Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

Diagnostic Microbiology and Infectious Disease 102 (2022) 115606

Contents lists available at ScienceDirect

Diagnostic Microbiology and Infectious Disease

journal homepage: www.elsevier.com/locate/diagmicrobio

SARS-CoV-2 variant detection with ADSSpike

Daniel Castañeda-Mogollón[a,b,c], Claire Kamaliddin[a,b,c], Laura Fine[a,b,c], Lisa K. Oberding[a,b,c], Dylan R. Pillai[a,b,c,d,*]

[a] *Cumming School of Medicine, Department of Pathology & Laboratory Medicine, the University of Calgary, Alberta, Canada*
[b] *Cumming School of Medicine, Department of Microbiology, Immunology, and Infectious Diseases, the University of Calgary, Canada*
[c] *Calvin, Phoebe & Joan Snyder Institute for Chronic Diseases, the University of Calgary, Calgary, Alberta, Canada*
[d] *Alberta Precision Laboratories, Diagnostic & Scientific Centre, Calgary, Alberta, Canada*

ARTICLE INFO

Article history:
Received 20 August 2021
Revised in revised form 6 November 2021
Accepted 16 November 2021
Available online 23 November 2021

Keywords:
SARS-CoV-2
Amplicon deep sequencing
Variants of concern
Variants of interest
S gene

ABSTRACT

The SARS-CoV-2 coronavirus pandemic has been an unprecedented challenge to global pandemic response and preparedness. With the continuous appearance of new SARS-CoV-2 variants, it is imperative to implement tools for genomic surveillance and diagnosis in order to decrease viral transmission and prevalence. The ADSSpike workflow was developed with the goal of identifying signature SNPs from the S gene associated with SARS-CoV-2 variants through amplicon deep sequencing. Seventy-two samples were sequenced, and 30 mutations were identified. Among those, signature SNPs were linked to 2 Zeta-VOI (P.2) samples and one to the Alpha-VOC (B.1.17). An average depth of 700 reads was found to properlycorrectly identify all SNPs and deletions pertinent to SARS-CoV-2 mutants. ADSSpike is the first workflow to provide a practical, cost-effective, and scalable solution to diagnose SARS-CoV-2 VOC/VOI in the clinical laboratory, adding a valuable tool to public health measures to fight the COVID-19 pandemic for approximately $41.85 USD/reaction.

© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

1. Introduction

The COVID-19 pandemic's causative agent is the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), first identified at the end of 2019 in Wuhan, China (Wu et al., 2020). Following the appearance of mutations in specific SARS-CoV-2 genes, different viral phenotypes have emerged as the main threat to epidemic control. These viral variants, known as variants of concern (VOC), and variants of interest (VOI) are mutants that are associated with either increased transmissibility, changes in epidemiological patterns or clinical presentation, increased virulence, decreased effectiveness of public health and social measures for epidemic control, and/or decreased effectiveness of available diagnostics, vaccines, and/or therapeutics.

Genomic surveillance of these mutants is essential for monitoring viral spread and implementing targeted measures to reduce VOC/VOI transmission. Current surveillance strategies rely on whole genome sequencing (WGS) for identification of VOC/VOI, or assays to monitor and detect known mutations. WGS using next generation sequencing (NGS) approaches are based on commercially available sequencing kits or international consortia-published protocols such as those from the ARTIC network (COVID-19 ATIC, 2020). The majority of the targeted assays rely on PCR amplification of a genomic region followed by amplicon Sanger sequencing (Cabral et al., 2021; Sanger Sequencing Solutions for SARS-CoV-2 Research - CA). Other assays are based on reverse transcription quantitative polymerase chain reaction (RT-qPCR) using TaqMan probes targeting specific SNPs (Bier et al., 2021), amplicon melting curve analysis, or high range melting analysis (Banada et al., 2021; Diaz-Garcia et al., 2021). These targeted assays are limited to currently known SNPs related to VOC/VOI and do not have the potential for rapid identification of new mutations, especially during a pandemic where turnaround times are a critical factor for molecular surveillance.

The SARS-CoV-2 S gene encodes for the surface glycoprotein of the virus which mediates viral adhesion and entry into human host cells. Mutations in the S-gene sequence in VOC/VOI have resulted in amino acid changes in the S protein that affect the binding to the human cell receptor ACE (Boehm et al., 2021; Hoffmann et al., 2020), antibody recognition, vaccine efficacy, and antibody therapy efficacy (Boehm et al., 2021). In addition, this gene represents an attractive target for SARS-CoV-2 VOC/VOI identification, as signature SNPs, insertions, and deletions can be easily identified in this region through amplicon deep sequencing (ADS).

Here, a novel workflow using ADS of the S gene (ADSSpike) was implemented to provide a scalable, high throughput, and unbiased identification workflow for VOC/VOI from clinical SARS-CoV-2

---

* Corresponding author. Tel.: +1-403-770-3578; fax: +1-403-770-3296.
E-mail address: drpillai@ucalgary.ca (D.R. Pillai).

https://doi.org/10.1016/j.diagmicrobio.2021.115606

0732-8893/© 2021 The Author(s). Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

samples. In addition, ADSSpike simplifies the ARTIC Illumina V3 protocol (COVID-19 ARTIC, 2020) by incorporating Illumina overhangs into PCR primers targeting 400 bp regions spanning the SARS-CoV-2 S gene, with the aim of reducing PCR cycles and thus chimera formation (Lahr and Katz, 2009; Sze and Schloss, 2019). To our knowledge, ADSSpike is one of the first workflows to identify SNPs by employing a SARS-CoV-2 S gene-targeted amplicon deep sequencing approach across SARS-CoV-2 specimens. A previous pipeline discusses a similar methodology; however several pitfalls compared to this pipeline were observed Fass et al., 2021.

