*Article*

# A Routine Sanger Sequencing Target Specific Mutation Assay For SARS-CoV-2 Variants Of Concern And Interest

**Sin Hang Lee ¹\***

1 Affiliation
Milford Molecular Diagnostics Laboratory, 2044 Bridgeport Avenue, Milford, CT 06460, U.S.A.

\* Correspondence: shlee01@snet.net

**Abstract:** As SARS-CoV-2 continues to spread among human populations, genetic changes occur and accumulate in the circulating virus. Some of these genetic changes have caused amino acid mutations, including deletions, which may have potential impact on critical SARS-CoV-2 countermeasures, including vaccines, therapeutics, and diagnostics. Considerable efforts have been made to categorize the amino acid mutations of the angiotensin-converting enzyme 2 (ACE2) receptor binding domain (RBD) of the spike (S) protein along with certain mutations in other regions within the S protein as specific variants in an attempt to study the relationship between these mutations and the biological behavior of the virus. However, the currently used whole genome sequencing surveillance technologies can test only a small fraction of the positive specimens with high viral loads and often generate uncertainties in nucleic acid sequencing that needs additional verification for precision determination of mutations. This article introduces a generic protocol to routinely sequence a 437-bp nested RT-PCR cDNA amplicon of the ACE2 RBD and a 490-bp nested RT-PCR cDNA amplicon of the N-terminal domain (NTD) of the S gene for detection of the amino acid mutations needed for accurate determination of all variants of concern and variants of interest according to the definitions published by the U.S. Centers for Disease Control and Prevention. This protocol was able to amplify both nucleic acid targets into cDNA amplicons to be used as templates for Sanger sequencing on all 16 clinical specimens that were positive for SARS-CoV-2.

**Keywords:** Delta variant; Variants of concern; Variants of interest; SARS-CoV-2; Spike protein; Nested RT-PCR; Sanger sequencing; Amino acid mutations; ACE2 RBD; N-terminal domain (NTD)

## 1. Introduction

The COVID-19 crisis has continued its pace. According to Real Time World Statistics, as of September 7, 2021, there were > 222 million cumulative human cases with > 4.5 million deaths due to COVID-19 since the outbreak [1]. In the meantime, numerous amino acid mutations of the spike (S) protein of the SARS-CoV-2, the causative agent of COVID-19, are being recognized as whole genome sequencing data generated by the next generation sequencing (NGS) technology have been used more widely for genomic surveillance. Great efforts have been made to categorize these S protein amino acid mutations or substitutions into specific groups, according to their combination profiles. A few of these groups are referred to as variants in an attempt to correlate these amino acid mutation profiles with possible increased transmissibility, increased virulence, or reduced effectiveness of vaccines against them [2, 3].

The U.S. Centers for Disease Control and Prevention (CDC) has selected four variants of concern, namely the Alpha, Beta, Delta and Gamma variants, to be closely monitored for their potential impact on critical SARS-CoV-2 countermeasures, including vaccines, therapeutics, and diagnostics. In addition, four variants of interest, namely the Eta, Iota, Kappa and Pango Lineage B.1.617.3 variant, are being monitored and characterized [2].

As reported in mass media, the SARS-CoV-2 Delta variant has spread around the world [4] and is becoming the variant of most concern [5]. However, the science and procedures of how to accurately test for Delta variant and to differentiate it from other variants for meaningful data analyses remain unclear. The CDC's definition for Delta variant depends on demonstration of a specific profile of amino acid mutations listed as T19R, (V70F\*), T95I, G142D, E156 del, F157del, R158G, (A222V\*), (W258L\*), L452R, T478K, D614G, P681R and D950N with possible additional K417N. When K417N is also detected, the variant is designated as Delta Plus [2]. The amino acid mutations with a \* in parenthesis (\*) indicate that the mutation of this particular amino acid may or may not occur, and will not affect a

Delta variant designation. The amino acid mutations not in a (*) are invariable mutations, which must be present in a Delta variant. But it is not known if all laboratories performing variant testing are generating unambiguous sequencing data to verify all these 15 potential amino acid mutations before making a diagnosis of Delta variant of SARS-CoV-2. A university laboratory director in California was quoted as claiming that an L452R mutation is often a telling sign and that about 94% of the samples analyzed by his laboratory that show that mutation are proven to be Delta [6]. "Right now, we are assuming any new case is Delta given the high probability", reportedly declared by an infectious disease specialist at the University of California in San Francisco [7]. It is generally believed that there is a "Lack of testing" for Delta variant and that "without adequate data, policymakers are just swinging in the dark," as stated by a clinical professor of population and public health sciences at the University of Southern California [8]. Therefore, there is an urgent need for a science-based routine testing method for accurate detection of the key S protein amino acid mutations on all samples positive for SARS-CoV-2 so that the Delta and other variants of concern or of interest can be properly and consistently identified for further analyses.

When RNA viruses are allowed to transmit from population to population, genetic change invariably occurs due to RNA polymerase copying errors, which may lead to single nucleotide nonsynonymous mutations and indel mutations. The wildtype Wuhan-Hu-1 SARS-CoV-2 spike protein has 1,273 amino acids encoded by a 3,822-base S gene. But as of August 23, 2021, the number of S protein amino acid mutations reported worldwide already reached 2,860 [9]. Even randomly mixing a small fraction of these mutations will result in an enormous number of combination profiles. Therefore, as a matter of necessity the CDC can only select the most prevalent profiles, for example, the mutation combinations listed in the definitions of variants of concern and interest, for analyses. However, in order to find a common denominator for meaningful statistical analyses, laboratories still need to follow a single, stringent, but easily achievable set of criteria to define each variant to be included for the analyses and there should not be any bias in selecting positive samples for variant determination in order to generate truly comparable data.

The currently used whole genome/NGS surveillance technology for SARS-CoV-2 genomic sequencing is not always successful, particularly when there is not enough viral load in the specimen [10, 11]. And the NGS technology is also known to be associated with computational errors and biases in base-calling [12]. Target specific mutation assays are needed to identify variants of concern [11] and for their differentiation from other variants.

Based on information available in the public domain, even the profiles used to identify Delta variant vary from laboratory to laboratory. According to the World Health Organization (WHO), the characteristic S protein amino acid mutations for Delta variant, also known as the B.1.617.2 lineage, are T19R, G142D, 157del, 158del, L452R, T478K, D614G, P681R and D950N [13]. Although both the WHO and CDC emphasize the presence of T478K mutation in the Delta variant, a search of the GenBank database failed to find a SARS-CoV-2 genomic sequence containing a combination of G142D, 157del, 158del, L452R and T478K with an intact E156 in the S gene. On the other hand, Public Health England has advised that P681R must be present in a Delta variant, but also stated that genotyping assay for B.1.617 cannot distinguish between Kappa, Delta and B.1.617.3 lineage and all results with P681R are treated as probable Delta given the current dominance of this lineage [14]. Therefore, there is a need to develop a uniform practical protocol for precise variant classification on all SARS-CoV-2 isolates.

This article introduces a simplified routine target amplicon sequencing assay on all nasopharyngeal swab samples, which are positive for SARS-CoV-2, to accurately determine the key S protein amino acid mutations that are used to characterize variants of concern and of interest according to the CDC's definitions.

## 2. Materials and Methods

SARS-CoV-2 S protein amino acids mutate constantly and unpredictably when the virus is allowed to continue transmitting from host to host. GISAID automatically updates its site of hCoV-19 spike glycoprotein mutation surveillance dashboard. The updates include spike protein changes in amino acid sequences of the ACE2 receptor binding domain (RBD) newly submitted to GISAID, displayed in structures organized by the most common clades. The August 24/25, 2021 dashboard data showed the new clades (Figure 1), all which contain mutations commonly used for Delta variant categorization.

