

## SARS-CoV-2 spike gene Sanger sequencing methodology to identify variants of concern

Fatimah S Alhamlan<sup>1,2,3</sup> , Dana M Bakheet<sup>2,4</sup>, Marie F Bohol<sup>1</sup>, Madain S Alsanea<sup>1</sup>, Basma M Alahideb<sup>1</sup>, Faten M Alhadeq<sup>1</sup>, Feda A Alsuwairi<sup>1</sup>, Maha A Al-Abdulkareem<sup>1</sup>, Mohamed S Asiri<sup>1</sup>, Reem S Almaghrabi<sup>5</sup>, Sarah A Altamimi<sup>3</sup>, Maysoun S Mutabagani<sup>3</sup>, Sahar I Althawadi<sup>3</sup> & Ahmed A Al-Qahtani<sup>\*,1,2</sup>

<sup>1</sup>Department of Infection & Immunity, King Faisal Specialist Hospital & Research Center, Riyadh, 11211, Saudi Arabia; <sup>2</sup>College of Medicine, Alfaisal University, Riyadh, 11211, Saudi Arabia; <sup>3</sup>Department of Pathology & Laboratory Medicine, King Faisal Specialist Hospital & Research Center, Riyadh, 11211, Saudi Arabia; <sup>4</sup>Center for Genomic Medicine, King Faisal Specialist Hospital & Research Center, Riyadh, 11211, Saudi Arabia; <sup>5</sup>Organ Transplant Center of Excellence, King Faisal Specialist Hospital & Research Center, Riyadh, 11211, Saudi Arabia; \*Author for correspondence: Tel.: +966 114 424 550; aqahatani@kfshrc.edu.sa

BioTechniques 74: 00–00 (February 2023) 10.2144/btn-2021-0114

First draft submitted: 30 November 2021; Accepted for publication: 3 January 2023; Published online: 16 February 2023

### ABSTRACT

The global demand for rapid identification of circulating SARS-CoV-2 variants of concern has led to a shortage of commercial kits. Therefore, this study aimed to develop and validate a rapid, cost-efficient genome sequencing protocol to identify circulating SARS-CoV-2 (variants of concern). Sets of primers flanking the SARS-CoV-2 spike gene were designed, verified and then validated using 282 nasopharyngeal positive samples for SARS-CoV-2. Protocol specificity was confirmed by comparing these results with SARS-CoV-2 whole-genome sequencing of the same samples. Out of 282 samples, 123 contained the alpha variant, 78 beta and 13 delta, which were indicted using in-house primers and next-generation sequencing; the numbers of variants found were 100% identical to the reference genome. This protocol is easily adaptable for detection of emerging variants during the pandemic.

### METHOD SUMMARY

The need for active genomic sequencing surveillance to rapidly identify circulating SARS-CoV-2 variants of concern is critical. Sanger sequencing method using in-house primers is an alternative approach that can be used in facilities with existing equipment.

### KEYWORDS:

alpha • beta • delta • genomic surveillance • oligonucleotide synthesis • PCR • Sanger sequencing • SARS-CoV-2 • VOCs

In March 2020, the WHO declared a global pandemic caused by SARS-CoV-2. As of 26 July 2021, the virus had infected more than 194 million people worldwide and caused the deaths of 4.16 million individuals [1]. As of July 2021, there were four SARS-CoV-2 variants of concern (VOCs) and nine variants under investigation. The VOCs were the alpha (B.1.1.7), beta (B.1.351), gamma (P.1) and delta (B.1.617.2) variants, where the variant alpha had been reported in 178 countries, beta in 123 countries, gamma in 75 countries and delta in 111 countries [2].

SARS-CoV-2 is a single-stranded RNA-enveloped virus. An RNA-based metagenomic next-generation sequencing (NGS) approach was applied to characterize its entire genome, which is 29,881 base pairs in length (GenBank no. MN908947), encoding 9860 amino acids [3]. The spike (S) gene accounts for 1273 amino acids and consists of a signal peptide (amino acids 1–13) located at the N-terminus, an S1 subunit (14–685 residues) and an S2 subunit (686–1273 residues); the last two regions are responsible for receptor binding and membrane fusion, respectively [4]. Thus, the spike gene plays an important role in the attachment of the viral particle to the receptor cells in the host, range determination and membrane fusion [5,6]. We focused on the spike gene for this study because of its many implications in disease transmissibility, infectiousness and immune escape, as well as for its potential as a therapeutic or diagnostic target.