2. Methodology

2.1. Ethics statement

Ethical approval was obtained from the Conjoint Health Research Ethics Board of the University of Calgary (REB 20-0567, REB 20-0402). All archived specimens were de-identified prior to analysis in this study.

2.2. Sample collection and nucleic acid extraction

Clinical nasopharyngeal swab and throat swab specimens ($n = 72$) were collected and tested for SARS-CoV-2 by Alberta Precision Laboratories between March 2020 and February 2021 as previously described (Pabbaraju et al., 2021). RNA was extracted from 90 to 120 μL of sample using the Qiagen QIAamp® Viral RNA Mini Kit (Cat. No/ID 52906, Qiagen, USA), following the manufacturer's protocol with slight modifications. Briefly, samples were digested with proteinase K for 10 minutes at 56°C prior to extraction, and treated with DNAse I (Promega) to remove any remaining DNA. Obtained RNA was eluted in 2 centrifugation rounds of 40 μL of nuclease-free water.

2.3. Amplicon deep sequencing (ADS) strategy

The ADS pipeline (Fig. 1) was designed to optimize sample preparation time. Briefly, extracted RNA was subjected to cDNA synthesis before PCR amplification. Fourteen pairs of primers spanning the S gene and its adjacent regions, generating ~400 bp amplicons, were designed and adapted from published protocols (COVID-19 ARTIC, 2020) (Supplementary Table 1). The PCR amplification step was designed to include the Illumina adapters (thus allowing PCR products to move directly to library preparation and sequencing while reducing PCR artifacts and chimeric reads. Illumina adaptors fused to the locus-specific primers is a strategy that has been used in the past from previous protocols and was adapted from current protocols for SNP calling and haplotype screening (Schnell et al., 2015; 16S Metagenomic Sequencing Library Preparation) with each primer pair). The performance of a pooling strategy was evaluated, comparing the sequencing results from PCR conducted either individually (iPCR, conducting 14 individual PCR reactions each yielding 400 bp amplicons), or as a pooled PCR (pPCR) strategy (one-pot PCR reaction using the 14 primer pairs, generating mixed 400 bp amplicons from each primer set). For each sample, the iPCR products were pooled prior to purification and subsequent analysis (Supplementary Fig. 1).

2.4. cDNA synthesis and PCR amplification of the SARS-CoV-2 S Gene

The cDNA synthesis and PCR amplification of the SARS-CoV-2 S gene are detailed in the Supplementary data. Briefly, the cDNA synthesis was performed using the LunaScript® RT Supermix (New England Biolabs, #E3010L) following the manufacturer's protocol. The S gene and its adjacent regions were amplified using 14 primer pairs, followed by a single PCR run and SPRIselect bead purification (Beckman Life Science, USA). Three nuclease-free water negative controls were used to assess potential amplicon contamination in the lab, as well as to help tuning the parameters for SNP calling (Table 1).

2.5. Library preparation and sequencing

Fifteen μL of pooled purified amplicons per sample was sent to the Center for Health Genomics and Informatics (CHGI) at the University of Calgary for library preparation and sequencing. Each sample was indexed using the Nextera XT Index Kit V2 along with KAPA HiFi polymerase. The indexed libraries were pooled and sequenced in an Illumina MiSeq instrument (San Diego, CA) in paired-end mode

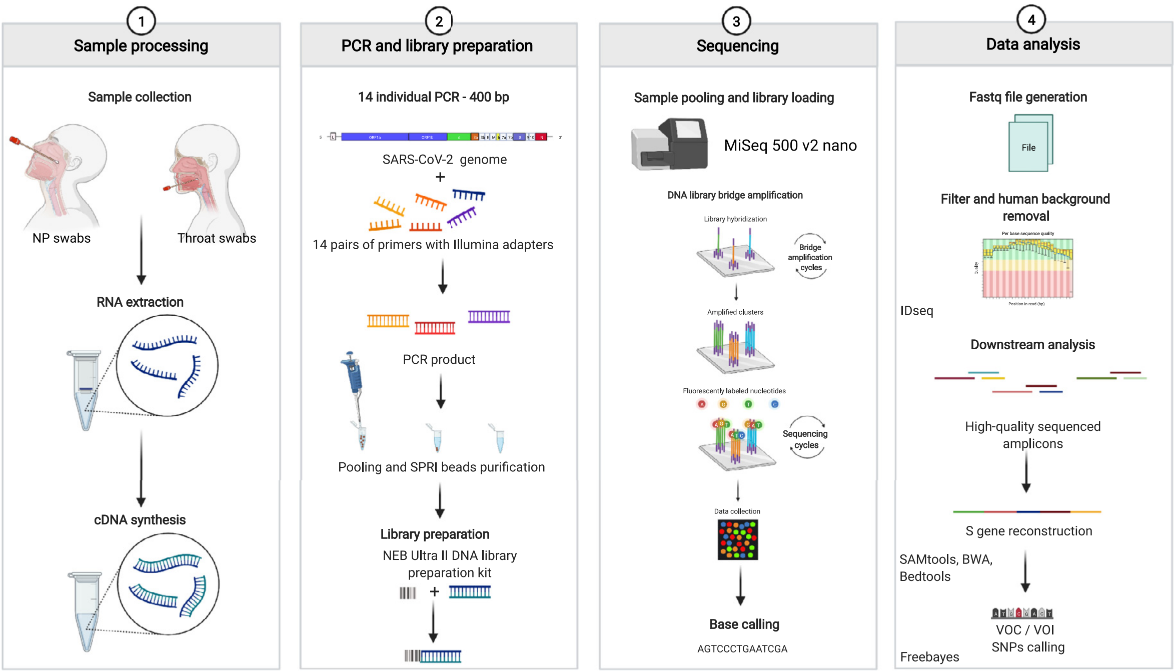
1
Sample processing
Sample collection
NP swabs
Throat swabs
RNA extraction
cDNA synthesis
2
PCR and library preparation
14 individual PCR - 400 bp
SARS-CoV-2 genome
+
14 pairs of primers with Illumina adapters
PCR product
Pooling and SPRI beads purification
Library preparation
NEB Ultra II DNA library preparation kit
+
3
Sequencing
Sample pooling and library loading
MiSeq 500 v2 nano
DNA library bridge amplification
Library hybridization
Bridge amplification cycles
Amplified clusters
Fluorescently labeled nucleotides
A G T C
Sequencing cycles
Data collection
Base calling
AGTCCCTGAATCGA
4
Data analysis
Fastq file generation
File
Filter and human background removal
Per base sequence quality
IDseq
Downstream analysis
High-quality sequenced amplicons
S gene reconstruction
SAMtools, BWA, Bedtools
VOC / VOI SNPs calling
Freebayes