Figure 1. This is a figure showing GISAID hCoV-19 S protein mutation surveillance dashboard data on August 24 and 25, 2021

**8-24-2021 Dashboard**

N501T E484Q K417N N501Y E484K N439K T478K S494P F490S S477N
in GH clade (B.1.*)
E484K T478K K417T S494P F490S S477N N501Y
in GR clade (B.1.1.1 & B.1.1.7)
N501T E484Q K417N E484K K417E T478K S494P G446V S477I Q498R
in G, GK & GV clades (B.1, B.1.617.2, AY.* & B.1.177)
E484K T478K K417T N501Y
in nonG clade (A, B & B.2)

**8-25-2021 Dashboard**

E484K N439K K417N T478K S494P S477N N501Y
in GH clade (B.1.*)
K417N K417T L455F N501Y E484K G447V F490I K417A S494P F490S
in GR clade (B.1.1.1 & B.1.1.7)
V445F E484Q K417N A475S Q493E R403T S494L A475V F456L G485V F490S G446V S477N
S477I F490L N501T P499L K417T G446A N501Y V445I L455F E484K E484V K458N Q493L
T478K S494P P499S
in G, GK & GV clades (B.1, B.1.617.2, AY.* & B.1.177)
T478K S477I
in nonG clade (A, B & B.2)

While the GISAID hCoV-19 S protein mutation surveillance focuses on the ACE2 RBD mutations, some researchers have pointed out that the Delta variant has several unique mutations in the ACE2 RBD and the N-terminal domain (NTD) of the spike protein. The mutations in the NTD, such as T19R, G142D, E156G, F157del and R158del are involved in the enhanced infectivity by the BNT162b2-immune sera. The neutralizing activity of sera from vaccinated individuals as well as convalescent COVID-19 patients decreases for the Delta variant compared to the wild-type SARS-CoV-2 [15-17]. Both ACE2 RBD and NTD mutations should be evaluated on all positive samples to understand the pathogenicity of the SARS-CoV-2 variants. The CDC's classifications and definitions of SARS-CoV-2 variants of interest and concern, including mutations in the NTD, are summarized in Figure 2.

Figure 2     This is a figure showing CDC's definitions of SARS-CoV-2 variants of interest and concern



In Figure 2, the WHO variant labels are presented in boldfaced Greek letter names on the far left, followed by their distinctive specific spike protein amino acid mutations, the Pango lineage and the locations of first identification [2]. The mutations in the ACE2 RBD are highlighted yellow and those in the NTD highlighted green.

Sanger sequencing is the technology used to verify genomic nucleotide sequences, the analyte of all nucleic acid amplification tests. The currently widely used whole genome/NGS technology is an emerging, not yet stable technology for general use in disease diagnosis. There is a strong opinion within the EuroGentest and the European Society of Human Genetics that for genes that are responsible for a significant proportion of the defects, the sensitivity should not be compromised by the transition from Sanger to NGS [18]. In addition, there is a high percentage of uncertainties of base calls associated with computational errors and biases in NGS [12]. While the NGS technique is widely applied, varying error rates have been observed [19]. The first genomic sequences of SARS-CoV-2 isolates

from patient specimens in China [20] and in the United States [21] were verified by Sanger sequencing to avoid base-calling errors. Since specific variant classification is based on certain key amino acid mutations in the S protein, which in turn depend on accurate determination of SNPs and indel mutations of the S gene sequence, Sanger sequencing is the method of choice if the information derived from variant testing is used to influence patient management and policy makings.

Sanger sequencing needs a properly prepared template, which is usually a PCR amplicon of the target nucleic acid, for example, a segment of the S protein gene.   In molecular diagnostics, the size of PCR amplicon of the target DNA or cDNA is usually <450 bp. Attempts to amplify big-sized templates in complex samples often lead to PCR failures [22]. It is technically impossible to amplify the entire 3,822-base S gene as one single amplicon to be used as Sanger sequencing template. 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 outbreak [23]. The CDC's standard diagnostic protocol for SARS 2003 recommended using 3 SARS-CoV specific primers to perform heminested PCR and to sequence the 348-bp heminested PCR amplicon "to verify the authenticity of the amplified product" [24]. With accurate diagnosis, prompt isolation of patients and early treatment, the SARS 2003 outbreak ended in June with 8,098 reported cases and 774 deaths worldwide [25] without a variant of concern reported. The method presented in this article followed the CDC's established SARS 2003 protocols [22, 24] to sequence two ~400-base segments of the S gene of SARS-CoV-2 for accurate determination of SNP and indel mutations, which are used to determine amino acid changes to further define variants.

## 2.1. Using amino acid mutations in ACE2 RBD and NTD for variant determination

Both the WHO and the CDC definitions of SARS-CoV-2 variants of concern and interest depend on the specific profiles of amino acid mutations from K417 to N501 in the ACE2 RBD, supplemented by mutations in other regions of the 1,273 amino acid chain of the spike protein [2, 13].

Based on information retrieved from the GenBank database, a sequence of 116 amino acids from T393 to Y508 highlighted yellow in Figure 3 contains the entire ACE2 RBD from K417 to N501. A sequence of 160 amino acids from M1 to Y160 highlighted green in Figure 3 covers the entire NTD whose mutations are used as additional characteristics for variant categorization [2]. Assuming the classification algorithms defined by the CDC (Figure2) to be valid and stringent, accurate determination of the mutations of the amino acids from S45 to R158 and from positions K417 to N501 should be adequate for variant categorization.

Figure 3    This is a figure showing the first 508 amino acids of SARS-CoV-2 S protein with highlighted NTD M1 to Y160 and ACE2 RBD



Figure 3 shows the sequence of the first 508 amino acids of the SARS-CoV-2 Wuhan-Hu-1 S protein (retrieved from the GenBank database-Seq ID# NC_045512.2). The ACE2 RBD region is highlighted yellow and the NTD green. The common amino acids whose mutations (Figures 1 and 2) are used for variant determination are typed in red.

## 2.2. Patient samples used for method development

The materials used for method development were residues of 16 nasopharyngeal swab specimens from patients with clinical respiratory infections. These were previously tested patient specimens without patient identifications and were purchased from Boca Biolistics Reference Laboratory, Pompano Beach, FL, a commercial reference material

laboratory endorsed by the U.S. Food and Drug Administration (FDA) as a supplier of clinical samples positive for SARS-CoV-2 by RT-qPCR assays. According to the commercial supplier, the swabs were immersed in VTM after collection and stored in freezer at -80°C temperature following the initial testing.

In the author's laboratory, these 16 swab rinse specimens were proven to contain SARS-CoV-2 genomic RNA by successful bi-directional Sanger sequencing of a 398-bp N gene cDNA PCR amplicon. These 16 sequencing-confirmed positive samples were among the 30 specimens that were purchased and were initially classified as positive by RT-qPCR tests granted emergency use authorization by the FDA for the presumptive qualitative detection of nucleic acid from the 2019-nCoV [26]. The general characteristics of these 30 swab specimens were previously published in detail elsewhere [27]. According to the commercial supplier, all these samples were re-tested with an EUA N gene RT-qPCR assay, showing a Ct value ranging from 14.55 to 36.71. Nevertheless, only 16 of the 30 samples were shown to contain SARS-CoV-2 genomic RNA by partial N gene sequencing [27].

### 2.3. RNA extraction from nasopharyngeal swab specimens

Instead of cell-free fluid samples, which are used for most RT-qPCR assays, cellular components are routinely included in the material being tested in this assay [28]. The initially published protocol was slightly modified. Briefly, about 1 mL of the residues of the nasopharyngeal swab rinse in VTM was transferred to a graduated 1.5 mL microcentrifuge tube and centrifuged at ~16,000× g for 5 min to pellet all cells and cellular debris. The supernatant was discarded except the last 0.2 mL, which was left in the test tube with the pellet. To each test tube containing the pellet with 0.2 mL supernatant, 200 μL of digestion buffer containing 1% sodium dodecyl sulfate, 20mM Tris-HCl (pH 7.6), 0.2M NaCl and 700 μg/mL proteinase K, was added. The mixture was digested for 1 hr in a heated shaker set at 47°C. After digestion, an equal volume (400 μL) of acidified 125:24:1 phenol:chloroform:isoamyl alcohol mixture (Thermo Fisher Scientific Inc.) was added to each tube. After vortexing twice for extraction and centrifugation at ~16,000× g for 5 min to separate the phases, the liquid in the phenol/chloroform phase was pipetted out and discarded. To the remaining aqueous phase solution 300 μL of acidified 125:24:1 phenol:chloroform:isoamyl alcohol mixture was added for a second extraction. After a second centrifugation at ~16,000× g for 5 min to separate the phases, 200 μL of the aqueous supernatant without any material at the interface was transferred to a new 1.5 mL microcentrifuge tube.

