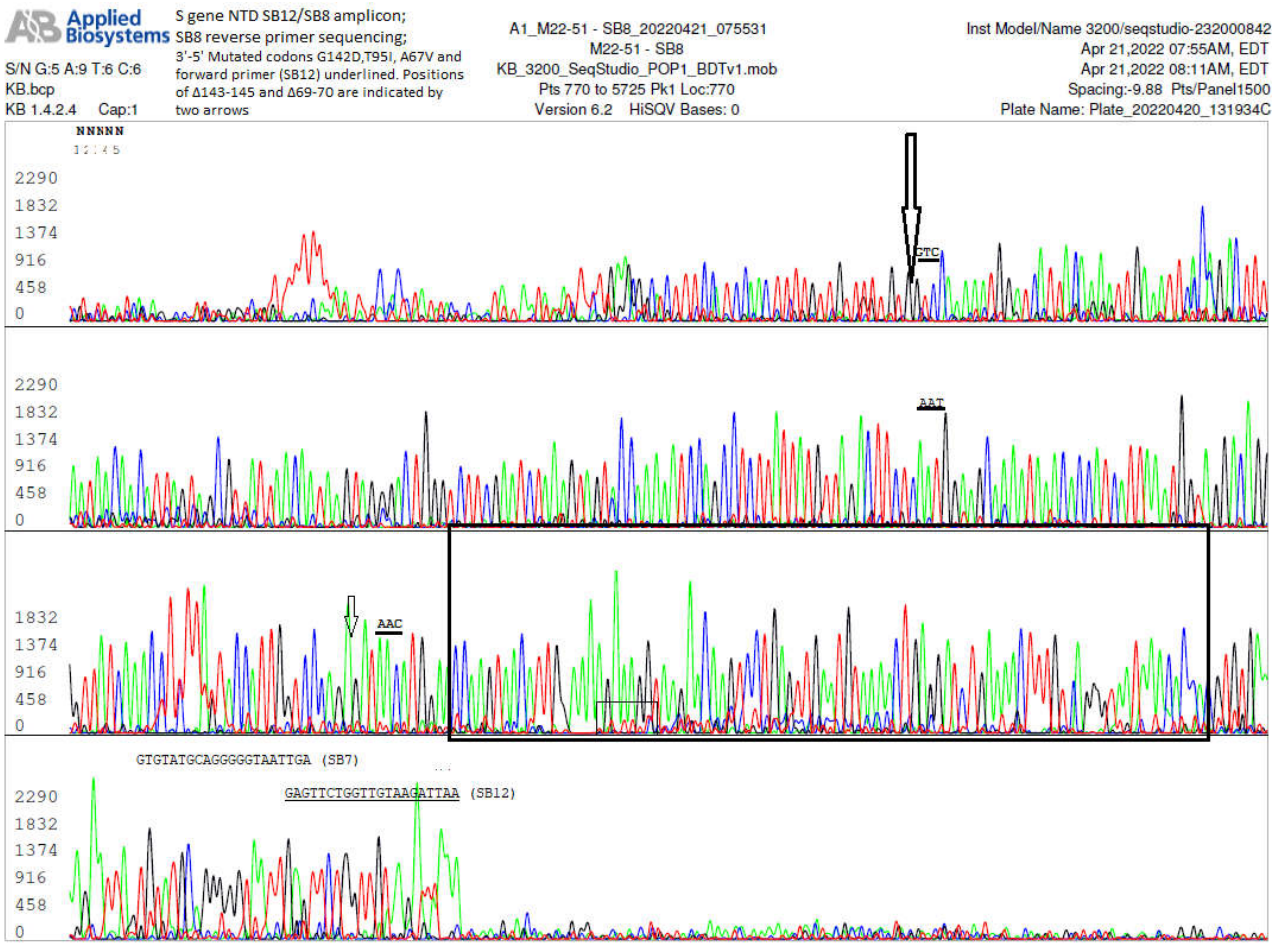
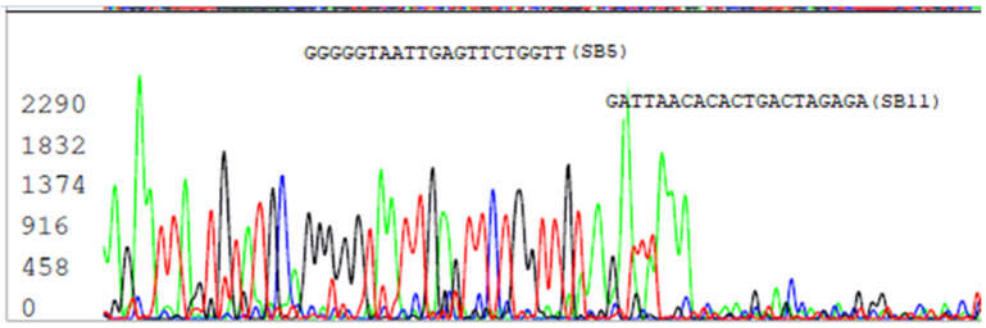


