Received: 19 February 2022 | Accepted: 9 April 2022

DOI: 10.1002/jmv.27770

RESEARCH ARTICLE

JOURNAL OF MEDICAL VIROLOGY WILEY

A pyrosequencing protocol for rapid identification of SARS-CoV-2 variants

Richard A. Allen[1] | Christopher L. Williams[2] | Yvonne Penrod[1] | Cindy McCloskey[1] | Kristin Carpenter-Azevedo[3] | Richard C. Huard[3] | Ewa King[3] | Samuel Terence Dunn[2]

[1]OU Health Laboratories, Oklahoma City, Oklahoma, USA

[2]Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, Oklahoma, USA

[3]Rhode Island Department of Health (RIDOH) Laboratories, Providence, Rhode Island, USA

Correspondence
Samuel Terence Dunn, Department of Pathology, University of Oklahoma Health Sciences Center, 940 Stanton L. Young Blvd., BMSB 451, Oklahoma City, OK 73104, USA.
Email: terence-dunn@ouhsc.edu

Abstract

Next-generation sequencing (NGS) is the primary method used to monitor the distribution and emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants around the world; however, it is costly and time-consuming to perform and is not widely available in low-resourced geographical regions. Pyrosequencing has the potential to augment surveillance efforts by providing information on specific targeted mutations for rapid identification of circulating and emerging variants. The current study describes the development of a reverse transcription (RT)-PCR-pyrosequencing assay targeting >65 spike protein gene (*S*) mutations of SARS-CoV-2, which permits differentiation of commonly reported variants currently circulating in the United States with a high degree of confidence. Variants typed using the assay included B.1.1.7 (Alpha), B.1.1.529 (Omicron), B.1.351 (Beta), B.1.375, B.1.427/429 (Epsilon), B.1.525 (Eta), B.1.526.1 (Iota), B.1.617.1 (Kappa), B.1.617.2 (Delta), B.1.621 (Mu), P1 (Gamma), and B.1.1 variants, all of which were confirmed by the NGS data. An electronic typing tool was developed to aid in the identification of variants based on mutations detected by pyrosequencing. The assay could provide an important typing tool for rapid identification of candidate patients for monoclonal antibody therapies and a method to supplement SARS-CoV-2 surveillance efforts by identification of circulating variants and novel emerging lineages.

KEYWORDS

Coronaviridae, coronavirus, monoclonal antibody treatment, pyrosequencing, SARS-CoV-2, severe acute respiratory syndrome, spike protein, variant

1 | INTRODUCTION

Recently, much effort has been devoted to the identification and monitoring of genetic variations occurring within the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2; family *Coronaviridae*, subfamily *Orthocoronacirinae*, genus *Betacoronavirus*, subgenus *Sarbecovirus*) genome, some of which have been related to higher transmissibility, virulence, vaccine evasion, and/or reduced response to monoclonal antibody treatment. In addition, some variants have been shown to compromise diagnostic tests.

Abbreviations: cDNA, complementary deoxyribose nucleic acid; Ct, threshold crossing point; NGS, next-generation sequencing; RLU, relative light units; RT-PCR, reverse transcriptase-polymerase chain reaction; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SNV, single nucleotide variation; SQA, sequence analysis; VBM, variants being monitored; VOC, variant of concern; VOHC, variant of high consequence.

Understanding which variants of the virus are circulating and when new variants emerge is crucial to inform local epidemiological investigations, formulate public health action, and track regional, national, and global trends. The emergence of multiple variants of SARS-CoV-2 has been documented throughout the COVID-19 pandemic. The US SARS-CoV-2 Interagency Group's variant classification scheme defines four classes of SARS-CoV-2 variants: Variant Being Monitored (VBM), Variant of Interest; Variant of Concern (VOC), and Variant of High Consequence (VOHC).[1] Currently, several dozen VBMs are circulating in the United States at low rates of prevalence and are considered not to pose a significant or imminent threat to public health. The Delta (B.1.617.2) and Omicron (B.1.1.529) variants and their sublineages, owing to their demonstrated increased relative transmissibility, varying severity of disease, and ability to evade vaccination-induced immunity, are the only designated VOCs currently circulating in the United States as of February 1, 2022. To date, no VOHC has been identified in the United States.