Fig. 1. Spike gene amplicon deep sequencing (ADS) pipeline named ADSSpike. The presented workflow was divided among 4 steps: sample processing, PCR and library preparation, sequencing, and data analysis (read filtering, alignment, SNP calling, and variant assessment).

Table 1
Summary of the clinical samples used in this study.

Variables	SARS-CoV-2 positive (n = 68)	SARS-CoV-2 negative (n = 4)	P-value
Anatomical swabbing site (n = 72)			0.3058
Nasopharyngeal (n = 50)	46	4	
Oropharyngeal (n = 22)	22	0	
Swab and transport media (n = 72)	68	4	0.2567
Saline (n = 16)	16	0	
Aptima (n = 19)	19	0	
E-swab (n = 5)	5	0	
UTM (n = 35)	28	4	
RT-PCR E gene Ct value (n = 68), mean ±SD [min, max]	22.83 ± 4.77 [17.28, 35.73]	N/A	

(2 × 250 bp), using the Illumina MiSeq Reagent Kit V2 Nano (500 cycles), for a total of 1 million reads, with a 5% spike of Enterobacteria phage PhiX control.

2.6. Read filtering and variant calling parameter optimization

Samples were submitted to the IDseq pipeline for adapter and primer trimming, removal of low-quality reads, and removal of host sequences (Kalantar et al., 2020). Reads were mapped using the Burrows Wheeler Aligner (BWA) (Li and Durbin, 2009) against a SARS-CoV-2 reference genome (GenBank MN908947.3 (Wu et al., 2020)). Samtools was used for file indexing and sorting (Li et al., 2009), followed by FreeBayes (Garrison and Marth, 2012) to perform SNP calling and insertion/deletion (indel) identification. To determine which parameters are best to identify SARS-CoV-2 VOC/VOI, a combination of parameters were employed and the positive predictive value (PPV) along with analytical sensitivity were computed. For this, a depth of coverage (number of reads supporting a SNP or deletion) between 5 to 40 reads was tested, along with a fraction depth between 15% to 90% (fraction of reads supporting the SNP or deletion); a Phred score of 20 was maintained across all parameter combinations. Assessment details of the PPV and sensitivity are available in the Supplementary data. Read depth and coverage, defined as the percentage of the S gene covered by at least one read, were assessed by BEDTools (Quinlan and Hall, 2010) and visualized with Tablet (Milne et al., 2010).

2.7. Statistical analysis

Descriptive statistics regarding depth and coverage were performed by including the mean ± standard deviation. Non-parametric Mann-Whitney tests were performed across the continuous variables, and a Fisher's exact test was performed across categories with discrete variables. A nonparametric Kruskal-Wallis test followed by multiple-paired comparisons were carried out to determine difference in significance amongst coverage and Ct value groups. All statistical tests were considered significant for adjusted p-values below 0.05. The statistical analysis and figure design were performed using GraphPad Prism for Mac.

3. Results

3.1. Included samples

Seventy-two samples were selected and tested with ADSSpike (68 were positive for SARS-CoV-2, and 4 were negative). This sample size was selected to aim for an average of 1,000 reads per amplicon per positive sample, for a total of 14,000 reads per positive sample. Of the tested samples, 50 were from nasopharyngeal swab, and 22 were from throat swab. The majority of the samples were stored in UTM (n = 35), followed by Aptima (n = 19), saline (n = 16), and E-swab (n = 5). Ct values for the E gene RT-PCR amongst the positive samples varied from 17.28 to 35.73, with a mean of 22.83. Samples were previously sequenced by an Illumina COVIDSeq SARS-CoV-2 NGS test, where 3 samples were VOC/VOI (2 samples were Zeta/P.2, one sample was Alpha/B.1.1.7).

3.1.1. Preliminary sequencing assessment

A significant difference was observed in the purified DNA concentration after S gene amplification when samples were grouped by SARS-CoV-2 diagnosis (positive: 13.8 ± 15.0 vs negative: 1.0 ± 0.2 ng/μL; P-value = 0.03, Mann-Whitney test). Amplicons from 79 samples, including negative controls and paired samples, were sequenced on Illumina MiSeq in paired-end mode. A total of 835,676 paired raw reads were retrieved; of all reads, 1,210,743 were properly mapped to the S gene reference sequence, with a median of 10,610 (Q1 and Q3: [4432, 14877]), and a mean of 15,325 reads per sample. The positive samples displayed an average read depth of 869.81 ± 477.37, while the negative samples had an average read depth of 30.68 ± 66.31 (Fig. 2A). Analysis of read depth by Ct value groups displayed an expected pattern where the lowest Ct values tended to have a higher read depth across the S gene (Fig. 2B). An average of 70.38% coverage was observed for all samples, with 76.37% ± 34.90% for the SARS-CoV-2 positive samples and 33.18% ± 13.46% for the negative samples. An analysis of coverage by Ct value grouping shows a similar trend, where there is a significantly different mean across groups (Kruskal-Wallis test; P-value < 0.0001) (Fig. 2C). A multiple comparison paired-analysis test suggests significant difference by coverage between Ct values of 15 to 20 vs 25 to 30 (adjusted P-value = 0.02), 15 to 20 vs 30 to 35 (adjusted P-value = 0.0002), and 20 to 25 vs 30 to 35 (adjusted P-value = 0.0003), detailed in the Supplementary data.