Figure 22. This electropherogram shows the dominant S gene NTD sequence with A67V, Δ69-70, T95I, G142D and Δ143-145 mutations, characteristic of an Omicron BA.1 subvariant in sample M22-51. The template was a nested RT-PCR product generated by a pair of SB12/SB8 PCR primers after the SB7/SB8 primer pair failed to produce a nested PCR product. The SB12 forward primer was used as the sequencing primer for Figure 22. The mutated codons are underlined and the positions of Δ69-70 and Δ143-145 are indicated by a small arrow and a big arrow, respectively. Under the dominant sequence peaks on which the computer depended for base calling, there are numerous colored low peaks. These colored low peaks are interpreted as representing minor subvariant sequences with multi-allelic SNPs, not sequencing noise because many of these low-peak sequences were reproduced in the reverse primer sequencing shown in Figure 23 A. The minor subvariant sequences with multi-allelic SNPs are most clearly demonstrated in the sequences framed by the two rectangular boxes. For example, there is a sequence of 4 green low "A" peaks in the small box under the dominant peaks of TTTC. These 4 green low "A" peaks were reproduced as 4 red low "T" peaks under the dominant peaks of GAAA in reverse complement within the small box in Figure 23 A. In Figure 22, the 21-base sequence of the SB8 reverse primer is underlined in the end of the sequence, indicating that the NTD region for the reverse primer-binding site in this sample does not have crowded SNPs that might suppress PCR amplification of the dominant sequence.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 14 June 2022 doi:10.20944/preprints202206.0192.v1



(A)



(B)

Figure 23. A and B. The electropherogram in (A) shows a reverse-complementary sequence generated from the same nested PCR product used to produce the sequence presented in Figure 22. Because of the massive segmental losses of sequencing signal, the computer was unable to perform its base-calling function. But a visual analysis successfully identified the sequences of Δ143-145, G142D, T95I, Δ69-70 and A67V mutations in 3′-5′ direction (mutated codons are underlined. The positions of Δ143-145 and Δ69-70 are indicated by a big arrow and a small arrow, respectively). Many colored low peaks under the high peaks of the dominant sequence were reproduced in reverse-complementary sequence. For example, there are 4 red low "T" peaks in the small rectangular box, against the 4 green low "A" peaks in the corresponding position located in the small box of Figure 22. These colored low peaks may represent one minor subvariant sequence with multi-allelic SNPs or numerous minor subvariant sequences each with one or several nucleotide mutations. They may be the products of exponential PCR amplification of a minor subvariant sequence with multi-allelic SNPs defined by the same PCR primer pair as those in the dominant sequence, or the products of numerous template-directed linear enzymatic extensions of one primer without a collaborating functional PCR primer from the opposite direction of the templates. The electropherogram in (B) is a subset of

 doi:10.20944/preprints202206.0192.v1

(A) with the primary RT-PCR primer SB5 and the primary RT-PCR primer SB11 aligned against their respective binding sites of the dominant sequence.