SARS-CoV-2 variant surveillance efforts in the United States have relied largely on next-generation sequencing (NGS) performed at the Centers for Disease Control and Prevention, in the 130 state and local public health laboratories, and in multiple commercial and academic diagnostic and research laboratories nationwide.[2] While NGS provides a comprehensive characterization of the genome, generally, it is costly, low-throughput, time-consuming, and prone to technical issues when examining samples of low or variable quality. Several alternative technologies, less costly and complex to perform and with more efficient workflow, have the potential to augment SARS-CoV-2 variant screening efforts by providing information on specific targeted mutations for rapid identification of circulating and emerging variants. In addition, the urgent need for more real-time SARS-CoV-2 genotyping systems was recently realized by the lack of efficacy of certain monoclonal antibodies against specific variants.[3]

Pyrosequencing technology is an attractive alternative sequencing-by-synthesis method to conventional Sanger-based sequencing that can be used to provide a rapid and accurate characterization of short DNA sequence variations. Since sequencing data are produced in real-time during synthesis rather than by electrophoretic separation of fluorescently labeled fragments post-synthesis, it is a much more rapid methodology and is less costly to perform; however, it is limited by the amount of sequence data generated in a single experiment (typically, 30–60 bases vs. 200–400 bases). One of two different pyrosequencing formats is adopted depending on the sequence to be analyzed. An allele quantification (AQ) genotyping assay dispenses nucleotides into reactions relative to a defined variant sequence and returns a percentage value of variant to normal nucleotide detected at the variable position. By contrast, a sequence analysis (SQA) assay can use either a cyclic dispensation of A, T, C, and G nucleotides to interrogate unknown sequences or a defined dispensation order of nucleotides to detect variant sequences without quantification.

This study describes a novel pyrosequencing protocol for rapid identification of a set of relevant sequence variations within the spike protein gene (*S*) that can be used to characterize SARS-CoV-2 variants currently circulating in the United States and potentially to screen for emerging variants. The protocol can be performed on residual RNA derived from specimens previously tested in various SARS-CoV-2 diagnostic tests, or on residual cDNA from those tests, and thus obviates the need for heightened biosafety containment during analysis.

2 | MATERIALS AND METHODS

2.1 | Ethics

The study was reviewed and approved by the University of Oklahoma Health Sciences Center Institutional Review Board (#14161).

2.2 | Specimens

RNA extracted from residual clinical specimens was obtained from Rhode Island Department of Health (RIDOH) Laboratories (Providence, RI) and from the OU Health Virology Laboratory (Oklahoma City, OK), following diagnostic testing using the TaqPath™ COVID-19 Combo Kit (Thermo Fisher Scientific). Original specimens were collected in various transport media from mid-January 2021 through mid-December 2021. Specimens were deidentifed before receipt at the OU Health Molecular Pathology Laboratory (Oklahoma City, OK). Variant determination from NGS data, C_t values, and information regarding *S* gene target failure during diagnostic testing were provided with specimens, when available. In total, 49 specimens were obtained from RIDOH and 32 from OU Health. Specimens were purposely selected for analysis that were representative of the range of different variants detected at those facilities. RNA was stored at −80°C for up to 3 months before analysis by reverse transcriptase-PCR (RT-PCR) and pyrosequencing. Methods for acquisition, diagnostic testing, and sequencing of specimens at RIDOH are described elsewhere.[4]

2.3 | PCR and pyrosequencing primer design

Spike protein gene mutations were selected for analysis in the pyrosequencing assay based on their ability (alone or in combination with other mutations) to differentiate common SARS-CoV-2 variants circulating in the United States during 2021.[1] PCR and sequencing primers (Table 1) were designed using PyroMark Assay Design Software 2.0.2.5 and the *S* gene sequence as specified in the Wuhan reference sequence (US National Center for Biotechnology Information Reference Sequence NC_045512.2). Primers were subjected to BLASTn (NCBI) analysis to ensure specificity to SARS-CoV-2 sequences. Some of the common *S* gene mutations identified by our pyrosequencing assay are presented in Figure 1.

TABLE 1 PCR and pyrosequencing primer sequences and pyrosequencing nucleotide dispensations