3.2. Assessment of parameter selection for variant calling

Seven parameter combinations of read depth and fraction depth were analyzed for SARS-CoV-2 variant calling (Table 2). Our results indicate that the highest PPV and sensitivity were obtained by 3 combinations: (1) read depth of 10 and depth fraction of 50%, (2) read depth of 40 and depth fraction of 60%, and (3) read depth of 40 and depth fraction of 70%. These combinations were able to capture all the signature SNPs pertinent to the Alpha/B.1.1.7 strain and one of the Zeta/P.2 VOI samples. Only one SNP was captured with the remaining Zeta/P.2 strain by any of the combinations. Upon closer inspection, no reads were generated in the region where the remaining 2 SNPs were expected for the Zeta sample. The rest of the parameter combinations were not stringent enough (sacrificing positive predictive value, and therefore analytical specificity), or were too robust to capture real and expected SNPs (sacrificing analytical sensitivity). Because a tie occurred amongst 3 parameter combinations, a variant screening procedure was performed across all sequenced samples.

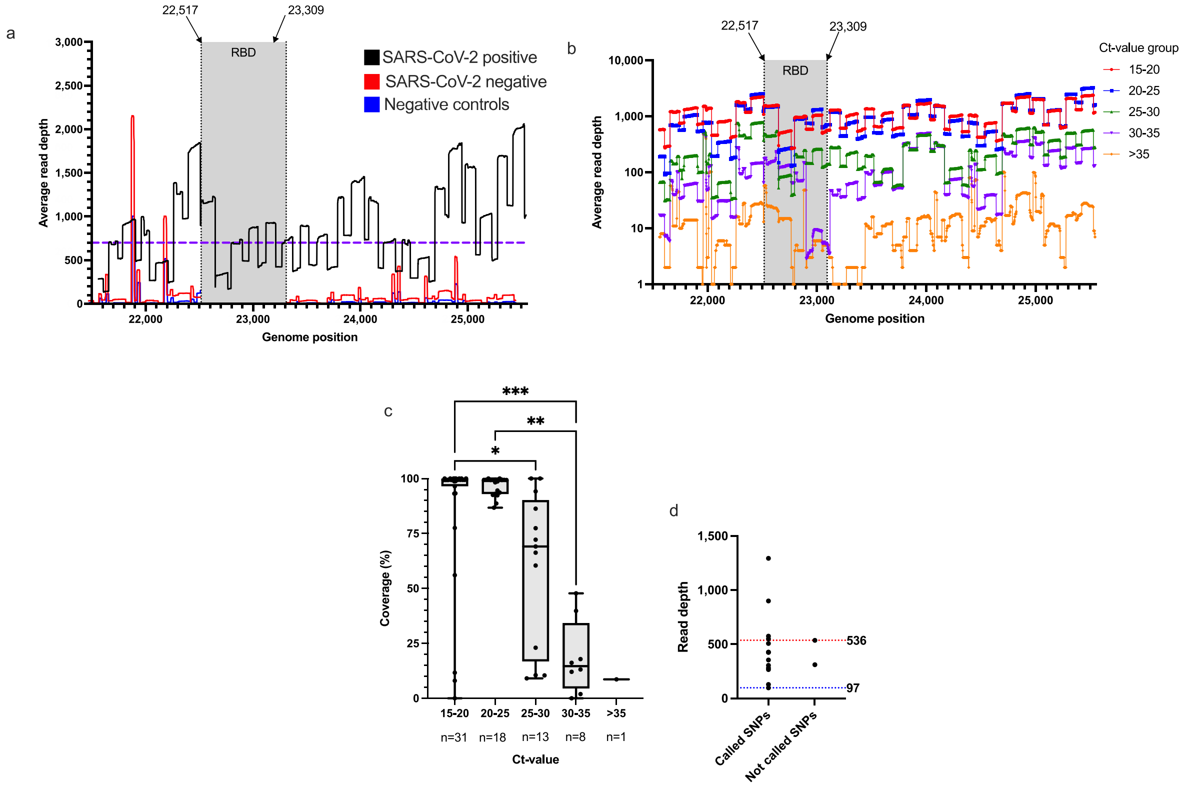

Fig. 2. Read depth and coverage assessment. (A) Plot displaying the mean read depth distribution by individual nucleotide position. Depth was assessed by SARS-CoV-2 positive iPCR samples, SARS-CoV-2 negative samples, and negative controls (NFWiPCR). The dashed purple line represents a 700 read depth. (B) Plot displaying the mean read depth distribution by individual nucleotide position amongst Ct value groups. Depth was assessed by SARS-CoV-2 positive iPCR samples. (C) Side by side box plot of coverage by Ct value groups assessed by a Kruskal-Wallis test and multiple-paired comparisons (*: adjusted *P*-value < 0.05; **: adjusted *P*-value < 0.01; ***: adjusted *P*-value < 0.001). (D) Dot plot displaying the position-specific depth of the identified SNPs for the VOC/VOI (left) vs missed SNPs. The blue dashed line represents the lowest depth for a called SNP, and the red dashed line represents the highest depth recorded for a missed SNP.