Comparing the primer-binding sites in the end of the sequence in Figure 22 and that in Figure 23 (A), both representing the bidirectional sequencing result of the same nested PCR products, showed that while the level of minor subvariant sequences with multi-allelic SNPs was not high enough to suppress the function of the SB8 reverse PCR primer when the SB8 primer was paired with an SB12 forward nested PCR primer, the crowded multi-allelic SNPs in the templates at the site binding the 3'-end part of an SB7 forward nested PCR primer (Figure 23 A) and the 3'-end part of the SB5 forward primary PCR primer (Figure 23 B) were probably the cause responsible for the failure of the RT-PCR amplification of the NTD by a pair of SB7/SB8 nested PCR primers on sample M22-51 [27]. Allele sequences are well-recognized PCR inhibitors [33,34]. Based on the locations of the primary forward RT-PCR primers shown in Figure 23 B, which is an excised part from Figure 23 A, there were less crowded multi-allelic SNPs at the site binding the 3'end of the SB11 primary forward RT-PCR primer than at the site binding the 3'end of the SB5 primary forward RT-PCR primer. With less inhibition by allele sequences, the SB11/SB6-initiated primary RT-PCR was able to generate enough dominant sequence copies to be used as the template in the SB12/SB8 nested PCR amplification although the minor sub-variant sequences were also co-amplified.

All laboratories using S gene target sequencing diagnostics for Omicron variants must be prepared to deal with minor subvariant sequences with multi-allelic SNPs and possible template mutations in the primer-binding sites both of which can cause PCR and/or sequencing failures.

4. Discussion

The SARS-CoV-2 Omicron variant with its subvariants, which are the currently dominant strains causing COVID-19 cases in the world, is characterized by its remarkably high number of cumulative mutations. The eCDC and the WHO have recommended partial Sanger or next-generation sequencing of the S gene RBD and NTD PCR amplicons to monitor their circulation in all countries [18]. Timely routine sequencing of all positive samples can reduce the >40% false-positive rates generated by the RT-qPCR test kits [26,27,35], in addition to offering a variant-diagnostic test, which is valuable for patient management and for policy-making. Rapid and accurate diagnosis of patients in the early stages of infection is the key step to curtail the COVID-19 pandemic. As pointed out by the then BMJ editor in chief in July 2021, "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" [36]. With rapid and accurate diagnostics by gene sequencing [21-23], SARS that emerged as a pandemic in 2002 in Guangdong, China, was stopped in 2003 within 7 months by applying travel restrictions and isolating individuals infected by SARS-CoV [37] before a variant could emerge to cause concern.

Diagnosis of SARS-CoV-2 variants needs nucleotide sequencing. Compared to NGS, Sanger sequencing of nested PCR products does not require high viral-load samples and does not need costly bioinformatic services for data analysis. In addition, the NGS technology is more prone to base-call errors and bias [38,39], which may require correction or verification by Sanger sequencing [40,41]. Therefore, Sanger sequencing is the method of choice for diagnostic laboratories. However, Sanger sequencing of PCR amplicon needs well-designed PCR primer sets, which must be periodically adjusted to accommodate newly emerging variants. The sequencing data presented in this paper emphasize the following key points:

 doi:10.20944/preprints202206.0192.v1

4.1. The N gene is the more reliable target than the S gene for amplicon sequencing-based diagnostics

As previously reported, testing a group of nasopharyngeal swab samples collected in October, 2020 in the U.S. prior to the emergence of any variant of concern by sequencing a 398-base N gene showed that there is a hypervariable 79-base stretch of nucleotide sequence corresponding to the position 28821 to 28899 (GenBank reference sequence NC_045512.2) [26]. This 79-base region is flanked by two segments of conserved sequence, which were used to design the PCR primers (referred to as Co1, Co8, Co4 and Co3 in Table 1) for the diagnostic N gene-sequencing assay. The N gene of the Omicron variant strains invariably harbors the R203K and G204R mutations. But these two mutations were already sporadically reported in nasopharyngeal swab specimens collected in the early part of 2020 [42] long before the appearance of the Omicron variant. Therefore, variant determination based on N gene sequencing alone is not reliable. Nevertheless, compared to the S gene RBD and NTD, this N gene target is more conserved, especially in the primer-binding sites, and is associated with a much lower level of homeologous minor subvariant sequences with multi-allelic SNPs, which may suppress RT-PCR amplification [33, 34].