PCR 1
Forward primer: 5′-CGT GGT GTT TAT TAC CCT GAC AAA-3′
Reverse primer: 5′-Biotin-AAT AAG TAG GGA CTG GGT CTT CG-3′
Amplicon length: 258 bp
Sequencing primer 1: 5′-TTA CTT GGT TCC ATG CTA-3′
DO: CTACATGTCTCTGAGCATGTGACTAGAGTGACGT
S2A: TACATGTCTCTGGGACCAATGG/TTAC/TTAAGAGGTTTGA/C/GTAAC
Analysis mode: AQ
Codon coverage: I68 to D80
Sequencing primer 2: 5′-CGT GGT GTT TAT TAC CCT GAC AAA-3′
DO: TACGTACTGTCTACATATGATGTGTATTGCTCGACTGAG
S2A: A/C/GTAACCCTGTCCTACCATTTAATGATGGTGTTTATTTTGCTTCCAC/TTGAGAAGTCTAACATAATAAGAGGC
Analysis mode: AQ
Codon coverage: D80 to E96
PCR 2
Forward primer: 5′-GTC CCT ACT TAT TGT TAA TAA CGC-3′
Reverse primer: 5′-Biotin-CCA TAA GAA AAG GCT GAG AGA CAT-3′
Amplicon length: 188 bp
Sequencing primer 1: 5′-TGT GAA TTT CAA TTT TGT AA-3′
DO: GTGATCATTGATGTACTCACAC
S2A: N/A
Analysis mode: SQA
Codon coverage: D138 to H146
Sequencing primer 2: 5′-CAC AAA AAC AAC AAA AGT T-3′
DO: TGCTACTGAAGTGACGTCAGAGTATC
S2A: GG/C/TATGG/AAAAGTGAGTT/CCA/GGAGTTTATTCT
Analysis mode: AQ
Codon coverage: W152 to S161
PCR 3
Forward primer: 5′-TTT TAC AGG CTG CGT TAT AGC TT-3′
Reverse primer: 5′-Biotin-GTT GCT GGT GCA TGT AGA AGT TC-3′
Amplicon length: 285 bp
Sequencing primer 1: 5′-CGT TAT AGC TTG GAA TTC T-3′
DO: CAACGATCTGACTCTAGTGTGTATATATAGCTGTATAGAT
S2A: AAC/A/GAATCTTGAT/CTCTAAGGTTGGTGGTAATTATAATTACCT/GGTA/TTAGATTG
Analysis mode: AQ
Codon coverage: N439 to L455
Sequencing primer 2: 5′-GAA ATC TAT CAG GCC G-3′
DO: GCTAGTCACACTGTATGTAGTCGAGTATGTACTCTACATCATATGTCACACTCGATATGTCACACTACATGT
S2A: GTANCAC/AACCTTGTAATGGTGTTNAAGGTTTTAATTGTTACTTTCCTTTACAAT/CCATATGGTTTCCAACCCACTA/TATGGTGTTGGTT
Analysis mode: AQ
Codon coverage: S477 to V503
PCR 4
Forward PCR primer: 5′-Biotin-TCC CTG TTG CTA TTC ATG CAG ATC-3′
Reverse PCR primer: 5′-TGA CAT AGT GTA GGC AAT GAT GGA-3′
Amplicon length: 236 bp
Sequencing primer 1: 5′-GTT TAA TAG GGG CTG AA-3′
DO: GCTATGTC
S2A: C/TATGTCAACAAC
Analysis mode: AQ
Codon coverage: H655 to V656
Sequencing primer 2: 5′-TGC GCT AGT TAT CAG ACT C-3′
DO: GACGTACTATCATCAGTCG
S2A: AG/C/TACTAATTCTCC/A/GTCGGCGGGCACGTAGT
Analysis mode: AQ
Codon coverage: Q677 to R682

Note: Nucleotide terminology follows International Union of Pure and Applied Chemistry recommendations.

Abbreviations: AQ, analysis mode used for quantification of different alleles; DO, dispensation order; S2A, sequence to analyze; SQA, analysis mode used for base-calling of unknown sequences.

2.4 | Reverse transcription-PCR

Eight microliters of RNA was reverse-transcribed in a 25 µl reaction at 25°C for 10 min, then at 50°C for 60 min and 85°C for 5 min, followed by a 4°C hold, using random nonamers and Invitrogen SuperScript™ III (Thermo Fisher Scientific). Following RT, 5 µl of cDNA were added to each of four PCRs containing 1× Clear GoTaq® Buffer (Promega), 2.5 mM $MgCl_2$, 0.125 mM each deoxynucleotide triphosphate, 0.2 µM of each primer, and 1.5 U GoTaq® Flexi DNA Polymerase (Promega). Each 50 µl reaction was subjected to an initial denaturation of 95°C for 2 min followed by 45 cycles of 95°C for 20 s, 58°C for 20 s, and 72°C for 20 s followed by a final extension of 72°C for 2 min and 4°C hold. Reagent and thermocycling conditions were optimized and standardized across the four PCRs for ease of PCR setup and workflow.

2.5 | Pyrosequencing

Ten microliters of amplified products from each of the four PCRs were sequenced separately on a PyroMark Q24 System (Qiagen) using

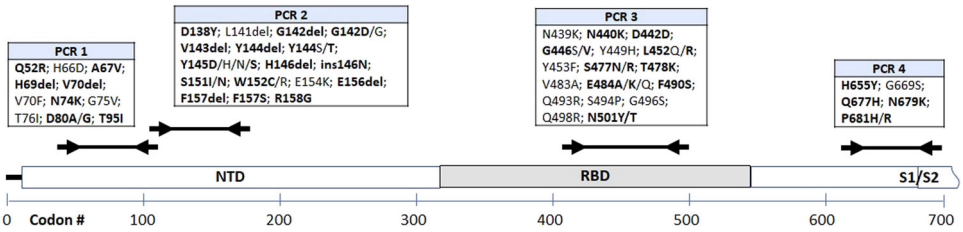

FIGURE 1 Schematic representation of the location of forward and reverse PCR primers (arrows) used to amplify four regions of the *S* gene of SARS-CoV-2. PCR 1 and PCR 2 target sequences corresponding to the N-terminal domain (NTD) of the S protein and overlap by 11 bases. PCR 2 resides within the receptor-binding domain (RBD), and PCR 4 overlies the junction between the S1 and S2 subunits (S1/S2). Common *S* gene mutations reported in SARS-CoV-2 variants, detectable by pyrosequencing the PCR products, using sequencing primers or forward PCR primers as listed in Table 1, are indicated in the respective boxes. Forty-two mutations detected in specimens analyzed in the current study appear in bold type. Other rare, novel mutations may be detected within targeted sequences but are not listed.