Genomic sequencing is the primary tool for revealing the virus blueprint and genetic code. Given the automation and commercialization of high-throughput DNA sequencing, conducting whole-genome sequencing using NGS technology would be ideal under normal situations with sufficient supplies. However, owing to low supplies during the COVID-19 pandemic and a lack of expertise in this technology in some countries, scientists worldwide have turned to the gold standard in genomic sequencing: Sanger sequencing [7]. Sanger sequencing (with polymerase chain reaction [PCR]) can be conducted using equipment that typically exists in most standard molecular or diagnostic laboratories worldwide. We therefore aimed in this work to provide a rapid and cost-efficient approach to detect circulating VOCs using a protocol that can be applied in any standard molecular or diagnostic laboratory. The present study successfully developed and validated a rapid, cost-efficient genome sequencing protocol to identify circulating SARS-CoV-2 (VOCs).

Table 1. Polymerase chain reaction primer sets for detection of SARS-CoV-2 variants of concern.

Primer name	Position	Forward sequence 5'–3'	Position	Reverse sequence 5'–3'	Amplicon Size (bp)
G1_KFSHRC	21645–21664	ACACTAATTCTTTCACACGT	22101–22082	TCAAGGTCCATAAGAAAAGG	457
G2(A)_KFSHRC	22733–22754	TGCTTACTAATGTCTATGCAG	23226–23207	GACTCAGTAAGAACACCTGT	494
G2(B)_KFSHRC	22882–22903	TCTTGATTCTAAGGTTGGTGGT	23519–23500	CCCCTATTAACAGCCTGCA	638
G2(C)_KFSHRC	23296–23315	TCCACAGACACTTGAGATTC	23873–23854	CTATTCCAGTTAAAGCACGG	578
G3_KFSHRC	24368–24387	GACTCACTTTCTCCACAGC	24993–24974	TCAGGTTGCAAAGGATCATA	626

## Materials & methods

### Design & synthesis of oligonucleotide primers

More than 200 complete SARS-CoV-2 genome sequences were retrieved from the Global Initiative on Sharing Avian Influenza Data database [8] in December 2020 and were aligned using the Clustal W algorithm of the MegAlign module to identify the conserved regions using DNASTAR software (DNASTAR, WI, USA). Oligonucleotide primers were designed for the SARS-CoV-2 spike genes to ensure maximal efficiency and specificity. The desired primers were designed using the consensus sequences from different SARS-CoV-2 isolates from around the world, and the genome sequence from the first virus detected in Wuhan was used as a reference (accession no. MN908947). Primer3 online tools (version 4.1.0) (<https://primer3.ut.ee/>) was used for primer design. Because a successful PCR assay requires efficient and specific amplification, the primers were assessed for several properties, including melting temperature, secondary structure and complementarity. Primers required a guanine–cytosine (GC) content of 50–60%, a melting temperature between 50°C and 65°C, a salt concentration of 50 mM and an oligonucleotide concentration of 300 nM. The specificity of the primers and final sequences were verified using *in silico* prediction analyses with the online Basic Local Alignment Search Tool (BLAST) (<https://blast.ncbi.nlm.nih.gov/blast.cgi>). None of our designed primers showed genomic crossreactivity with other viruses, the human genome or other probable interfering genomes in the BLAST database analysis [9]. Primers, targeting the S gene, were synthesized in-house using Machine MerMade 192E and were purified using desalting method at the Oligonucleotide Synthesis Unit, Center for Genomic Medicine at the King Faisal Specialist Hospital and Research Center (KFSHRC).

### Sample collection & ethical considerations

In total, an identified 282 nasopharyngeal swabs were collected from the The College of American Pathologists (CAP)-accredited diagnostic laboratory in the Pathology and Laboratory Medicine Department at KFSHRC, Saudi Arabia, from December 2020 to April 2021. These samples were confirmed positive for the presence of SARS-CoV-2 with Ct values ranging from 14 to 40 using a Roter-Gene PCR cyclor with RealStar SARS-CoV-2 RT-PCR Kit (Altona Diagnostics, Hamburg, Germany).