4.2. Co-existing minor Omicron subvariants and their effects on variant diagnosis

After the SARS-CoV-2 has circulated for more than 2 years from population to population since its outbreak, numerous mutated subvariant sequences with multi-allelic SNPs [9] have been accumulated in the virus and show up in some of the emerging variant isolates displaying genetic diversity within single infected hosts [43, 44]. The number of these minor subvariant sequences with multi-allelic SNPs in some clinical samples positive for Omicron may be extremely high. Most surveillance studies relying on testing high-viral-load samples with NGS only consider a single consensus sequence for each infected person for statistic purpose and ignore the co-existing subvariant sequences. Bioinformatic analysis of the NGS data is often based on alignment, or mapping, of reads against a reference sequence followed by the consensus extraction by majority voting. If the studied virus sequence is divergent from the chosen reference sequence, the reads covering the regions of divergence could not be aligned correctly or might be discarded, which will bias the resulting consensus [45]. It has been reported that the key sites of the S gene known to harbor mutations of interest, such as the K417N/T, E484K and N501Y in the RBD, are in the regions of low read coverage when NGS is used and may need target Sanger sequencing to recover [46]. For diagnostic purpose, routine target Sanger sequencing of the RBD is a better approach to determine variants [18]. However, Sanger sequencing can generate a dominant sequence for accurate mutation analysis only if the level of minor subvariant sequences with multi-allelic SNPs in the sample is not high enough to suppress the PCR amplification of the dominant target sequence to be used as the sequencing template or to interfere with the Sanger sequencing process. The failure of RT-PCR amplification of a target RBD or NTD sequence may be due to mutations affecting the primer-binding sites in the SARS-CoV-2 genome or due to the presence of an overwhelming amount of minor subvariant sequences with multi-allelic SNPs. If a target PCR amplicon band is visualized at agarose gel electrophoresis, the failure to generate a readable electropherogram may be mitigated by performing a bidirectional sequencing (Figures 5-7; Figures 22 and 23). If no S gene target PCR amplicon is visualized at gel electrophoresis when the N gene sequencing of the sample is positive for SARS-CoV-2, a new PCR primer set for the S gene amplification may be required (Figures 20-23).

4.3. Possible Omicron BA.2 and BA.1 S gene recombinant

Target Sanger sequencing can reveal important information that NGS misses, as demonstrated in Section 3.3.

As a whole, Omicron subvariants have a high number of mutations in the spike protein gene and these mutations mainly occur in the NTD and RBD of the S gene [47]. Determination of the mutations in the NTD and RBD in any given sample usually generates

 doi:10.20944/preprints202206.0192.v1

enough information for accurate diagnosis of the key Omicron subvariants. However, there are exceptions. For example, in one specimen, M22-75, Sanger sequencing showed S371F, S373P, S375F, T376A, D405N, R408S, K417N, N440K, S477N, T478K, E484A, Q493R, Q498R, N501Y and Y505H (Figure 11), the 15 typical mutations in the RBD for a BA.2 subvariant, but A67V, Δ69-70, T95I, G142D, and Δ143-145 (Figure 12), a profile of mutations in the NTD characteristic of a BA.1 subvariant. In addition, there were 3 single nucleotide mutations, one nucleotide deletion and one overlapping single-nucleotide polymorphism (SNP), converting the wildtype base "A" to "G" in an allele sequence to create a new nonsynonymous N74R mutation in the new allele.

After next-generation sequencing of an aliquot of sample M22-75 generated a consensus sequence of Omicron BA.2 (Figure 15), repeated bidirectional Sanger sequencing of the original sample confirmed the previous findings of a BA.1 NTD sequence with all the mutations and deletion (Figure 17). In addition, new PCR primers were used to generate templates for Sanger sequencing to confirm that the M22-75 sample contained a BA.2 NTD sequence (Figure 16) and a BA.2 RBD (Figure 19) but did not harbor a BA.1 RBD sequence (Figure 18). Therefore, the co-existence of a BA.1 NTD and a BA.2 NTD in one sample was not due to co-infection by an Omicron BA.1 and an Omicron BA.2 subvariants, as reported by others [48].