PyroMark Gold Q24 Reagents and the corresponding sequencing primers and nucleotide dispensation orders as shown in Figure 1. Analyses were performed using the AQ mode of the PyroMark Q24 Software and the corresponding sequence to analyze (S2A) or the SQA mode (Figure 1). In some cases where mutations underlying the 3′-end of the sequencing primer resulted in aberrant or no sequencing peaks, the forward PCR primer was used as the sequencing primer; nucleotide dispensation orders for these modified sequencing reactions are not shown but can be easily determined using the SARS-CoV-2 reference sequence and PyroMark Q24 Software. To ensure the reliability of automated base calls by the PyroMark Q24 Software v2.0.8, operators visually inspected pyrograms for evidence of mutant or aberrant peaks by side-by-side comparison with the pyrograms of the Wuhan-Hu-1 reference sequence and other specimens run concurrently.

2.6 | Typing tool

A web-based application (SARS2-TYPER) was developed to aid operators in typing SARS-CoV-2 variants following the identification of specific mutations in pyrograms. Identified mutations are selected from a list of common mutations found within regions of the *S* gene to generate a "Best-Fit Variant(s)" output. Specific variants and/or subvariants can be manually excluded from the search when prevalence rates are low to allow for greater discrimination of variants. Lists of mutations associated with variants and subvariants for inclusion in the typing application were compiled by review of multiple publicly available databases (e.g., CoVariants,[5] Outbreak.info,[6] CoV-RBD[7]) and the literature. A copy of the SARS2-TYPER can be accessed on GitHub (https://boredboar.github.io/covidtyper/) together with the source code (https://github.com/BoredBoar/covidtyper).

3 | RESULTS

The quality of sequence data revealed in pyrograms using products from individual PCRs was very good and easily interpreted for the majority of specimens; typically, single-height peaks exceeded 70 relative light units (RLUS) in amplitude with little to no background "noise." A few specimens, presumably due to lower amounts and/or quality of viral RNA, produced consistently low-amplitude peaks (single-height peaks ~10 RLUs); nevertheless, usually the sequence of these specimens was unambiguous and easily interpreted. One specimen produced exceptionally poor-quality pyrograms for all targets and another failed to provide any sequencing data, despite repeated attempts to amplify and sequence these specimens. All specimens had C_t values for viral-targets below 28 cycles when initially tested using the *Taq*Path™ COVID-19 Combo Kit, and C_t values of compromised specimens did not appear significantly elevated relative to other specimens; nevertheless, the poor performance of these specimens in the pyrosequencing assay is likely due to low amounts and/or quality of viral RNA.

Seventy-nine of 81 specimens provided sufficient quality of sequencing data of targeted *S* gene regions for appropriate assignment of variant type using the electronic SARS2-TYPER application. Comparing pyrosequencing data and available NGS data for corresponding specimens, all applicable mutations were identified and all specimens correctly genotyped. SARS-CoV-2 sequences identified included the Wuhan Hu-1 reference and B.1.1.7 (Alpha), B.1.1.529 (Omicron) B.1.351 (Beta), B.1.375, B.1.427/429 (Epsilon), B.1.525 (Eta), B.1.526.1 (Iota), B.1.617.1 (Kappa), B.1.617.2 (Delta), B.1.621 (Mu), and B.1.1 (contained only E484K and Q677H mutations for regions interrogated) variants. Common mutations associated with these and other variants and detectable using the pyrosequencing assay are indicated in Table 1. Some specimens revealed additional or missing mutations from those described in original variants. For instance, one B.1.351 subvariant had an N501T mutation instead of the standard N501Y mutation, one B.1.1.7 subvariant demonstrated a P681R mutation rather than the standard P681H mutation (i.e., Q4 subvariant) and one B.1.617.2 variant lacked H69del, V70del, G142D, Y144del and E884K/Q mutations (i.e., Delta V (3 + 2) subvariant).

Mutations specifically targeted by the assays had defined nucleotide dispensations set to detect the corresponding variable sequences at those locations, and therefore, produced pyrogram patterns that were easily interpreted (Figures 2A, B and 3A, B).