To the 200 μL of phenol/chloroform-extracted aqueous solution, 20 μL of 3M sodium acetate (pH 5.2) and 570 μL of 95% ethanol were added. The mixture was placed into a cold metal block in a -15 to -20°C freezer for 20 min. The precipitated nucleic acids were centrifuged at ~16,000× g for 5 min and washed with 700 μL of cold 70% ethanol.

After a final centrifugation at ~16,000× g for 5 min, the 70% ethanol was completely removed with a fine-tip pipette, and the microcentrifuge tubes with opened caps were put into a vacuum chamber for 10 minutes to evaporate the residual ethanol. The nucleic acids in each tube were dissolved in 50 μL of UltraPure™ DEPC-Treated Water (ThermoFisher Scientific). All samples were tested immediately or stored at -80°C until testing.

### 2.4. PCR primers

As reported by the CDC, nested PCR is the necessary step to generate SARS-CoV-2 cDNA amplicons to be used as the templates for Sanger sequencing [21]. Three sets of primary and nested PCR primers were used to generate various amplicons for Sanger sequencing. The sequences, the sizes of the amplicons and the reference positions of the major primers used in this study are listed in Figure 4.

Figure 4        This is a figure showing sequences of 3 sets of primary and nested PCR primers and their positions in reference Seq # NC_045512.2

```
SB5 5'-AACCAGAACTCAATTACCCCC Primary Forw              21619-21639
SB6 5'-TTTGAAATTACCCTGTTTTCC Primary Rev    505 bp    22103-22123
SB7 5'-TCAATTACCCCCTGCATACAC Nested  Forw              21628-21648
SB8 5'-ATTACCCTGTTTTCCTTCAAG Nested  Rev    490 bp    22097-22117


SS1 5'-TGTGTTGCTGATTATTCTGTC Primary Forw              22643-22663
SS2 5'-AAAGTACTACTACTCTGTATG Primary Rev    460 bp    23082-23102
SS3 5'-ATTCTGTCCTATATAATTCCG Nested  Forw              22656-22676
SS4 5'-TACTCTGTATGGTTGGTAACC Nested  Rev    437 bp    23072-23092

VF1 5'-AATCATTACTACAGACAACAC Primary Forw              24901-24921
VF2 5'-CAATCAAGCCAGCTATAAAAC Primary Rev    338 bp    25218-25238
VF3 5'-AGACAACACATTTGTGTCTGG Nested  Forw              24913-24933
VF4 5'-GCTATAAAACCTAGCCAAATG Nested  Rev    315 bp    25207-25227
```

*2.5. PCR Conditions*

To initiate the primary RT-PCR, a total volume of 25 µL mixture was made in a PCR tube containing 20 µL of ready-to-use LoTemp® PCR mix with denaturing chemicals (HiFi DNA Tech, LLC, Trumbull, CT, USA), 1 µL (200 units) of Invitrogen SuperScript III Reverse Transcriptase, 1 µL (40 units) of Ambion™ RNase Inhibitor, 0.1 µL of Invitrogen 1 M DTT (dithiothreitol), 1 µL of 10 µmolar primary PCR forward primer in TE buffer, 1 µL of 10 µmolar primary PCR reverse primer in TE buffer and 1 µL of sample RNA extract, 1 µL of RNA extract from a patient sample previously tested positive for SARS-Cov-2 S gene as positive control, and 1 µL of M.B. grade water as negative control. 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.

To perform the nested PCR, a trace (about 0.2 µL) of primary PCR products was transferred by a micro-glass rod into a 25 µL volume of complete PCR mixture containing 20 µL of ready-to-use LoTemp® mix, 1 µL of 10 µmolar nested PCR forward primer, 1 µL of 10 µmolar nested PCR reverse primer and 3 µL of M.B. grade water. 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.

Aliquots of all primary and nested PCR products were analyzed by agarose gel electrophoresis. The nested PCR products yielding a band of expected size on the gel plate were subject to Sanger sequencing without further purification.

*2.6. DNA sequencing*

About 0.2 µL of the nested PCR products showing an expected band on agarose gel electrophoresis was transferred by a micro-glass rod into a Sanger reaction tube containing 1 µL of 10 µmolar sequencing primer (one of the nested PCR primers), 1 µL of BigDye® Terminator (v 1.1/Sequencing Standard Kit), 3.5 µL 5× buffer, and 14.5 µL M.B. grade 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 with a CentriSep column (Princeton Separations, Adelphia, NJ, USA), the reaction mixture was loaded in an Applied Biosystems SeqStudio Genetic Analyzer for sequence analysis. Sequence alignments were performed against the standard sequences stored in the GenBank database by on-line BLAST alignment analysis, and by visual comparison with the SARS-CoV-2 wild-type S gene sequence (Seq # NC_045512.2) for possible SNPs and indels.

## 3. Results

*3.1. The samples positive for N gene also contained an intact S gene*

Since mutations are widely scattered in the S protein amino acid chain among the variants (Figure 2) and PCR amplification of different specific segments of the 3,822-base S gene may be needed for Sanger sequencing and for differentiation of emerging variants, it is important to confirm that all RNA extracts from the clinical specimens, which were positive for an N gene segment [27], also contained an intact S gene. One way to achieve this goal

without performing an entire S gene sequencing was to use the SB7/SB8 nested PCR primer set to amplify a 490 bp cDNA at position 21628-22117 (Figure 4) and the VF3/VF4 nested PCR primer set to amplify a 315 bp cDNA at position 24913-25227 (Figure 4) on all 16 nasopharyngeal samples previously tested positive for the N gene. Representative parts of these two sequences from one sample are shown in Figure 5.

Figure 5    This is a figure showing two panels of sequencing electropherograms as evidence of intact S gene in a sample



5' CACTAATTCTTTCACACGTGGTGTTTATTACCCTGACAAAGTTTTCAGATCCTCAGTTTTATATTCAACTCA (21646-21717)

5' GTTGTAAACATTCAAAAAGAAATTGACCGCCTCAATGAGGTTGCCAAGAATTTAAATGAATCTCTCATCGATCTCCAAGAACTTGGAAAGTATGA GCAGTATATAAAATGGCCATGGTACATTTGGCTAGGTTTTATAGC  (25088-25227)|
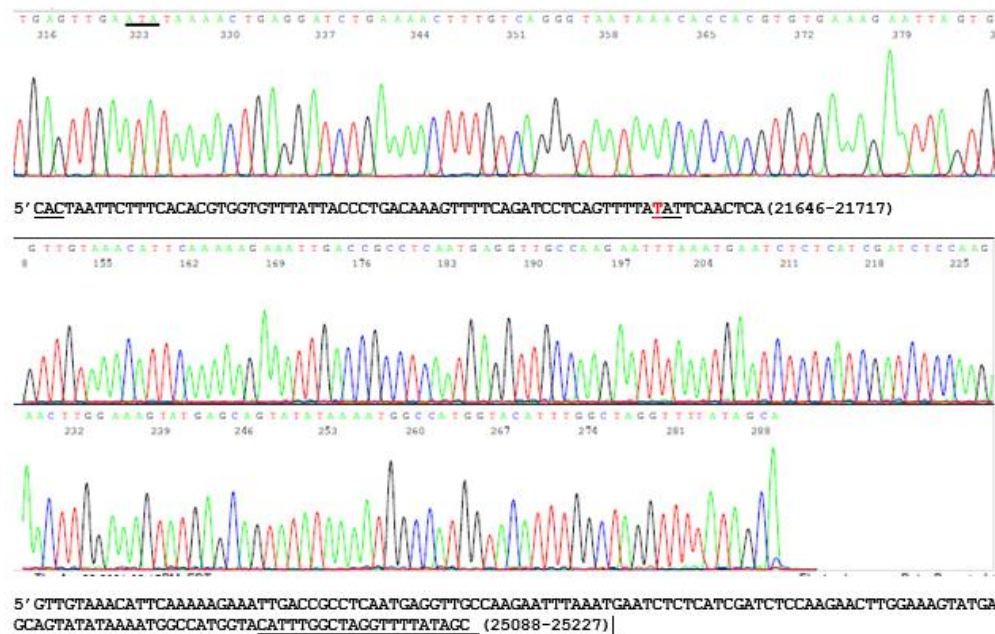
The upper panel of Figure 5 was excised from an electropherogram of a 490 bp amplicon sequence of the S gene defined by nested PCR primers SB7 and SB8 (Figure 4). The computer-generated sequence has been converted to 5'-3' reading that was re-typed under the upper electropherogram with the last 3 bases of the SB7 forward PCR primer "CAC" underlined. The number 21646-21717 indicates the position of this segment of sequence in the SARS-CoV-2 genome. The letter "T" in red means that the wild-type nucleotide in this position has undergone a nonsynonymous mutation causing an H49Y amino acid mutation (CAT>TAT). The lower panel was excised from an electropherogram of a 315 bp amplicon sequence of the S gene defined by nested PCR primers VF3 and VF4 (Figure 4). The computer-generated sequence in 5'-3' reading direction is re-typed under the electropherogram with the entire 21-base reverse PCR primer underlined. The number 25088-25227 indicates the position of this segment of sequence in the SARS-CoV-2 genome.