In other words, on sample M22-75 target partial Sanger sequencing has demonstrated that one host was infected with at least 3 Omicron subvariants, which share one BA.2 RBD sequence, but harbor 3 different S gene NTD sequences, namely a typical BA.2 NTD sequence, an atypical BA.1 NTD sequence and an atypical BA.1 NTD sequence with an extra N74R mutation. The two atypical NTD sequences do not fully match with any SARS-CoV-2 genomic sequences annotated in the GenBank database (Figure 14), and as allele sequences their clinical significance remains unknown. However, there is no direct sequencing evidence that these BA.1 NTD sequences are in fact connected to a BA.2 RBD in one single PCR amplicon of about 1,500 bp in size. Demonstration of such connection in one template would be the irrefutable proof for a BA.1/BA.2 S gene recombination, but is difficult to accomplish in diagnostic RT-PCR with patient samples.

5. Conclusions

The eCDC and the WHO recommend partial Sanger sequencing of the SARS-CoV-2 S gene RBD and NTD on the PCR-positive samples in diagnostic laboratories as a practical means to determine variants of concern to monitor a possible increased transmissibility, increased virulence, or reduced effectiveness of vaccines against them. This paper presented bidirectional Sanger sequencing of 3 gene targets on five selected Omicron variant patient samples to show the potential interfering effects of co-existing minor subvariant sequences with multi-allelic SNPs on Sanger sequencing of RT-PCR products and the need to adjust the PCR primer sequences when the BA.2 NTD is also a target for PCR amplification in order to bypass the LPPA24S mutation that is unique for the BA.2 and BA4/BA.5 sub-lineages. Unlike next-generation sequencing, which focuses on deriving consensus sequence, Sanger sequencing may reveal a BA.1 NTD sequence in a sample containing a BA.2 RBD, depending on the PCR primers used to amplify the "target". Infection with more than one variant or with a recombinant variant may be encountered more often if Sanger sequencing is implemented in diagnostic laboratories as recommended.

Funding: This research received no external funding.

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: Four patients have granted written consent to use their specimens for research and to publish the results with concealed patient identification.

Data Availability Statement: Not applicable.

Acknowledgments: The author thanks Wilda Garayua for her technical assistance. The author also thanks Dr. Claire Pearson of the Connecticut Department of Public Health Katherine A. Kelley State Public Health Laboratory for performing NGS on specimen M22-75 and providing the FASTA file part of which forms the image presented in Figure 14.

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.