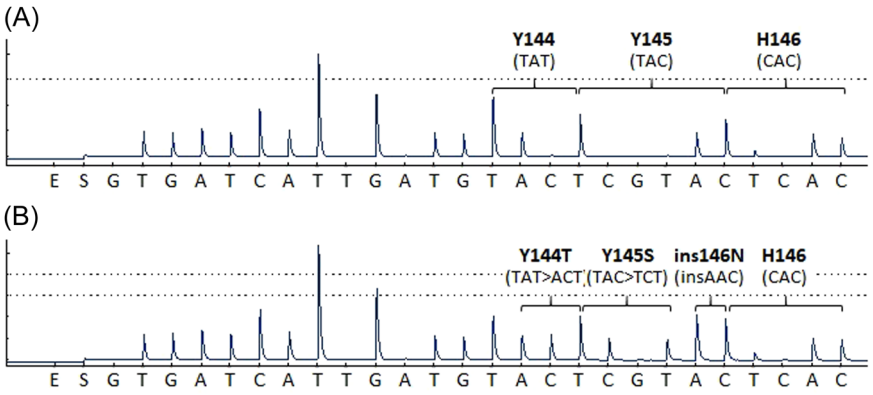

FIGURE 2 Pyrograms for two specimens amplified in PCR 2 and pyrosequenced using sequencing primer 1. (A) Specimen with normal (nonmutated) sequence, 5′-TGA TCC ATT TGG TGT TAT TAC CAC-3′. (B) Specimen with Y144T, Y145S, and 146N insertion mutations, 5′-TGA TCC ATT TGG TGT ACT TCT AAC CAC-3′.

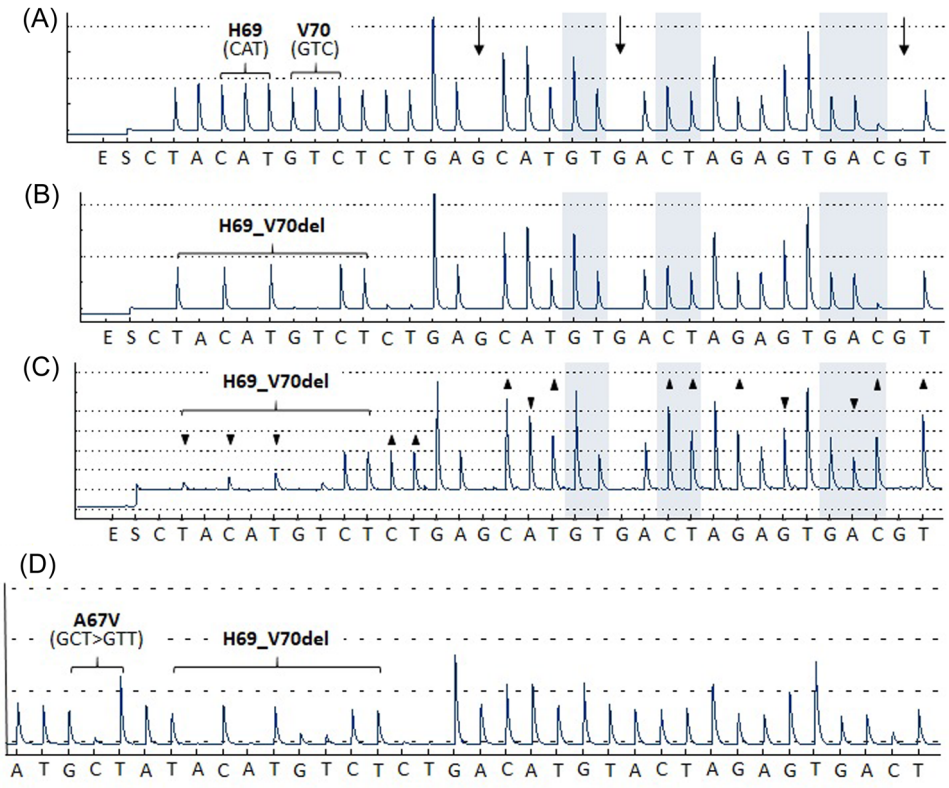

FIGURE 3 Pyrograms demonstrating the interference of a sequencing primer used to pyrosequence codons I68-D80. (A) Specimen without an H69_V70del mutation; note the presence of CAT and GTC peaks corresponding to codons H69-V70. Shaded boxes are for analyzing mutations in codons G75, T76, and D80 in AQ mode. Arrows indicate "blank" nucleotide dispensations (i.e., no peaks anticipated). (B) Specimen with H69_V70del, which entails the deletion of the third nucleotide of codon 68 through to the second nucleotide of codon 70. (C) Specimen with an S67V mutation underlying the 3′-end of the sequencing primer producing an aberrant uninterpretable pattern of peaks, some elevated (▲) or decreased (▼) in expected amplitude. (D) Same specimen as analyzed in (C) but sequenced using a primer located further 5′ (not all sequence is shown); note the presence of S67V and H69_V70del mutations. The peak pattern 3′ to H69_V70del is now similar to that shown in (A) and (B); however, blank nucleotide dispensations used in the analysis of specimens in (A) and (B) were omitted from the analysis of the specimen in (D) and specimen in (D) was analyzed in SQA mode.