Since the two sequences illustrated in Figure 5 are >3,000 nucleotides apart within the S gene of SARS-CoV-2 genome, their presence in one sample supported the interpretation that the sample being tested contained an intact S gene and was suitable as the material for development of methods for S gene target specific mutation assays.

### 3.2. Limitations of the size of diagnostic RT-PCR amplicon

Initially, attempts were made to design primary and nested PCR primers to amplify a 1,524-base segment of the S gene encoding the first 508 amino acids of the SARS-CoV-2 spike protein (Figure 3), including the NTD and the ACE2 RBD in a single amplicon. It has been reported that under certain conditions the entire >1,500-base bacterial 16S rRNA gene can be amplified by PCR [29, 30]. However, all attempts failed. A single >1,500-bp S gene cDNA PCR amplicon could not be generated from the nasopharyngeal swab samples used for this study.

### 3.3. Target amplicon sequencing of the ACE2 RBD region

According to the CDC's definitions, all SARS-CoV-2 variants of concern and of interest contain at least one amino acid mutation in the S protein ACE2 RBD from K417 to N501 (Figure 2). But R403T has also appeared recently

at the GISAID hCoV-19 S protein mutation surveillance dashboard along with other mutations for emerging variant characterization [Figure 1]. Therefore, a diagnostic base-calling electropherogram must contain a 297-base unambiguous sequence covering 99 amino acid codons (nucleotide position 22769 -23065). For routine diagnostic convenience, these 297 bases must be present in one single computer-generated sequence on an electropherogram to confirm that the positive isolate is not a variant of concern or interest, or to provide mutation information for variant identification. To fulfill these requirements, a pair of SS1/SS2 primary RT-PCR primers and a pair of SS3/SS4 nested PCR primers (Figure 4) were selected to amplify a 460 bp primary PCR cDNA amplicon and a 437 bp nested PCR amplicon, respectively. These two pairs of primers were proven to be successful for amplification of a 437 bp nested PCR amplicon to be used as sequencing templates from all 16 samples proven to contain a segment of N gene sequence. One of these electropherograms showing the sequence encompassing the codons from R403 to N501 is presented in Figure 6.

Figure 6    This is a figure of computer-generated electropherogram showing the codons of 19 spike protein amino acids in the ACE2 RBD region commonly used to detect and define SARS-CoV-2 variants
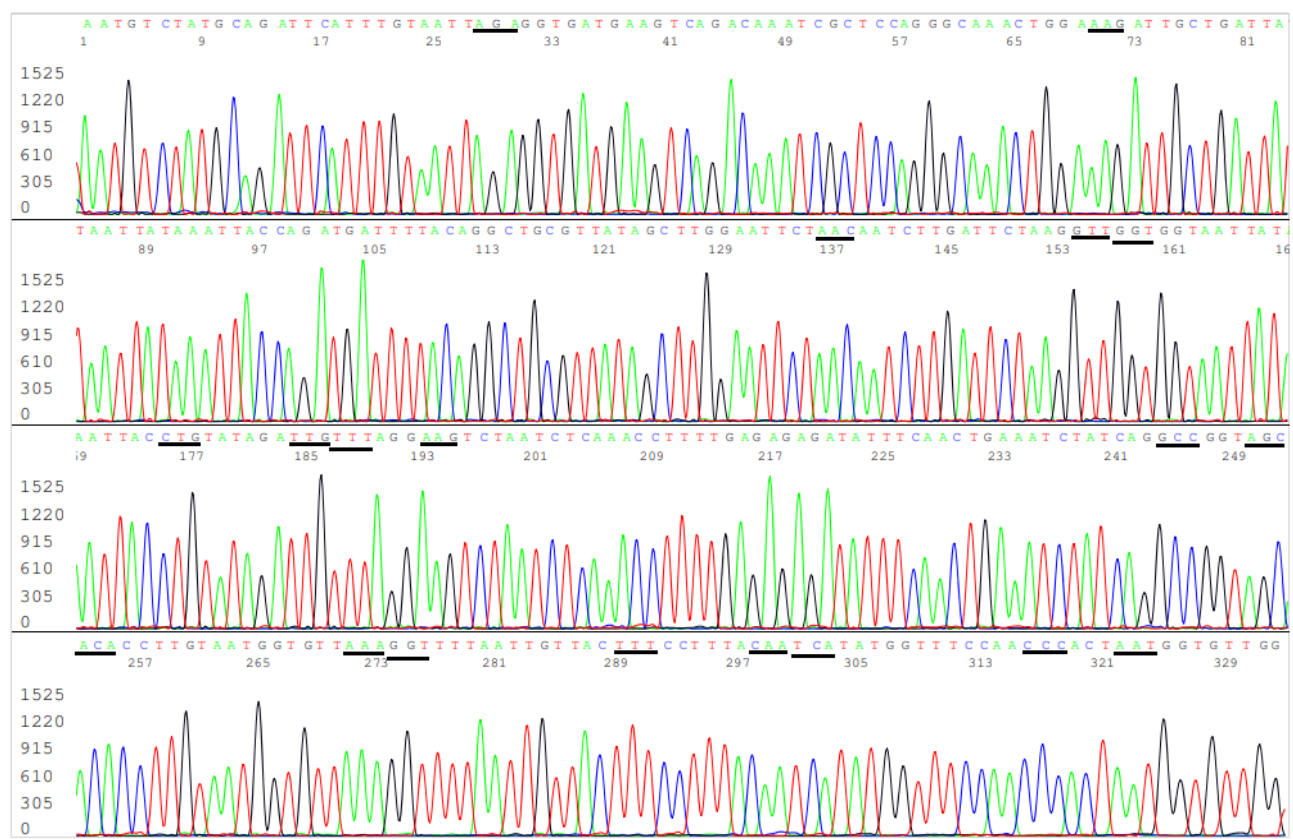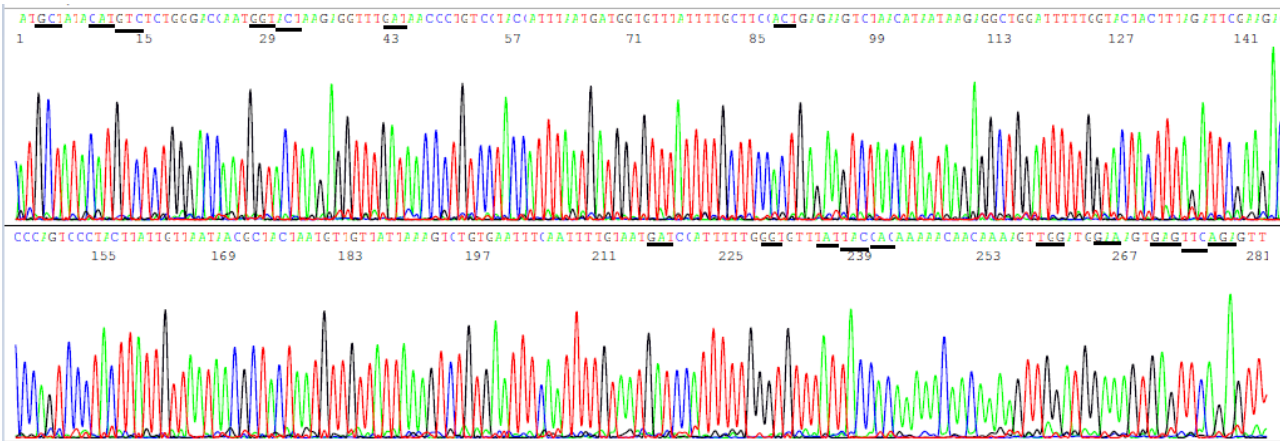


In Figure 6, the sequencing electropherogram shows 19 underlined codons of R403, K417, N439, V445, G446, L452, L455, F456, K458, A475, S477, T478, E484K (GAA>AAA mutation), G485, F490, Q493, S494, P499 and N501 in the ACE2 RBD region of the SARS-CoV-2 spike protein gene. Nonsynonymous mutations of the nucleotides in these 19 codons are routinely monitored for surveillance by GISAID (Figure 1).