Variant calling between the top 3 parameters was performed. Our analysis revealed that a read depth of 10 and a depth fraction of 50% recorded 121 SNPs and deletions. A read depth of 40 and depth fraction of 60% revealed a total of 119 SNPs and deletions, and a read depth of 40 and depth fraction of 70% showed 118 SNPs and deletions. After analysing each of the called variants amongst these sets of combinations, our analysis revealed that 4 false positives were detected by the first approach, 3 false positives and 2 false negatives by the second, and 3 false positives and 3 false negatives by the third. This suggests that a read depth of 10 and a depth fraction of 50% has a PPV of 96.69%, while a read depth of 40 with a depth fraction of 60% and 70% had a PPV of 97.45% and 97.43%, respectively (Supplementary table 2). Nevertheless, when employing a read depth of 40% and a fraction depth of 60% and 70%, 2 and 3 false negatives were observed, decreasing the sensitivity to 98.29% and 97.43%, respectively. This revealed that a read depth of 10 and a depth fraction of 50% had gave the optimum values for SNP calling and SARS-CoV-2 identification, and were therefore used for variant analysis.

3.3. *Read depth as a proxy for accurate SNP calling*

The previously sequenced VOC/VOI samples, P739 (Alpha/B.1.1.7-VOC; Ct value 27.17), P743 (Zeta/P.2-VOI; Ct value 23.68), and P744 (Zeta/P.2-VOI; Ctvalue 30.17), were used as controls to assess the identification of VOC/VOI from the S gene sequences. The samples had average read depths of 713, 699, and 406, respectively. The P739 sample contained all S gene mutations pertinent to the Alpha/B.1.1.7-VOC (Δ69/70, Δ144/145, N501Y, A570D, D614G, P681H, T716I, S982A, and D1118H). The P743 sample contained all the S gene mutations associated with the Zeta/P.2-VOI (E484K, D614G, and V1176F); in contrast, the P744 sample only showed the V1176F VOC-flagged mutant, suggesting a minimum average read depth of 700 may be required to identify all SNPs pertinent to VOC/VOI. A depth of 97 was

Table 2
Parameter iteration for PPV and sensitivity calculation in SARS-CoV-2 VOC/VOI.

Parameter combination	Alpha/B.1.1.7 VOC PPV and sensitivity	Zeta/P.2 VOI (sample 1) PPV and sensitivity	Zeta/P.2 VOI (sample 2) PPV and sensitivity
Read depth = 5; Depth fraction = 15%	22.5%; 100%	9.67%; 100%	1.47%; 33.33%
Read depth = 10; Depth fraction = 33%	100%; 100%	100%; 100%	16.67%; 33.33%
Read depth = 10; Depth fraction = 50%	100%; 100%	100%; 100%	100%; 33.33%
Read depth = 40; Depth fraction = 60%	100%; 100%	100%; 100%	100%; 33.33%
Read depth = 40; Depth fraction = 70%	100%; 100%	100%; 100%	100%; 33.33%
Read depth = 40; Depth fraction = 80%	100%; 88.89%	100%; 100%	100%; 33.33%
Read depth = 40; Depth fraction = 90%	100%; 77.78%	100%; 66.67%	100%; 33.33%

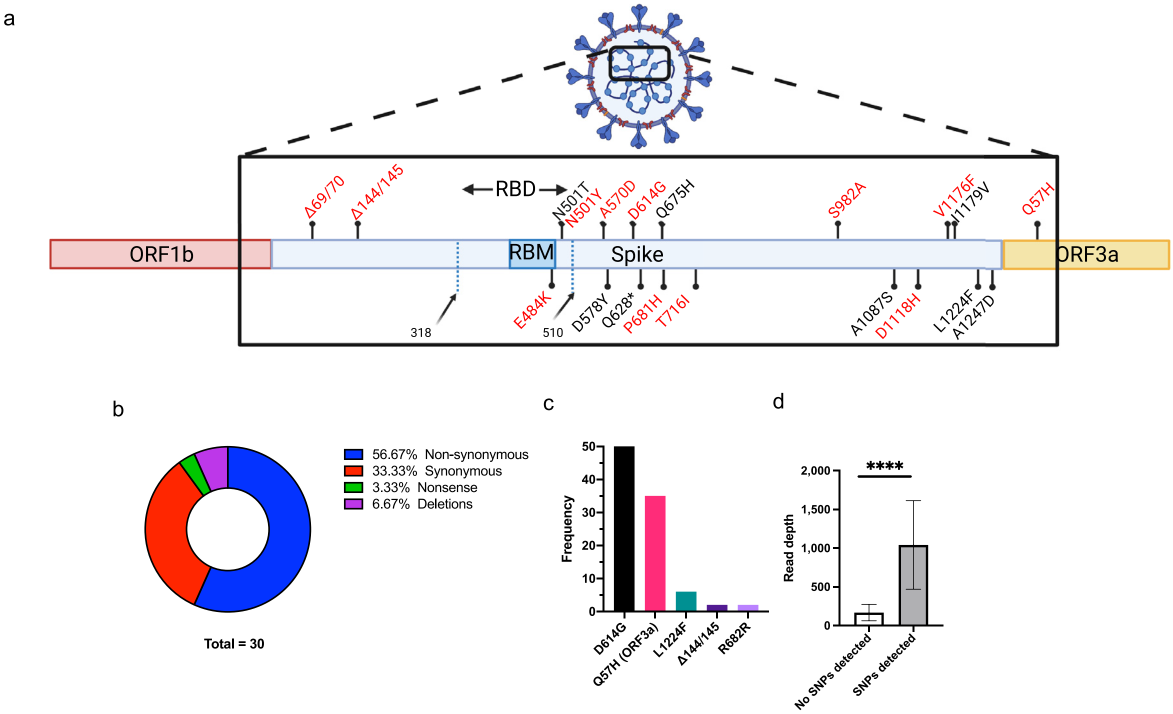

Fig. 3. Identified SNPs indels SARS-CoV-2 positive samples. (A) Non synonymous mutations, deletions, and nonsense mutants identified amongst the 64 iPCR SARS-CoV-2 positive samples in the S and flanking ORF3a genes region. Blue dashed lines represent the receptor binding domain (RBD) delimited by the nucleotides 318 and 510, and its receptor binding motif (RBM). Red mutants are flagged for their presence in VOC/VOI. (B) Pie chart of the SNP type distribution. (C) Side-by-side bar chart of the read depth by SARS-CoV-2 positive samples with no SNPs detected vs samples with at least one identified SNP (D) Box plot of the 5 most prevalent mutants identified across all positive samples (n = 72).

found to be the lowest acceptable for accurate SNP calling, while a depth of 536 was the highest observed for missing a SNP (Fig. 2D). We also observed that 55/76 (72.36%) samples had an 50% coverage, a requirement for submission to the GISAID database (Shu and McCauley, 2017).