Occasionally, however, unexpected mutations appeared within analyzed sequences of specimens. For instance, one B.1.1.7 specimen had a novel N74K mutation, one B.1.525 specimen had a novel G446V mutation, and one B.1.621 specimen had a D442 GAT > GAC silent mutation. While interpretation of most of these incidental mutations was straightforward, others, in the absence of the appropriate nucleotide dispensation order, were challenging to characterize. Nevertheless, these mutations did not compromise any overall assignment of variant type for these specimens.

The importance of careful design of primers was amply demonstrated by a number of reactions producing low or no PCR and/or sequencing products. In reactions used to detect the

60_70del mutation (Table 1: PCR 1, sequencing primer 1), six samples initially produced aberrant peaks throughout pyrograms, precluding accurate interpretation of sequence (Figure 3C). When the forward (unbiotinylated) PCR primer was used to pyrosequence the same specimens, an A67V mutation (GCT to GTT), underlying the third nucleotide from the 3′-end of the original sequencing primer, was revealed in all cases. The A67V mutation presumably interferes with the ability of the original sequencing primer to hybridize to fully complementary sequences during pyrosequencing, leading to the observed aberrant profiles in these samples. In addition to A67V and 60_70del mutations, all six of these samples harbored Q52R and E484K mutations, which effectively designated them as B.1.525 variants. A similar strategy was undertaken to resolve several other problematic cases, including two B.1.526.1 specimens and one B.1.429 specimen that failed to provide sequencing data due to an S151N mutation underlying the 3′-end of the sequencing primer and one specimen each of B.1.526.1, B.1.427, and B.1.429 variants and two B.1.351 specimens that failed to amplify using standard PCR 2 primers. In the latter five cases, the forward (unbiotinylated) PCR primer for PCR 1 was used in combination with the reverse (biotinylated) PCR primer from PCR 2 to amplify the specimens, and then products were pyrosequenced using sequencing primers normally used for PCR 2. As expected, mutation-laden Omicron specimens interfered with the performance of some pyrosequencing reactions, mostly leading to an inability to characterize some 3′-end sequences of PCR products; however, more than adequate numbers of mutations were identified to accurately type these specimens without the need to modify the assays.

4 | DISCUSSION

High-throughput and inexpensive surveillance methods that can deliver accurate and timely sequence data are needed by national, state, and local public health agencies to track the emergence and geographical spread of SARS-CoV-2 variants around the world. This is particularly important for tracking lineages harboring mutations that are associated with increased transmission, severity of disease and/or vaccine breakthrough. In addition, since not all monoclonal antibody (mAb) therapies work effectively on all variants, characterization of specific variants infecting individual patients may help identify candidates for specific urgent mAb therapies.

It is undeniable that NGS offers the greatest sensitivity for the detection of circulating and emerging SARS-CoV-2 variants; however, it is a costly, highly complex, and laborious technology and is unavailable in many locations, primarily due to a lack of resources and expertise. There is a critical need for new technical surveillance methods that are simple and accurate, offer at least moderate specimen throughput, and can be implemented in lower-resource communities. Several existing technologies that are well-suited to the detection of single nucleotide variations (SNVs) and small insertion and deletion sequences have the potential to complement and augment NGS surveillance efforts by their application in the interrogation of certain targeted sequences within the genomes of virus variants. While such genotyping methods still require specific molecular instrumentation and technical expertise, they are relatively less complex and much cheaper to perform than NGS, and can provide rapid identification of specific variants within a population with a high degree of confidence.

Several commercial SARS-CoV-2 genotyping assays are available, including the MassARRAY® SARS-CoV-2 Panel (Agena Biosciences), which uses RT-PCR followed by single-base extension of mutation-specific probes and mass spectrometry, the TaqMan SARS-CoV-2 Mutation Panel (Thermo Fisher Scientific), which uses RT-PCR and TaqMan SNP technology, the Allplex™ SARS-CoV-2 Variants I and II Assays (SeeGene), which use RT-PCR and a Tagging Oligonucleotide Cleavage and Extension (TOCE™) technology, and CoviDetect™ Variants and CoviDetect™ 4-plex (PentaBase), which uses real-time RT-PCR and DNA melt curve analysis. In addition, a variety of laboratory-developed tests have been described: Harper et al.[8] developed a SARS-CoV-2 variant genotyping RT-PCR allele-specific assay, targeting a set of 19 SNV markers, to provide genotype identification of UK variants with good discriminatory power; Diaz-Garcia et al.[9] designed a post-PCR, small-amplicon, high-resolution, melting analysis and an amplification-refractory mutation system assay to differentiate samples containing GR and non-GR clade SARS-CoV-2 viruses; and Vogels et al.[10] described a PCR assay targeting two mutations (3675–3677del in *ORF1a* and 69–70del in *S*) to differentiate B.1.351, P.1 and B.1.1.7.