References

1. Coronavirus Updates. Worldometer. Available online: https://www.worldometers.info/coronavirus / (accessed on 6 May 2022).
2. Denison, M. R., Graham, R. L., Donaldson, E. F., Eckerle, L. D., & Baric, R. S. (2011). Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. RNA biology, 8(2), 270–279. https://doi.org/10.4161/rna.8.2.15013
3. 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
4. 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
5. Mallapathy S. The hunt for the origins of omicron. https://media.nature.com/original/magazine-assets/d41586-022-00215-2/d41586-022-00215-2.pdf
6. Wright, C. F., Morelli, M. J., Thébaud, G., Knowles, N. J., Herzyk, P., Paton, D. J., Haydon, D. T., & King, D. P. (2011). Beyond the consensus: dissecting within-host viral population diversity of foot-and-mouth disease virus by using next-generation genome sequencing. Journal of virology, 85(5), 2266–2275. https://doi.org/10.1128/JVI.01396-10
7. Nikitin, N., Petrova, E., Trifonova, E. & Karpova, O. Infuenza virus aerosols in the air and their infectiousness. Adv. Virol. 2014, 20 (2014).
8. Basu, S. Computational characterization of inhaled droplet transport to the nasopharynx. Sci Rep 11, 6652 (2021). https://doi.org/10.1038/s41598-021-85765-7
9. 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 Europeen sur les maladies transmissibles = European communicable disease bulletin, 25(22), 2000746. https://doi.org/10.2807/1560-7917.ES.2020.25.22.2000746
10. Tonkin-Hill, G., Martincorena, I., Amato, R., Lawson, A., Gerstung, M., Johnston, I., Jackson, D. K., Park, N., Lensing, S. V., Quail, M. A., Gonçalves, S., Ariani, C., Spencer Chapman, M., Hamilton, W. L., Meredith, L. W., Hall, G., Jahun, A. S., Chaudhry, Y., Hosmillo, M., Pinckert, M. L., … Wellcome Sanger Institute COVID-19 Surveillance Team (2021). Patterns of within-host genetic diversity in SARS-CoV-2. eLife, 10, e66857. https://doi.org/10.7554/eLife.66857
11. Lythgoe, K. A., Hall, M., Ferretti, L., de Cesare, M., MacIntyre-Cockett, G., Trebes, A., Andersson, M., Otecko, N., Wise, E. L., Moore, N., Lynch, J., Kidd, S., Cortes, N., Mori, M., Williams, R., Vernet, G., Justice, A., Green, A., Nicholls, S. M., Ansari, M. A., … Golubchik, T. (2021). SARS-CoV-2 within-host diversity and transmission. Science (New York, N.Y.), 372(6539), eabg0821. https://doi.org/10.1126/science.abg0821
12. CDC. SARS-CoV-2 Variant Classifications and Definitions. Available online: https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html (accessed on 15 August 2021).
13. WHO. Tracking SARS-CoV-2 Variants. Available online: https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/(accessed on 25 August 2021).
14. Walker, A. S., Vihta, K. D., Gethings, O., Pritchard, E., Jones, J., House, T., Bell, I., Bell, J. I., Newton, J. N., Farrar, J., Diamond, I., Studley, R., Rourke, E., Hay, J., Hopkins, S., Crook, D., Peto, T., Matthews, P. C., Eyre, D. W., Stoesser, N., … Covid-19 Infection Survey Team (2021). Tracking the Emergence of SARS-CoV-2 Alpha Variant in the United Kingdom. The New England journal of medicine, 385(27), 2582–2585. https://doi.org/10.1056/NEJMc2103227
15. eCDC. Situation Updates on Covid-19. SARS-CoV-2 variants of concern as of 5 May 2022. https://www.ecdc.europa.eu/en/covid-19/variants-concern
16. CNN news report. Newer, fitter descendants of Omicron variant begin to drive their own coronavirus waves. By Brenda Goodman. Updated 5:26 PM ET, Tue May 3, 2022. https://www.cnn.com/2022/05/03/health/fitter-omicron-descendants-covid-variants/index.html
17. 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 May 2022).
18. 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
19. Alaofi, A. L., & Shahid, M. (2021). Mutations of SARS-CoV-2 RBD May Alter Its Molecular Structure to Improve Its Infection Efficiency. Biomolecules, 11(9), 1273. https://doi.org/10.3390/biom11091273