*3.4. Target amplicon sequencing of the NTD region*

The amino acids in the NTD region used by the CDC to define variants spans from L5 to R158, a segment of 154 amino acids with a coding nucleic acid sequence of 462 bases. Attempts to generate a 569 bp nested PCR amplicon from the 16 clinical samples known to be positive for SARS-CoV-2 by partial N gene sequencing were successful in 9 samples only (9/16). By necessity, the sizes of the primary RT-PCR amplicon and the nested PCR amplicon were reduced to 505 bp and 490 bp, respectively, to gain PCR sensitivity while using the SB5/SB6 pair for the primary PCR primers and the SB7/SB8 pair as the nested PCR primers to generate a 490 bp amplicon (Figure 4) as the template for Sanger sequencing from all 16 samples. This 490 bp amplicon includes 17 key amino acids in a region from A67 to R158, i.e., 92 codons with a 276-base sequence. Mutations in these 17 amino acids are used by the CDC [2] to help

distinguish variants of concern and interest for surveillance. Since several deletions are involved in these mutation profiles and a bi-directional Sanger may be needed for verification of some of these deletions, the size of the NTD nested PCR amplicon is longer than necessary for a one-directional reading so that the key mutation sites are not placed too close to the PCR primer sites in case a bi-directional sequencing is needed to confirm a SNP or a deletion toward the 3' end of a nested PCR primer site. A typical computer-generated electropherogram showing the codons of the 17 amino acids in the NTD, which the CDC uses to help define variants, is presented in Figure 7.

Figure 7        This is a figure of computer-generated electropherogram showing 17 codons of S protein amino acids in the NTD region commonly used to help distinguish SARS-CoV-2 variants



In Figure 7, the sequencing electropherogram shows 17 underlined codons of amino acids, A67, H69, V70, G75, T76, D80, T95, D138, G142, Y144, Y145, H146, W152, E154, E156, F157 and R158 of the SARS-CoV-2 spike protein in the NTD, which may mutate in different variants of concern and of interest. These 17 amino acid mutations along with the ACE2 RBD amino acid mutations are used for variant categorization (Figure 2).

*3.5. Determination of variants by sequencing of the ACE2 RBD and NTD*

According to the CDC's definitions, all SARS-CoV-2 variants of concern or interest must have at least one amino acid substitution in the ACE2 RBD of the S protein (Figure 2). Assuming the CDC's variant classification algorithms to be valid and stringent, the profiles of the amino acid mutations in the NTD listed in Table 1 in combination of the mutations in the ACE2 RBD can be used for accurate variant categorization.

Table 1        Summary of using ACE2 RBD/NTD sequencing for variant categorization

| WHO Name Valiant | Pango lineage | ACE2 RBD mutations | NTD mutations | Location of first identification |
|---|---|---|---|---|
| Alpha | B.1.1.7 | N501Y | 69del, 70del, 144del | United Kingdom |
| Beta | B.1.351 | K417N, E484K, N501Y | D80A | South Africa |
| Gamma | P.1 | K417T, E484K, N501Y | D138Y | Japan/Brazil |
| Delta | B.1.617.2 | L452R, T478K | T95I, G142D, E156del, F157del, R158G | India |
| Delta plus | B.1.617.2.1 | K417N, L452R, T478K | T95I, G142D, E156del, F157del, R158G | India |
| Epsilon | B.1.427 | L452R | | California, USA |
| Epsilon | B.1.429 | L452R | W152C | California, USA |
| Eta | B.1.525 | E484K | A67V, 69del, 70del, 144del | U. K./Nigeria |
| Iota | B.1.526 | E484K | T95I | New York, USA |
| Kappa | B.1.Variant617.1 | L452R, E484Q | G142D, E154K | India |
| Kappa | B.1.617.3 | L452R, E484Q | G142D | India |
| Lambda | C.37 | L452Q, F490S | G75V, T76I | Peru |

## 4. Discussion

*4.1. Accurate categorization of VOCs and VOIs on all samples positive for SARS-CoV-2*

This article introduces a generic target specific mutation assay for accurate detection of variants of concern (VOCs) and variants of interest (VOIs) by sequencing two nested RT-PCR amplicons of the SARS-CoV-2 spike protein gene, one located in the ACE2 RBD and one in the NTD region. Since the sample being tested includes digestate of virus-infected cells, instead of cell-free fluid only [28], more viral genome copies are available for testing in this assay as compared to other commercial assays. In addition, the nested RT-PCR technology routinely amplifies the target nucleic acid for a total of 60 cycles to raise detection sensitivity. Therefore, this target specific mutation assay can determine the amino acid mutations accurately in samples with low viral loads when the whole genome/NGS surveillance technologies fail.

Traditionally, the CDC recommends sequencing of a 348-bp nested RT-PCR amplicon to verify the authenticity of the amplified product [23, 24] in molecular testing for SARS-CoV. In molecular diagnostics, PCR amplification of a target DNA > 400 bases long may risk loss of test sensitivity because human nasopharyngeal swab specimens invariably contain PCR inhibitors of unpredictable kinds and quantities. PCR inhibitors co-extracted from the complex clinical specimens tend to interfere with enzymatic primer extension. But in the assay recommended in this article the RNA samples for RT-PCR are prepared by phenol/chloroform extraction, which has removed most of the PCR inhibitors carried over from the clinical specimens. Therefore, a > 400-base RNA segment can be readily amplified to be used as the sequencing template. Phenol-chloroform has been shown to be a $10^6$ times more sensitive extraction method than the popular commercial QIAamp blood kit in detection of HBV DNA in serum samples [31].

Any meaningful correlative analysis linking variants to clinical and epidemiological data must be based on precise determination of S protein amino acid mutations, which are the basis for variant categorization. Variant testing must be conducted routinely on all samples positive for SARS-CoV2. The current surveillance programs select less than 5% of the positive samples with high viral loads for variant testing by whole genome/NGS; it is generating highly biased and potentially misleading information based on which public policy with impacts on society and economy is made. A high SARS-CoV-2 viral load in a clinical specimen is not invariably associated with disease severity [32].

### 4.2. *The current confusion of Delta variant testing*

When RNA viruses are allowed to pass from host to host, the viruses invariably acquire genome mutations due to RNA polymerase copying errors. Obviously, only those mutations that can be passed down to descendant viruses in subsequently infected individuals can be observed, documented, and reported in the literature [33]. The WHO and the U.S. CDC have selected the mutations of 8 amino acids, namely the K417, L452, S477, T478, E484, F490, S494 and N501, in the spike protein ACE2 RBD as the key mutations to create a limited number of VOCs and VOIs for surveillance purpose (Figure 2). The WHO and CDC seem to advise that the absence of mutations in these 8 amino acids rules out VOCs or VOIs although such advice has not been clearly stated on record.

As SARS-CoV-2 continues to spread, more new amino acid mutations in the ACE2 RBD have been accumulated to the circulating strains and reported to GISAID (Figure 1). These new profiles of mutations may involve some of the 8 amino acids used to define the VOCs and VOIs. Some of these new profiles may contain a mixture of mutations each of which is considered unique for a specific variant, such as the E484K, T478K, K417T and N501Y (Figure 1 and Figure 2). It is not clear if these new profiles are considered to be VOCs or as Delta variants if there is a T478K mutation.

According to the official classification algorithms, the T478K mutation in the RBD is unique to the Delta variant; by definition the spike protein of Delta variant contains eight mutations, including four mutations in the NTD (T19R, G142D, 156–157del, and R158G), two in the RBD (L452R and T478K), one mutation close to the furin-cleavage site (P681R) and one in the S2 region (D950N) [2, 34]. The Delta variant has been reported to become the most dominant SARS-CoV-2 worldwide in the summer of 2021. In the United States during the week of 8/22/2021– 8/28/2021, 99.1% of the SARS-CoV-2 isolates were classified as Delta variants [35].