We developed a set of pyrosequencing tests for genotyping SARS-CoV-2 variants by targeting four regions of the *S* gene, including the N-terminal domain, receptor binding domain, and S1/S2 cleavage site, known to contain mutations that would allow for the characterization of variants currently circulating in the United States, with a high degree of confidence. Assay workflow involved reverse transcription of residual RNA from SARS-CoV-2 diagnostic assays, followed by four separate PCR amplifications and then eight separate pyrosequencing reactions. Forty-two different mutations were detected and characterized in the specimens analyzed. Unfortunately, as encountered during the development of our assays, unanticipated mutations occasionally occurred in the *S* gene that led to reduced or failed PCR amplification and/or low quality or failed pyrosequencing reactions. Initially, some of these incidental mutations impacting assay performance were addressed by redesigning/moving primers; however, subsequently, we were able to resolve many of the same problems by substituting primer sets within individual assays (e.g., sequencing with the reverse PCR primer), or even between assays (e.g., using a forward PCR primer from one PCR together with the reverse biotinylated primer of another PCR to amplify cDNA). Other consequences of these unexpected base changes in viral sequence are that pyrosequencing reactions become temporarily stalled due to a divergence in viral sequence from the set nucleotide dispensation order and truncation of expected sequences. Usually, such changes were easily resolved by re-pyrosequencing PCR products using a modified dispensation order to accommodate the mutation and/or by extending the number of nucleotide

dispensations after the mutation to ensure appropriate coverage of sequences.

The web-based SARS2-TYPER application developed as part of this project proved efficient (significantly reducing technician time apportioned to variant assignment) and accurate (generating results that were concordant with lineage assignments made using NGS data in all cases). Because certain SARS-CoV-2 variants, in particular B.1.617.2 and B.1.1.7, have large numbers of subvariants defined by various combinations of mutations, not all subvariants could be easily accommodated in the typing application without diminishing its capacity to assign a best-fit variant(s) with confidence. To increase accuracy, a function was incorporated into the application whereby those variants/subvariants with very low prevalence within the community can be effectively disabled. Regional tracking data for SARS-CoV-2 variants can be readily accessed on various websites, (e.g., CDC, WHO, GISAID [CoVizu], and state and local health departments) and used to inform such decisions.

One advantage of our pyrosequencing assay over most other SARS-CoV-2 variant genotyping assays is the ability to identify novel mutations as they arise within interrogated sequences. Assays using allele-discriminating probes are likely to fail in the detection of newly emerging mutations or, if detected, may be unable to characterize the mutation. By contrast, typically, pyrosequencing is able to detect and characterize new mutations if they occur within regions being sequenced, although occasionally their characterization may require slight revision to the assay (e.g., additional dispensations or revision to the dispensation order) in which they occur. As the SARS-CoV-2 virus genome continues to evolve, it is likely that the spectrum of mutations featured in our assay, and those of others, will need to change to identify and differentiate variants circulating in the future. Accordingly, to continue to provide useful discriminatory capability and remain a relevant surveillance tool, our assay will likely need to undergo periodic revisions, with optimized addition and/or subtraction of PCRs and/or pyrosequencing reactions. While our PCR-pyrosequencing assay has the ability to augment epidemiological surveillance of newly emerging variants, given the limited portion of viral genome sequenced, we recognize that NGS is best suited for this role.

While NGS offers considerably greater amounts of sequence data, our pyrosequencing assay offers a moderately rapid, low-cost method for the identification of SARS-CoV-2 variants within a population. Since therapeutic efficacy of mAbs differs significantly with variant type, the assay shows particular promise in providing clinicians with a rapid means for identification of patients who are likely to respond to specific urgent mAb therapies. Considering that most NGS methods currently deployed for epidemiological surveillance are not validated for clinical purposes, and the often-protracted turnaround times for test results, NGS is not well-suited for such urgent clinical laboratory testing. By contrast, an appropriately validated PCR-pyrosequencing assay performed in a CLIA-certified laboratory could address much of these urgent testing demands. Residual RNA from SARS-CoV-2 RT-PCR diagnostic assays can be used directly in the RT-PCR-pyrosequencing assay, obviating the need for re-extraction and biocontainment. The assay in its current configuration takes approximately 9 h to process a batch of approximately 20 specimens, including approximately 4 h hands-on time for the RT-PCR (1 h), pyrosequencing (1 h), and pyrogram interpretation (2 h). The technical complexity of the assay is much less than that encountered in NGS protocols and can be easily accomplished by most technologists competent in performing other PCR-based assays. Modifications to the standard protocol to resolve occasional problem specimens (e.g., using the forward PCR primer to pyrosequence) can be rapidly performed using residual PCR products, which does not significantly increase processing times; alternatively, such specimens may be processed for NGS. Significant improvements to workflow and specimen throughput likely could be achieved with use of a more rapid thermal cycling platform, the Pyromark Q96 ID, PyroMark Q48 Autoprep, and/or an automated fluid-handling system. Multiplexing PCRs or redesigning PCR formats (e.g., the forward primer from PCR 1 and reverse primer from PCR 2 can be used to produce a single PCR product for subsequent pyrosequencing) also has the potential to conserve reagents and supplies and simplify workflow.