20. Klinakis, A., Cournia, Z., & Rampias, T. (2021). N-terminal domain mutations of the spike protein are structurally implicated in epitope recognition in emerging SARS-CoV-2 strains. Computational and structural biotechnology journal, 19, 5556–5567. Advance online publication. https://doi.org/10.1016/j.csbj.2021.10.004
21. 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. [CrossRef]
22. Ksiazek, T.G.; Erdman, D.; Goldsmith, C.S.; Zaki, S.R.; Peret, T.; Emery, S.; Tong, S.; Urbani, C.; Comer, J.A.; Lim, W.; et al. A novel coronavirus associated with severe acute respiratory syndrome. N. Engl. J. Med. 2003, 348, 1953–1966. [CrossRef]
23. CDC. SARS-CoV Specific RT-PCR Primers. Available online: https://www.who.int/csr/sars/CDCprimers.pdf?ua=1 (accessed on 7 September 2021).
24. McCarty, S.C.; Atlas, R.M. Effect of amplicon size on PCR detection of bacteria exposed to chlorine. PCR Methods Appl. 1993, 3, 181–185. [CrossRef]
25. 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.
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. Evidence-Based Evaluation of PCR Diagnostics for SARS-Cov-2 and The Omicron Variants by Sanger Sequencing. Preprints 2022, 2022040091 https://www.preprints.org/manuscript/202204.0091/v1 (doi: 10.20944/preprints202204.0091.v1).
28. Andrew Rambaut. Cov-lineages/pango-designation. Proposal to split B.1.1.529 to incorporate a newly characterised sibling lineage #361 https://github.com/cov-lineages/pango-designation/issues/361
29. eCDC. Threat Assessment Brief: Implications of the emergence and spread of the SARS-CoV-2 B.1.1. 529 variant of concern (Omicron) for the EU/EEA. https://www.ecdc.europa.eu/en/publications-data/threat-assessment-brief-emergence-sars-cov-2-variant-b.1.1.529
30. Cao, Y., Yisimayi, A., Jian, F. et al. BA.2.12.1, BA.4 and BA.5 escape antibodies elicited by Omicron infection. bioRxiv 2022.04.30.489997; doi: https://doi.org/10.1101/2022.04.30.489997
31. Deadline News. By Tom Tapp. New Omicron Variant BA.2.12.1 Now Dominant In New York, Driving Infections. https://deadline.com/2022/04/new-omicron-variant-ba-12-1-dominant-new-york-1235010160/
32. Zimmermann, F., Urban, M., Krüger, C., Walter, M., Wölfel, R., & Zwirglmaier, K. (2022). In vitro evaluation of the effect of mutations in primer binding sites on detection of SARS-CoV-2 by RT-qPCR. Journal of virological methods, 299, 114352. https://doi.org/10.1016/j.jviromet.2021.114352
33. McCord, B.; Pionzio, A.; Thompson, B. Analysis of the Effect of a Variety of PCR Inhibitors on the Amplification of DNA using Real Time PCR, Melt Curves and STR Analysis. The U.S. Department of Justice. Document No. 249148. September 2015 https://www.ojp.gov/pdffiles1/nij/grants/249148.pdf
34. Ikegawa, S., Mabuchi, A., Ogawa, M., & Ikeda, T. (2002). Allele-specific PCR amplification due to sequence identity between a PCR primer and an amplicon: is direct sequencing so reliable?. Human genetics, 110(6), 606–608. https://doi.org/10.1007/s00439-002-0735-1
35. 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
36. Godlee, F. Caution, vaccines, testing: the only way forward. BMJ 2021;374:n1781 http://dx.doi.org/10.1136/bmj.n1781
37. 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
38. Dohm, J. C., Lottaz, C., Borodina, T., & Himmelbauer, H. (2008). Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic acids research, 36(16), e105. https://doi.org/10.1093/nar/gkn425
39. Meacham, F., Boffelli, D., Dhahbi, J., Martin, D. I., Singer, M., & Pachter, L. (2011). Identification and correction of systematic error in high-throughput sequence data. BMC bioinformatics, 12, 451. https://doi.org/10.1186/1471-2105-12-451
40. 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]
41. 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]
42. 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
43. 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
44. 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

45. Maurier, F., Beury, D., Fléchon, L., Varré, J. S., Touzet, H., Goffard, A., Hot, D., & Caboche, S. (2019). A complete protocol for whole-genome sequencing of virus from clinical samples: Application to coronavirus OC43. Virology, 531, 141–148. https://doi.org/10.1016/j.virol.2019.03.006
46. Singh, L., San, J. E., Tegally, H., Brzoska, P. M., Anyaneji, U. J., Wilkinson, E., Clark, L., Giandhari, J., Pillay, S., Lessells, R. J., Martin, D. P., Furtado, M., Kiran, A. M., & de Oliveira, T. (2022). Targeted Sanger sequencing to recover key mutations in SARS-CoV-2 variant genome assemblies produced by next-generation sequencing. Microbial genomics, 8(3), 10.1099/mgen.0.000774. https://doi.org/10.1099/mgen.0.000774
47. Ou, J., Lan, W., Wu, X., Zhao, T., Duan, B., Yang, P., Ren, Y., Quan, L., Zhao, W., Seto, D., Chodosh, J., Luo, Z., Wu, J., & Zhang, Q. (2022). Tracking SARS-CoV-2 Omicron diverse spike gene mutations identifies multiple inter-variant recombination events. Signal transduction and targeted therapy, 7(1), 138. https://doi.org/10.1038/s41392-022-00992-2
48. Vatteroni ML, Capria A-L, Spezia PG, Frateschi S, Pistello M. Co-infection with SARS-CoV-2 omicron BA.1 and BA.2 subvariants in a non-vaccinated woman. Lancet Microbe 2022; published online May 24. https://doi.org/10.1016/S2666-5247(22)00119-7