However, the WHO's definition for Delta variant is a profile of T19R, G142D, 157del and 158del in the NTD plus L452R and T478K in the ACE2 RBD [13] and the Public Health England uses P681R as the key mutation to define Delta variant [14]. It is not clear which classification algorithm is being used to define Delta variants in different parts of the world. In the United Stated, some specialists simply assumed "any new case is Delta" [7]. Since the

whole genome/NGS surveillance technology tends to generate uncertainties of base calls associated with computational errors and biases [12], it is not known how many of the new cases have been erroneously classified as Delta variants as a result of computational errors and biases. Based on information available in the public domain, the sequence data for variant surveillance have not been verified by Sanger sequencing as stringently as those used to identify the initial Wuhan-Hu-1 SARS-CoV-2 strain [20, 21].

*4.3. Accumulation of mutations in viruses is a function of passages*

When RNA viruses are subjected to passages as in serial culture transfers, accumulation of mutations will occur [36]. The same biological process takes place among humans in the current COVID-19 pandemic.

During the 2003 SARS outbreak, the CDC instituted 3 SARS-CoV specific RT-PCR primers to amplify a 348-bp PCR product for sequencing confirmation of molecular detection of SARS-CoV [24]. The accurately diagnosed SARS patients were isolated properly and treated promptly. The outbreak in 2003 ended in June. There was no SARS-CoV variant of concern in the 2003 outbreak because the epidemic ended too soon for accumulation of a significant number of mutations in the circulating viruses.

During the 2020 COVID-19 outbreak, it took 11 months for the first variant of concern, an Alpha variant of SARS-CoV-2, to develop and to be isolated from a 58-year-old human male on November 24, 2020 in England, United Kingdom [37]. Accumulation of amino acid mutations and emerging of SARS-CoV-2 variants of concern were probably the result of uncontrolled transmission of the RNA virus among populations.

For example, E484K is the unique mutation in combination with K417N and N501Y in the ACE2 RBD that is used to define the Beta variant, the so-called South Africa variant, first reported in December 2020 [38]. However, a search of the SARS-CoV-2 genomic sequence database in the GenBank revealed that a solitary E484K mutation in the ACE2 RBD without a concomitant K417N or N501Y was already reported to the GenBank from a specimen collected on March 15, 2020 in Utah, USA (Sequence ID: MW190617), from a specimen collected on May 22, 2020 in Illinois (Sequence ID: MT772530), from a specimen collected on August 17, 2020 in Utah (Sequence ID: MW420419), from a specimen collected on October 28, 2020 in Minnesota (Sequence ID: MW349062), and from two specimens collected in November, 2020 in Virginia (Sequence ID: MW338781 and Sequence ID: MW411871). The same solitary E484K might have been in the SARS-CoV-2 strains circulating unknowingly in South Africa before December, 2020. Some of the "Beta variant" isolates might have been locally developed in the United States.

A search of the SARS-CoV-2 genomic sequence database in the GenBank also revealed that a solitary T478K mutation without concomitant mutations in the ACE2 RBD or the NTD had been reported to the GenBank on numerous occasions from different states before the summer of 2021, for example, in a specimen collected on January 13, 2021 in Utah (Sequence ID: MZ490259), in a specimen collected on March 22, 2021 in Arizona (Sequence ID: MZ914771), in a specimen collected on April 24, 2021 in Alaska Interior (Sequence ID: MZ643206), and in a specimen collected on May 6, 2021 in Alaska Anchorage (Sequence ID: MZ927507). In addition to the solitary T478K mutation in the ACE2 RBD with a wild-type NTD sequence and a wild-type D950, the sequences mentioned above also contain a P681H mutation instead of a P681R that is used to define Delta variant. It is not clear if these isolates are being classified as Delta variant. They are certainly not the descendent of the Delta variant originating in India. According to the currently accepted classification algorithms, P681H only occurs in Alpha variant (Figure 2).

The GenBank database contains numerous SARS-CoV-2 spike protein amino acid mutation profiles, which may be mistaken as Delta variant if stringent variant classification algorithm is not followed. A few potential sequence profiles that can be mistaken for a Delta variant are listed as follows:

- L452R and T478K without concomitant mutations in the NTD -Sequence ID: MZ637393.

- E156del, F157del, R158G without concomitant mutations in the ACE2 RBD-Sequence ID: MZ340544.

- G142D, E156del, F157del, R158G without concomitant mutations in the ACE2 RBD-Sequence ID: MZ341068.

- T95I, G142D, E156del, F157del, R158G, E484K-Sequence ID: MZ531409.

- T95I, L452R-Sequence ID: MZ086521

*4.4. The Delta variant scare is not supported by facts*

Currently, there is a coronavirus Delta variant scare being generated in the United States to the point that the created public anxiety may have a negative impact on the U.S. economic recovery from the pandemic [39] although even the CDC does not know exactly how many U.S. coronavirus deaths are attributable to the Delta variant infections [40]. Nevertheless, according to the data published up to 5 July 2021 by Public Health England the system recorded a total number of 170,063 cases of Delta variant infection and 259 deaths among this group of patients [41] with a mortality rate of 0.15%. In the same document, there were 225,864 cases with Alpha variant infection and 4,264 deaths in the same group with a mortality rate of 1.89%. So, the Alpha variant is at least 10 times more deadly than the Delta variant.

For comparison, the Chinese data show that up to March 3, 2020 before any variants of concern emerged, there were 80,270 confirmed COVID-19 cases with 2,981 deaths in China, most of which were from the epicenter of the outbreak [42]. The mortality rate of the wild-type Wuhan-Hu-1 SARS-CoV-2 infections is 2981/80270=3.71%, which is about twice as high as the mortality rate of Alpha variant infections.

Therefore, the Delta variant is not more dangerous than the wild-type Wuhan-Hu-1 strain or the Alpha variant. The high number of Delta variants being reported in the literature may have resulted from over-extrapolation bias based on sequencing of a very limited number of specially selected samples with surveillance testing methods of uncertain accuracy in unregulated laboratories. Generally, surveillance testing using sequencing technology to identify SARS-CoV-2 genetic variants can be performed in a facility that is NOT CLIA certified, provided that patient-specific results are not reported to (1) the individual who was tested or (2) their health care provider [43]. There are no quality control measures to identify potential flaws in coronavirus variant testing in the United States because the surveillance testing results are not for patient management even though they are being used as the basis for formulation of public health policies.

The high number of Delta variants being reported to the government for surveillance purpose may simply indicate that many SARS-Cov-2 strains with certain amino acid mutations described in the CDC' definition for Delta variant (Figure 2) have acquired a genetic profile that enables them to have a higher replication rate in the host than the others, but their pathogenicity may have been reduced to the level of that of the common human coronaviruses as those of types 229E, NL63, OC43, and HKU1[44]. This heterogeneous group of SARS-CoV-2 strains may have been detected more often because there are more virus copies in the samples being tested resulting from their higher replication rates. A higher rate of being detected does not necessarily indicate that the virus variant is more transmissible unless actual movement of the variant among close contacts has been studied by epidemiological tracing research supported by accurate variant testing. Transmissibility of a virus is primarily determined by the infectivity of the pathogen [45], not the viral load of the donor.

*4.5. Routine sequencing on all positive samples for variant determination*

According to the CDC update for the week ending August 28, 2021, the combined proportion of cases attributed to Delta variant is estimated to be greater than 99% in the United States. The national proportions of Alpha, Beta, and Gamma are estimated to be well below 0.5%. It is expected that Delta will continue to be the predominant circulating variant [46]. However, the 99% attribution to Delta is an estimate. Laboratories may use different profiles of amino acid mutations to define Delta variant. Some reports were based on assumptions only [7].

Notably, some researchers in the field use a profile of T19R, G142D, E156G, F157del, R158del, L452R, T478K, D614G, P681R and D950N in the spike protein to define Delta variant by following the GISAID database [47]. According to the latter system of classification, Delta variant lacks T95I, and has E156G and R158del [47], the two mutations that are not in line with the CDC's definition for Delta variant (Figure 2). A GenBank Sequence ID# OU534154 also lists an NTD/ACE2 RBD sequence containing G142D, E156G, F157del, R158del, L452R and T478K with neither T19R nor T95I in the NTD. Based on the various issues discussed above, the actual number of Delta variants categorized according to the stringent CDC definition is unknown. All statistics based on correlations between Delta variant and its biological characteristics are highly questionable because the SARS-CoV-2 isolates currently classified as

Delta variants may actually consist of numerous genetic variants. There is no common denominator for any meaningful comparative discussions.