Some of the limitations of our assay have been discussed above and include the inability to monitor effectively for emerging lineages given the limited amount of genome coverage as compared to NGS, the occasional interference in assay performance from incidental variant sequences underlying primers, and the low throughput and high hands-on time. In addition, pyrosequencing technology platforms are not widely available in public health or clinical laboratories, which certainly will limit any wholesale application of the assay.

AUTHOR CONTRIBUTIONS

Richard A. Allen conducted RT-PCR-pyrosequencing and analysis of specimens and helped draft the manuscript. Christopher L. Williams developed the electronic genotyping application. Primary specimen diagnostic testing, NGS, specimen selection, and RNA extractions were coordinated by Yvonne Penrod and Cindy McCloskey at OU Health, and Richard C. Huard, Kristin Carpenter-Azevedo, and Ewa King at RIDOH. S. Terence Dunn was responsible for the study concept and design, data analysis and writing of the manuscript. All authors critically reviewed the manuscript before submission.

ACKNOWLEDGMENTS

We wish to express our gratitude to the staff of the OU Health Virology Laboratory, Oklahoma City, OK and the RIDOH Laboratories, Providence, RI for their efforts in the routine processing of diagnostic samples and generation of data used in this study.

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ETHICS STATEMENT

This study was performed in compliance with relevant laws and institutional guidelines and in accordance with the ethical standards of the Declaration of Helsinki. The study was reviewed by the University of Oklahoma Health Sciences Center Institutional Review Board (#14161).

REFERENCES

1. SARS-CoV-2 variant classifications and definitions. Centers for Disease Control and Prevention website. Accessed January 21, 2022. https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html (last accessed January 21, 2022).
2. Becker SJ, Taylor J, Sharfstein JM. Identifying and tracking SARS-CoV-2 variants—a challenge and an opportunity. *N Eng J Med*. 2021;385:389-391.
3. Coronavirus (COVID-19) update: FDA limits use of certain monoclonal antibodies to treat COVID-19 due to the omicron variant. U.S. Food and Drug Administration website. Accessed February 7, 2021. https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-limits-use-certain-monoclonal-antibodies-treat-covid-19-due-omicron
4. Kantor R, Novitsky V, Carpenter-Azevedo K, et al. SARS-CoV-2 variants in Rhode Island. *Rhode Island Med J*. 2021;104(7):16-20.
5. Hodcroft EB. CoVariants: SARS-CoV-2 mutations and variants of interest. Accessed January 21, 2022. https://covariants.org
6. Latif AA, Mullen JL, Manar Alkuzweny M, et al. Lineage Comparison.outbreak.info. Accessed January 21, 2022. https://outbreak.info/compare-lineages?pango=Lambda&sub=true
7. Tzou PL, Tao K, Nouhin J, et al. Coronavirus Antiviral Research Database (CoV-RDB): an online database designed to facilitate comparisons between candidate anti-coronavirus compounds. *Viruses*. 2020;12(9):1006. doi:10.3390/v12091006
8. Harper H, Burridge A, Winfield M, et al. Detecting SARS-CoV-2 variants with SNP genotyping. *PLoS One*. 2021;16:e0243185. doi:10.1371/journal.pone.0243185
9. Diaz-Garcia H, Guzmán-Ortiz AL, Angeles-Floriano T, et al. Genotyping of the major SARS-CoV-2 clade by short-amplicon high-resolution melting (SA-HRM) analysis. *Genes*. 2021;12:531. doi:10.3390/genes12040531
10. Vogels CBF, Breban MI, Ott IM, et al. Multiplex qPCR discriminates variants of concern to enhance global surveillance of SARS-CoV-2. *PLoS Biol*. 2021;19:e3001236. doi:10.1371/journal.pbio.3001236

How to cite this article: Allen RA, Williams CL, Penrod Y, et al. A pyrosequencing protocol for rapid identification of SARS-CoV-2 variants. *J Med Virol*. 2022;94:3661-3668. doi:10.1002/jmv.27770