### Nucleic acid extraction & cDNA synthesis

In total, 282 samples obtained by nasopharyngeal swab and collected in universal viral transport medium were used in the nucleic acid extraction with a MagMAX Viral/Pathogen Nucleic Acid Isolation Kit (catalog no. A42352, Thermo Fisher Scientific, MA, USA) using a KingFisher Flex system (catalog no. 5400610, Thermo Fisher Scientific). Because cDNA synthesis is required to prepare patient samples for subsequent PCR, SuperScript IV VILO Master Mix (catalog no. 11766050, Thermo Fisher Scientific) was used. Briefly, a mixture of 8 µl of RNA extract (~50 ng/µl), 4 µl of Master Mix and 8 µl of nuclease-free water was incubated at 25°C for 10 min. The RNA was reverse transcribed by incubating the mixture at 50°C for 10 min, and the reaction was terminated at 85°C for 5 min, and then placed on ice.

The quality and the quantity of the RNA and cDNA samples were determined using a NanoDrop spectrophotometer. The ratio of sample absorbance at wavelengths of 260 and 280 nm was obtained to assess the nucleic acid purity (~2.0 for RNA and ~1.8 for cDNA).

### PCR & Sanger sequencing

The primer sets used are given in Table 1. For a total reaction volume of 25 µl, 12.5 µl of 2× GoTaq Master Mix (Promega; Madison, WI, USA), 2 µl of cDNA, 0.25 µl of sense primer (20 µM), 0.25 µl of antisense primer (20 µM) and 10 µl of nuclease-free water were used. The PCR cycle was conducted using a Veriti 96-Well Fast Thermal Cycler (catalog no. 4375305, Thermo Fisher Scientific) as follows: 95°C for 5 min; then 45 cycles of 95°C for 30 s, 58°C for 30 s and 72°C for 30 s; followed by 72°C for 5 min. The positive control was a sample containing SARS-CoV-2, and the negative control was water. All PCR conditions were the same.

The PCR products were separated using 1.5% agarose gel electrophoresis and visualized using UV light (Gel Doc EZ System, Bio-Rad, CA, USA) (Figure 1). The retained PCR products were purified using AMPure XP and then sent for Sanger sequencing to the Core Facility of the Center for Genomic Medicine at KFSHRC. For Sanger sequencing, sequence chromatograms of 282 spike gene sequences were aligned using the Lasergene Suite for sequence analysis (DNASTAR), with the WH-human1 SARS-CoV-2 sequence (MN908947) used as a reference [10]. Forward AB1 file and reverse were added to the SeqMan Pro to be assembled then a contig was generated. Using the MegAlign program, the generated contig was added with the reference sequence then aligned by Clustal W Method to report the mutations.