3.4. SNP calling and variant confirmation

Sixty-eight samples from SARS-CoV-2 positive patients were successfully sequenced. The negative samples did not display any SNPs (coverage = 33.18% ± 11.55%; depth: 30.68 ± 66.31). Amongst the positive samples, a total of 30 unique mutations were identified after applying the thresholds for SNP calling, with a total frequency of 121 (Fig. 3A). Of these, 10/30 were synonymous (total frequency of 15/121), 17/30 were nonsynonymous (total frequency of 102/121), one was a nonsense mutation (total frequency of 1/121), and 2 were indels (total frequency of 3/121) (Fig. 3B). Fifteen SARS-CoV-2 positive samples did not display any SNPs. These samples featured a significantly lower read depth (mean = 168.05; SD = 105.71) compared to those with at least 1 identified SNP (mean = 1040.39; SD = 71.86) (P-value < 0.0001 Mann-Whitney test) (Fig. 3C).

The most commonly identified mutants were D614G (frequency of 50/68 or 73.52%), followed by the Q57H mutant in the ORF3a gene (frequency of 30/68 or 44.11%), L1224F (frequency 6/68 or 8.82%), deletion 144/145 (frequency of 2/68 or 2.94%), and the synonymous R628R mutant (frequency of 2/68 or 2.94%) (Fig. 3D). In addition, 4 independent samples had the A21583G synonymous mutant encoding for L7L in the S gene, a variant that has not been recorded in the literature.

3.5. Pooling strategy assessment

An alternative strategy was evaluated in order to test the efficacy of a simplified experimental workflow (running one iPCR for each of the 14 primer pairs vs pPCR with 14 primers pairs). A highly significant difference between the aligned read depths was observed, with a mean depth of 560±237.88 for the iPCR samples and 312.87 ± 372.72 for the pPCR samples (P-value < 0.0001; paired t-test). On the 3 samples tested with the 2 approaches, a total of 5 mutations were identified in the S gene, with a positive percent agreement of 80%. One mutant was missed (Δ144/145) and had a read depth of 422 at the flanking regions of the indel by the iPCR approach (Supplementary Table 3).

4. Discussion

The close monitoring of SARS-CoV-2 mutants is imperative, especially for the continuous development of therapeutics, vaccines, and for better understanding of pathogenesis. This study reports the development of ADSSpike, a practical workflow for selective ADS of the SARS-CoV-2 S gene. The overall workflow is an adaptation of the international ARTIC V3 consortium protocol (COVID-19 ARTIC, 2020) and is similar to the HiSpike pipeline (Fass et al., 2021). One key difference to HiSpike is the use of its relaxed parameters for SNP calling. Indeed, screening for SNPs pertinent to SARS-CoV VOC/VOI with the parameters employed by HiSpike (read depth of 15), revealed a PPV below 25% in each case VOC/VOI analyzed (Table 2). In addition, the HiSpike workflow has no mention of the fraction of reads employed for base calling, a parameter of uttermost importance in the identification of multiple SARS-CoV-2 strains in a single host. Despite mentioning HiSpike as a cost-effective tool, no cost description is provided, which adds uncertainty to the methodology employed. The HiSpike study does not provide enough information on SNPs detected amongst negative controls or SARS-CoV-2 negative samples, and a detailed cost analysis of the pipeline is missing. On the other hand, similar results were recorded with the findings by Fass et al. in terms of coverage decline when samples had a Ct value of 30 or higher. ADSSpike was able to identify SNPs and deletions in clinical specimens that tested positive for SARS-CoV-2, and no non-specific sequences were retrieved from the negative control samples, showing the specificity of the workflow. The results presented here suggest ADSSpike is an accurate SNP detection workflow regardless of the SARS-CoV-2 lineage.

ADSSpike provides a practical, cost-effective, and scalable solution to diagnose and monitor SARS-CoV-2 VOC/VOI in the clinical laboratory, adding a valuable tool to public health measures combatting the COVID-19 pandemic for approximately $41.85 USD per reaction. Current approaches to identify VOC/VOI are either based on capillary sequencing, WGS, or targeted PCR or RT-PCR assays. Targeted assays are commercially available but only cover existing mutations, meaning new SNPs can be missed (Wang et al., 2021). A comparison of cost, time, and overall advantages and limitations of SARS-CoV-2 VOC/VOI methods was performed across current methodologies (Supplementary table 4).

SARS-CoV-2 VOC detection using WGS by capillary sequencing or NGS has been widely used for surveillance (Goncalves Cabecinhas et al., 2021; Miller et al., 2020; Pillay et al., 2020; Shaibu et al., 2021). In contrast to ADS, amplicon deep sequencing focuses on a specific gene or genomic region, providing a greater sequencing depth and coverage while reducing costs and data burden. Sanger-based approaches have been previously employed (Bezerra et al., 2021; Goes et al., 2021). These are either focused on a small window in the S gene, and hence missing relevant SNPs flanking said region, or target SNPs prevalent to specific VOC/VOIs. While SNPs can be identified using Sanger sequencing, mutations present in a minority of the sample population sequenced can be missed (Davidson et al., 2012; Gómez-Romero et al., 2018).