In order to fully realize the potential of genomic epidemiology, there is a need for routine sequencing of viral nucleic acid established in parallel with COVID-19 testing [48], on all positive samples, including those with low viral loads. Even with high viral load samples, it took several months for the CDC to accurately verify the entire ~30,000-base sequence of a SARS-Cov-2 whole genome, using both the NGS and the nested PCR/Sanger sequencing technology [21]. Such an approach, even used to sequence the entire 3,822-base spike protein gene, is not practical in routine diagnostic works because the common RT-PCR amplicon size in SARS-CoV diagnostic testing is ~348 bp in size [24]. If a NGS technology is used for the diagnostic work, under certain circumstances it may need to sequence as many as 10 PCR amplicons to verify or to correct the base-calling uncertainties generated by the computational errors and biases of the NGS technology [12] in a gene target of 3,822 bases long among numerous non-target nucleic acids in a nasopharyngeal swab sample.

This article proposes routine sequencing of a 437-bp nested PCR cDNA amplicon of the S gene ACE2 RBD (Figure 6) on all samples that are positive for a SARS-CoV-2 RNA gene. If there is no amino acid mutation in the RBD, the SARS-CoV-2 detected is not a VOC or a VOI. If the RBD sequencing shows any amino acid mutations, an additional 490-bp nested PCR cDNA amplicon of the S gene NTD is sequenced (Figure 7). Since a properly executed computer-generated sequencing electropherogram does not have ambiguous base calls, the codons of the amino acids in the ACE2 RBD and in the NTD can be easily determined without the need of bioinformatic services.

Assuming the CDC's definitions based on amino acid mutations for variant determination to be valid and stringent, categorization of mutations of the amino acid codons underlined in Figures 6 and 7 can be used as the basis for accurate variant determination of all VOCs and VOIs currently recognized (Figure 2 and Table 1). Even the recently reported Mu valiant (PANGO lineage B.1.621), which is characterized by a combination of R346K, E484K, N501Y, D614G and P681H [49], can be distinguished from other VOCs and VOIs because there are no concomitant mutations in the NTD sequence in the presence of only E484K and N501Y in the ACE2 RBD sequence for this variant.

By the same token, a newly reported South Africa variant with PANGO lineage C.1.2, which contains multiple substitutions (R190S, D215G, N484K, N501Y, H655Y and T859N) and deletions (Y144del, L242-A243del) within the spike protein [50], can be distinguished from other VOCs and VOIs by demonstration of only N484K and N501Y in the ACE2 RBD and a Y144del in the NTD without other concomitant mutations in the two amplicons targeted for sequencing.

## 5. Conclusions

As SARS-CoV-2 is allowed to continue spreading through human populations, amino acid mutations occur and the mutations tend to accumulate in the circulating virus. When the mutations and deletions involve the amino acids of the S protein, especially those in the ACE2 RBD, the changes of the amino acid sequence may have potential impact on critical SARS-CoV-2 countermeasures, including vaccines, therapeutics, and diagnostics. Some of the amino acid mutations have been selectively grouped together as specific variants in an attempt to link these mutations and the biological behavior of the SARS-CoV-2. However, the whole genome sequencing technologies currently used for surveillance of a small fraction of the clinical specimens with high viral loads only cannot test specimens with low viral loads accurately, and may generate uncertainties in the DNA sequence, which is crucial for variant categorization. The protocol presented in this article is able to routinely sequence a 437-bp nested RT-PCR cDNA amplicon of the ACE2 RBD and a 490-bp nested RT-PCR cDNA amplicon of the N-terminal domain (NTD) of the S gene for detection of the amino acid mutations needed for accurate determination of all variants of concern and variants of interest in samples positive for SARS-CoV-2 regardless of their viral loads. In order to fully realize the potential of genomic epidemiology, there is a need for routine sequencing of viral nucleic acid established in parallel with COVID-19 testing on all positive samples, including those with low viral loads.

## Author Contributions

**S.H.L.: Conceptualization, study design, methodology, data analysis, manuscript preparation and final review**

## References

1. Coronavirus Updates. Worldometer. https://www.worldometers.info/coronavirus/ (accessed on 7 September, 2021).
2. CDC. SARS-CoV-2 Variant Classifications and Definitions. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html (accessed on 15 August, 2021).
3. WHO. Tracking SARS-CoV-2 variants. https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/ (accessed on 25 August, 2021).
4. Yen Nee Lee. CNBC. 3 charts show how far Covid delta variant has spread around the world. PUBLISHED THU, AUG 5 2021 https://www.cnbc.com/2021/08/06/covid-charts-show-how-far-delta-variant-has-spread-around-the-world.html (accessed on 7 September, 2021).
5. WHO. Episode #45 - Delta variant. 5 July 2021. Science conversation. https://www.who.int/emergencies/diseases/novel-coronavirus-2019/media-resources/science-in-5/episode-45---delta-variant (accessed on 7 September, 2021).
6. Angelo Fichera. Factcheck. Sequencing Used to Identify Delta, Other Coronavirus Variants. August 4, 2021. https://www.factcheck.org/2021/08/scicheck-sequencing-used-to-identify-delta-other-coronavirus-variants/ (accessed on 7 September, 2021).
7. Alix Martichoux, Nexstar Media Wire. How do you know if you have the delta variant of COVID-19? Aug 8, 2021. https://www.news10.com/news/how-do-you-know-if-you-have-the-delta-variant-of-covid-19/ (accessed on 7 September, 2021).
8. Denise Chow. NBC News. Where's the data on delta? Lack of testing, info makes it hard to see virus's full scope. Aug. 19, 2021 https://www.nbcnews.com/science/science-news/delta-variant-response-hindered-covid-test-limitations-lack-data-rcna1692 (accessed on 7 September, 2021).
9. China National Center for Bioinformation. 2019 Novel Coronavirus Resource (2019nCoVR). https://ngdc.cncb.ac.cn/ncov// (accessed on 23 August, 2021).
10. Australian Government Department of Health. Coronavirus Disease 2019 (COVID-19) CDNA National Guidelines for Public Health Units. 24 June 2021. https://www1.health.gov.au/internet/main/publishing.nsf/Content/7A8654A8CB144F5FCA2584F8001F91E2/$File/COVID-19-SoNG-v4.7.pdf (accessed on 20 August, 2021).
11. Buchan, B.W.; Wolk, D.M.; Yao, J.D. Rapid Communication. SARS-CoV-2 Variant Testing. Ver.1, released 4.28.2021. The Association for Molecular Pathology (AMP), Infectious Diseases Society of America (IDSA), and Pan American Society for Clinical Virology (PASCV). https://www.amp.org/AMP/assets/File/clinical-

practice/COVID/AMP_RC_VariantTestingforSARSCOV2_4_28_21.pdf?pass=91    (accessed on 7 September, 2021)

12. Abnizova, I.; te Boekhorst, R.; Orlov, Y. Computational Errors and Biases in Short Read Next Generation Sequencing. *J. Proteomics Bioinform.* **2017**, *10,* 1-17. doi: 10.4172/jpb.1000420      https://www.long-dom.org/open-access/computational-errors-and-biases-in-short-read-next-generationsequencing-jpb-1000420.pdf    (accessed on 7 September, 2021)

13. WHO. Coronavirus Disease (COVID-19): Weekly Epidemiological Update (11 May 2021) https://reliefweb.int/report/world/coronavirus-disease-covid-19-weekly-epidemiological-update-11-may-2021    (accessed on 25 August, 2021)

14. Public Health England. SARS-CoV-2 variants of concern and variants under investigation in England. Technical briefing 15. 11 June 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/993879/Variants_of_Concern_VOC_Technical_Briefing_15.pdf    (accessed on 25 August, 2021)

15. Liu, Y.; Arase, N.; Kishikawa, J.; Hirose, M.; Li, S.; et al. The SARS-CoV-2 Delta variant is poised to acquire complete resistance to wild-type spike vaccines. bioRxiv 2021.08.22.457114; doi: https://doi.org/10.1101/2021.08.22.457114    (accessed on 7 September, 2021)

16. Li, D.; Edwards, R.J.; Manne, K.; Martinez, D.R.; Schafer, A.; Alam, S.M.; et al. In vitro and in vivo functions of SARS-CoV-2 infection enhancing and neutralizing antibodies. *Cell* **2021**, *184*, 4203-4219 e4232.