ADSSpike data analysis employs thresholds based on high-quality reads, Phred score, coverage, and cut-offs of 50% of reads supporting the SNPs, as previously described (Kishikawa et al., 2019; Lerch et al., 2017; Lerch et al., 2019; Phred Quality Score - an overview | ScienceDirect Topics; Song et al., 2016). Careful evaluation and SNP screening showed these parameters to accurately call signature SNPs pertinent to VOC/VOIs while minimizing the number of false negatives. Whilst employing these parameters for variant calling, a total of 15 SARS-CoV-2 positive samples did not show any SNPs, suggesting either an identical gene to the reference employed, or insufficient to detect mutations. Because of this, the read depth analysis performed across SARS-CoV-2 positive samples suggests that a read depth below 170 is not likely to detect any SNPs. The analysis performed here suggests an average of 700 reads spanning the S gene as a minimum to detect SNPs and correlate them to SARS-CoV-2 variants. Moreover, a workflow is proposed that could increase the likelihood of detecting SNPs, and potentially a VOC/VOI, based on the initial average read depth and coverage (Supplementary Fig. 2).

A limitation of the study is that the recommended 700 read depth described is based on a small sample size (3 VOC/VOI samples) and is therefore subject to a potential selection bias. To summarize, ADSSpike is one of the first workflows to target the SARS-CoV-2 S gene for VOC-VOI detection by NGS amplicon deep sequencing. Additionally, ADSSpike was able to detect the A21583G SNP in 4 independent samples, a mutation that has not been recorded in the literature. This suggests that the presented workflow can identify SNPs in particular population clusters that could be assessed for increased risk, spread, and overall transmissibility. Moreover, the method here proposed can be used as a faster and more feasible approach than WGS, with even greater potential possible through use of a one pot primer-pooled approach.

5. Conclusions

The emergence of SARS-CoV-2 variants calls for molecular surveillance tools. The ADSSpike workflow provides a cost-effective, scalable and practical solution for SARS-CoV-2 variant identification, including VOC/VOIs. Additionally, the experimental design has been validated with the proper controls and parameter tuning to increase its reliability for SNP calling. The primer mixture approach to amplify the S gene may reduce the time to SNP detection, which is imperative during a pandemic.

Data availability

The data that supports the results presented here are available upon reasonable request.

Author contributions

DCM, CK, LO, and DP contributed towards the experimental design and conceptualization. CK and LO performed the laboratory work. DCM performed the bioinformatics and statistical analysis. LF performed the cost analysis. DCM, CK, and LF wrote the original draft. DCM, CK, LO, LF, and DP reviewed and edited the manuscript. DP led and supervised the project. All authors read and approved the final manuscript.

Acknowledgments

We would like to thank the participants of the study as well as the U Calgary DNA sequencing facility and the Illumina support team.

Financial support

This work was funded by Canadian Institutes of Health Research (NFRFR-2019-00015) and Genome261 Canada (Alberta, GC 2020-03-6).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.diagmicrobio.2021.115606.

References

16S Metagenomic Sequencing Library Preparation n.d. [20th July 2021]. Available: https://support.illumina.com/downloads/16s_metagenomic_sequencing_library_preparation.html.

Banada P, Green R, Banik S, Chopoorian A, Streck D, Jones R, et al. A simple RT-PCR melting temperature assay to rapidly screen for widely circulating SARS-CoV-2 variants. MedRxiv 2021. doi: 10.1101/2021.03.05.21252709.

Bezerra MF, Machado LC, De Carvalho V do CV, Docena C, Brandão-Filho SP, Ayres CFJ, et al. A sanger-based approach for scaling up screening of SARS-CoV-2 variants of interest and concern. Infect Genet Evol 2021;92:104910. doi: 10.1016/j.meegid.2021.104910.

Bier C, Edelmann A, Theil K, Schwarzer R, Deichner M, Gessner A, et al. Multi-site evaluation of SARS-CoV-2 spike mutation detection using a multiplex real-time RT-PCR assay. MedRxiv 2021. doi: 10.1101/2021.05.05.21254713.

Boehm E, Kronig I, Neher RA, Eckerle I, Vetter P, Kaiser L, et al. Novel SARS-CoV-2 variants: the pandemics within the pandemic. Clin Microbiol Infect Off Publ Eur Soc Clin Microbiol Infect Dis 2021. doi: 10.1016/j.cmi.2021.05.022.

Cabral GB, Ahagon CM, Lopez-Lopes GIS, Hussein IM, Guimaraes PM, Cilli A, et al. P1 variants and key amino acid mutations at the Spike gene identified using Sanger protocols. MedRxiv 2021. doi: 10.1101/2021.03.21.21253158.

Davidson CJ, Zeringer E, Champion KJ, Gauthier M-P, Wang F, Boonyaratanakornkit J, et al. Improving the limit of detection for Sanger sequencing: a comparison of methodologies for KRAS variant detection. BioTechniques 2012;53:182–8. doi: 10.2144/000113913.

Diaz-Garcia H, Guzmán-Ortiz AL, Angeles-Floriano T, Parra-Ortega I, López-Martínez B, Martínez-Saucedo M, 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.

Fass E, Valenci GZ, Rubinstein M, Freidlin PJ, Rosencwaig S, Kutikov I, et al. HiSpike: a high-throughput cost effective sequencing method for the SARS-CoV-2 spike gene 2021. doi: 10.1101/2021.03.02.21252290.

Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. Q-Bio 2012.arXiv:1207.3907.

Goes LR, Siqueira JD, Garrido MM, Alves BM, Pereira ACPM, Cicala C, et al. New infections by SARS-CoV-2 variants of concern after natural infections and postvaccination in Rio de Janeiro, Brazil. Infect Genet Evol 2021;94: 104998. doi: 10.1016/j.meegid.2021.104998.

Gómez-Romero L, Palacios-Flores K, Reyes J, García D, Boege M, Dávila G, et al. Precise detection of de novo single nucleotide variants in human genomes. Proc Natl Acad Sci 2018;115:5516–21. doi: 10.1073/pnas.1802244115.

Goncalves Cabecinhas AR, Roloff T, Stange M, Bertelli C, Huber M, Ramette A, et al. SARS-CoV-2 N501Y introductions and transmissions in Switzerland from beginning of october 2020 to february 2021—implementation of Swiss-wide diagnostic screening and whole genome sequencing. Microorganisms 2021;9:677. doi: 10.3390/microorganisms9040677.

Hoffmann M, Kleine-Weber H, Schroeder S, Krüger N, Herrler T, Erichsen S, et al. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and Is blocked by a clinically proven protease inhibitor. Cell 2020;181:271–80. doi: 10.1016/j.cell.2020.02.052.

Kalantar KL, Carvalho T, de Bourcy CFA, Dimitrov B, Dingle G, Egger R, et al. IDseq—An open source cloud-based pipeline and analysis service for metagenomic pathogen detection and monitoring. GigaScience 2020;9:. doi: 10.1093/gigascience/giaa111.

Kishikawa T, Momozawa Y, Ozeki T, Mushiroda T, Inohara H, Kamatani Y, et al. Empirical evaluation of variant calling accuracy using ultra-deep whole-genome sequencing data. Sci Rep 2019;9:1784. doi: 10.1038/s41598-018-38346-0.

Lahr DJG, Katz LA. Reducing the impact of PCR-mediated recombination in molecular evolution and environmental studies using a new-generation high-fidelity DNA polymerase. BioTechniques 2009;47:857–66. doi: 10.2144/000113219.

Lerch A, Koepfli C, Hofmann NE, Kattenberg JH, Rosanas-Urgell A, Betuela I, et al. Longitudinal tracking and quantification of individual Plasmodium falciparum clones in complex infections. Sci Rep 2019;9:3333. doi: 10.1038/s41598-019-39656-7.

Lerch A, Koepfli C, Hofmann NE, Messerli C, Wilcox S, Kattenberg JH, et al. Development of amplicon deep sequencing markers and data analysis pipeline for genotyping multi-clonal malaria infections. BMC Genomics 2017;18:864. doi: 10.1186/s12864-017-4260-y.

Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinforma Oxf Engl 2009;25:1754–60. doi: 10.1093/bioinformatics/btp324.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics 2009;25:2078–9. doi: 10.1093/bioinformatics/btp352.

Miller D, Martin MA, Harel N, Tirosh O, Kustin T, Meir M, et al. Full genome viral sequences inform patterns of SARS-CoV-2 spread into and within Israel. Nat Commun 2020;11:5518. doi: 10.1038/s41467-020-19248-0.

Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, et al. Tablet—next generation sequence assembly visualization. Bioinformatics 2010;26:401–2. doi: 10.1093/bioinformatics/btp666.

Pabbaraju K, Wong AA, Douesnard M, Ma R, Gill K, Dieu P, et al. Development and validation of RT-PCR assays for testing for SARS-CoV-2. Off J Assoc Med Microbiol Infect Dis Can 2021: e20200026. doi: 10.3138/jammi-2020-0026.

Phred Quality Score - an overview | ScienceDirect Topics n.d. [22nd July 2021]. Available https://www.sciencedirect.com/topics/biochemistry-genetics-and-molecular-biology/phred-quality-score.

Pillay S, Giandhari J, Tegally H, Wilkinson E, Chimukangara B, Lessells R, et al. Whole genome sequencing of SARS-CoV-2: adapting illumina protocols for quick and accurate outbreak investigation during a pandemic. Genes 2020;11:949. doi: 10.3390/genes11080949.

Quinlan AR, Hall IM. BED tools: a flexible suite of utilities for comparing genomic features. Bioinformatics 2010;26:841–2. doi: 10.1093/bioinformatics/btq033.

COVID-19 ARTIC v3 Illumina library construction and sequencing protocol 2020. [21st December 2021]. Available https://doi.org/10.17504/protocols.io.bgxjjxkn.

Sanger Sequencing Solutions for SARS-CoV-2 Research - CA n.d. [20th July 2021]. Available www.thermofisher.com/ca/en/home/life-science/sequencing/sanger-sequencing/applications/sars-cov-2-research.html. (accessed July 20, 2021).

Schnell IB, Bohmann K, Gilbert MTP. Tag jumps illuminated – reducing sequence-to-sample misidentifications in metabarcoding studies. Mol Ecol Resour 2015;15:1289–303. doi: 10.1111/1755-0998.12402.

Shaibu JO, Onwuamah CK, James AB, Okwuraiwe AP, Amoo OS, Salu OB, et al. Full length genomic sanger sequencing and phylogenetic analysis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Nigeria. PLOS ONE 2021;16: e0243271. doi: 10.1371/journal.pone.0243271.

Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality. Eurosurveillance 2017;22:30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494.

Song K, Li L, Zhang G. Coverage recommendation for genotyping analysis of highly heterologous species using next-generation sequencing technology. Sci Rep 2016;6:35736. doi: 10.1038/srep35736.

Sze MA, Schloss PD. The impact of DNA polymerase and number of rounds of amplification in PCR on 16S rRNA gene sequence data. BioRxiv 2019: 565598. doi: 10.1101/565598.

Wang H, Jean S, Eltringham R, Madison J, Snyder P, Tu H, et al. Mutation-specific SARS-CoV-2 PCR screen: rapid and accurate detection of variants of concern and the identification of a newly emerging variant with spike L452R mutation. J Clin Microbiol 2021;59: e0092621. doi: 10.1128/JCM.00926-21.

Wu F, Zhao S, Yu B, Chen Y-M, Wang W, Song Z-G, et al. A new coronavirus associated with human respiratory disease in China. Nature 2020;579:265–9. doi: 10.1038/s41586-020-2008-3.