17. Liu, C.; Ginn, H.M.; Dejnirattisai, W.; Supasa, P.; Wang, B.; Tuekprakhon, A.; et al. Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum. *Cell* **2021**, *184*, 4220-4236 e4213.

18. Matthijs, G.; Souche, E.; Alders, M.; Corveleyn, A.; Eck, S.; Feenstra, I.; Race, V.; Sistermans, E.; Sturm, M.; Weiss, M.; et al. Guidelines for diagnostic next-generation sequencing. *Eur. J. Hum. Genet.* **2016**, *24*, 2–5.

19. Pfeiffer, F.; Gröber, C.; Blank, M.; Händler, K.; Beyer, M.; Schultze, J.L.; Mayer, G. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Sci Rep.* **2018,** *8,* 10950. doi: 10.1038/s41598-018-29325-6. PMID: 30026539; PMCID: PMC6053417.

20. Ren, L.L.; Wang, Y.M.; Wu, Z.Q.; Xiang, Z.C.; Guo, L.; Xu, T.; et al. Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study. *Chin Med J (Engl).* **2020**, *133*, 1015-1024.

21. Harcourt, J.; Tamin, A.; Lu, X.; Kamili, S.; Sakthivel, S.K.; Murray, J.; et al. Severe Acute Respiratory Syndrome Coronavirus 2 from Patient with Coronavirus Disease, United States. *Emerg. Infect. Dis.* **2020**, *26,* 1266-1273.

22. 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.

23. 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.

24. CDC. SARS-CoV Specific RT-PCR Primers. https://www.who.int/csr/sars/CDCprimers.pdf?ua=1    (accessed on 7 September, 2021)

25. CDC. Severe Acute Respiratory Syndrome (SARS). https://www.cdc.gov/sars/about/faq.html    (accessed on 7 September, 2021)

26. FDA to CDC. Letter dated December 1, 2020.    https://www.fda.gov/media/134919/download    (accessed on 7 September, 2021)

27.  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 Studies & Opinions.* **2021**, *2*, 230–278.

28. 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.

29. Clarridge, J.E. 3rd. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. *Clin. Microbiol. Rev.* **2004**, *17*, 840-862.

30. Zeng, Y.H.; Koblížek, M.; Li, Y.X.; Liu, Y.P.; Feng, F.Y.; Ji, J.D.; et al.   Long PCR-RFLP of 16S-ITS-23S rRNA genes: a high-resolution molecular tool for bacterial genotyping. *J. Appl. Microbiol.* **2013**, *114*, 433-447.

31. Kramvis, A.; Bukofzer, S.; Kew, M.C. Comparison of hepatitis B virus DNA extractions from serum by the QI-Aamp blood kit, GeneReleaser, and the phenol-chloroform method. *J. Clin. Microbiol.* **1996**, *34*, 2731-2733.

32. Abdulrahman, A.; Mallah, S.I.; Alqahtani, M. COVID-19 viral load not associated with disease severity: findings from a retrospective cohort study. *BMC Infect. Dis.* **2021**, *21*, 688. doi: 10.1186/s12879-021-06376-1. PMID: 34271860; PMCID: PMC8284033.

33. Sanjuán, R.; Domingo-Calap, P. Mechanisms of viral mutation. *Cell Mol. Life Sci.* **2016**, *73*, 4433–4448.

34. Planas, D.; Veyer, D.; Baidaliuk, A.; Staropoli, I.; Guivel-Benhassine, F.; Rajah, M.M.; et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* **2021**, *596*, 276-280.

35. CDC. COVID data tracker. 8/22/2021– 8/28/2021. https://covid.cdc.gov/covid-data-tracker/#variant-proportions   (accessed on 7 September, 2021)

36. Peck, K.M.; Lauring, A.S. Complexities of Viral Mutation Rates. *J. Virol.* **2018**, *92*, e01031-17. doi: 10.1128/JVI.01031-17. PMID: 29720522; PMCID: PMC6026756.

37. BEI Resources. NR-54000 SARS-Related Coronavirus 2, Isolate hCoV19/England/204820464/ 2020 (Viruses) https://www.beiresources.org/Catalog/animalviruses/NR-54000.aspx   (accessed on 7 September, 2021)

38. Reuters. SA reaches grim milestone of 1 million Covid-19 cases. Dec 27, 2020. https://www.iol.co.za/news/south-africa/kwazulu-natal/sa-reaches-grim-milestone-of-1-million-covid-19-cases-a01906de-5442-451c-932f-84d6fc381b84   (accessed on 7 September, 2021)

39. Oliver O'Connell. Independent News. Will the Delta variant scare American diners and shoppers into staying home? Monday 26 July 2021. https://www.independent.co.uk/news/world/americas/covid-delat-variant-us-economy-b1890897.html   (accessed on 7 September, 2021)

40. Fact Check-The Delta variant death toll is not zero in the United States, as posts claim. https://www.reuters.com/article/factcheck-delta-zero/fact-check-the-delta-variant-death-toll-is-not-zero-in-the-united-states-as-posts-claim-idUSL1N2OO2I3   (accessed on 7 September, 2021)

41. Public Health England. SARS-CoV-2 variants of concern and variants under investigation in England. Technical briefing 18. 9 July 2021. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1001358/Variants_of_Concern_VOC_Technical_Briefing_18.pdf   (accessed on 7 September, 2021)

42. Zhang, Z.; Yao, W.; Wang, Y.; Long, C.; Fu, X. Wuhan and Hubei COVID-19 mortality analysis reveals the critical role of timely supply of medical resources. *J. Infect.* **2020**, *81*, 147-178. doi: 10.1016/j.jinf.2020.03.018. Epub 2020 Mar 21. PMID: 32209384; PMCID: PMC7163181.

43. Centers for Medicare and Medicaid Services. US Government.   CLIA SARS-CoV-2 Variant Testing Frequently Asked Question. Date: 3/19/2021. Does a facility that performs surveillance testing to identify

SARS- CoV-2 genetic variants need a CLIA certificate? https://www.cms.gov/files/document/clia-sars-cov-2-variant.pdf?ACSTrackingID=USCDC_2146-DM52811&ACSTrackingLabel=Laboratory%20Up-date%3A%20CMS%20Posts%20FAQ%20for%20Reporting%20Sequencing%20Results%20for%20SARS-CoV-2%20Variants&deliveryName=USCDC_2146-DM52811      (accessed on 7 September, 2021)

44. CDC. Common Human Coronaviruses. https://www.cdc.gov/coronavirus/general-information.html      (accessed on 7 September, 2021)

45. Leung, N.H.L. Transmissibility and transmission of respiratory viruses. *Nat. Rev. Microbiol.* **2021**, *19,* 528-545.

46. CDC. Covid Data Tracker Weekly Review. Interpretive Summary for September 3, 2021. https://www.cdc.gov/coronavirus/2019-ncov/covid-data/covidview/index.html    (accessed on 7 September, 2021)

47. Liu, Y.; Liu, J.; Johnson, B.A.; Xia, H.; Ku, Z.; Schindewolf, C.; et al. Delta spike P681R mutation enhances SARS-CoV-2 fitness over Alpha variant. bioRxiv 2021.08.12.456173; doi: https://doi.org/10.1101/2021.08.12.456173      (accessed on 7 September, 2021)

48. Urhan, A.; Abeel, T. Emergence of novel SARS-CoV-2 variants in the Netherlands. *Sci Rep.* **2021,** *11*, 6625. doi: 10.1038/s41598-021-85363-7. PMID: 33758205; PMCID: PMC7988010.

49. eCDC. SARS-CoV-2 variants of concern as of 6 September 2021. https://www.ecdc.europa.eu/en/covid-19/variants-concern    (accessed on 7 September, 2021)

50. Scheepers, C.; Everatt, J.; Amoako, D.G.; Mnguni, A.; Ismail, A.; Mahlangu, B.; et al. The continuous evolution of SARS-CoV-2 in South Africa: a new lineage with rapid accumulation of mutations of concern and global detection. https://www.medrxiv.org/content/10.1101/2021.08.20.21262342v2    (accessed on 7 September, 2021)