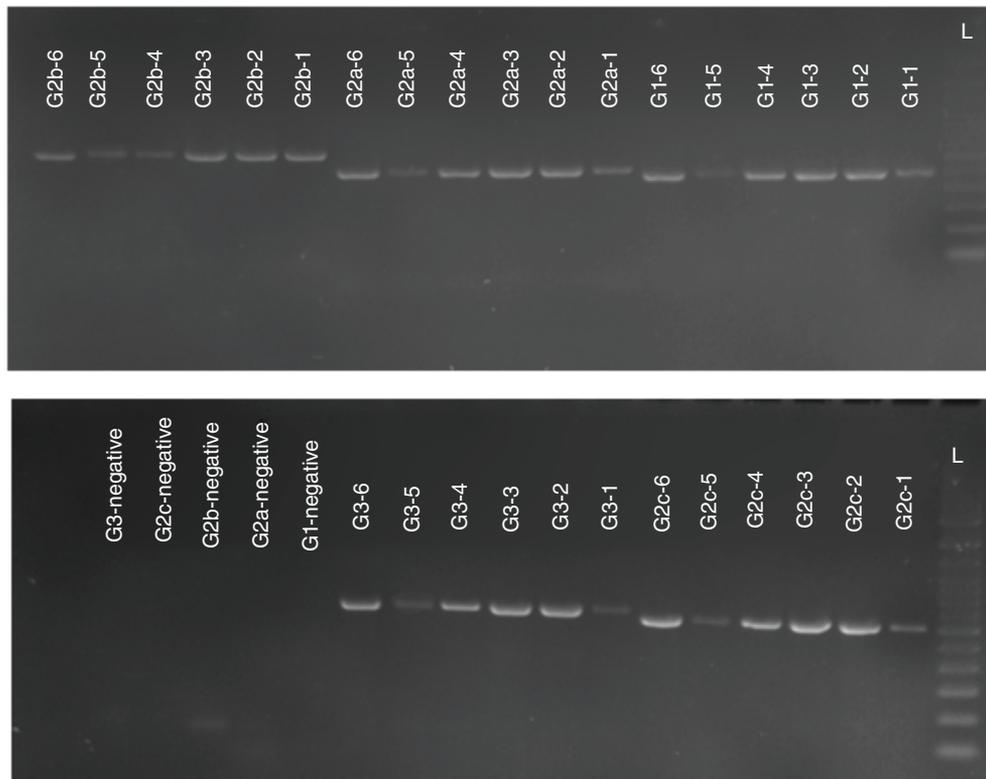


Figure 1. PCR products visualized on a 1.5% agarose gel stained with ethidium bromide. L: 100 bp DNA ladder.

### SARS-CoV-2 whole-genome sequencing

Portions of the same 282 nasopharyngeal samples used for nucleic acid extraction and cDNA synthesis were also subjected to nucleic acid extraction using a MagMAX™ Viral/Pathogen Nucleic Acid Isolation Kit (catalog no. A42352, Thermo Fisher Scientific, MA, USA) for whole-genome sequencing using the Ion AmpliSeq SARS-CoV-2 Research Panel with the Ion Torrent S5 platform by following the manufacturer's instructions (Thermo Fisher Scientific). For SARS-CoV-2 whole-genome sequencing, variants were detected using the S5 Torrent Suite™ software Variant Caller, version 5.12.

### Results & discussion

Our primers and protocol were validated by first determining that the bands with the corresponding amplicon size were visualized for positive control samples, whereas for negative samples, the band was absent. Of samples from 282 patients, 282 samples showed the correct band of the corresponding size; however, four samples with high Ct values ( $\geq 38$ ) showed faint bands. These results indicated that our primers were specific and provided positive results with samples that were previously confirmed as being positive. It is worth mentioning that the band size does not reveal the VOCs infection. Sanger sequencing is required to read out the amplicon and its corresponding VOC. Our results were 100% concordant with those of the diagnostic laboratory at KFSHRC, which used the DiaPlexQ™ Novel Coronavirus (2019-nCoV) Detection Kit. The PCR product of 282 samples, including the samples with faint bands, were sent to the sequencing core facility to perform Sanger sequencing.

The output sequence of 1401 amplicons from 282 patient samples were analyzed using DNASTAR Lasergene Suite. Multiple sequence alignment of the spike gene was conducted with the reference genome (GenBank Protein Accession: QHD43416.1). The sequencing results were as follows: 123 samples contained the alpha variant, 78 samples contained the beta variant, no sample contained the gamma variant (it is not circulating in our community), 13 samples contained the delta (B.1.617.2), 64 samples were non-VOCs that belonged to none of these variants and four samples had a poor sequence. The major amino acid variations detected were as follows: 69–70Del, 144Del, N501Y, A570D, P681H, T716I, S982A and D1118H for the alpha variant; D80A, K417N, E484K, N501Y and A701V for the beta variant; and G142D, 156–157Del, R158G, L452R, T478K, P681R and D950N for the delta variant (Figure 2). Moreover, the remaining samples were non-VOCs that belonged to none of these variants and had 99.97% identity with the reference genome. These mutations are not currently considered VOCs nor variants under investigation. Of note, the D614G mutation was the most predominant in this



cohort. Using the generated primers and developed protocol, we were able to sequence >98% of the positive samples with accurate mutation calls that matched the NGS data.

To test the accuracy of and to validate the PCR and Sanger sequencing results, parallel tests to detect SARS-CoV-2 were conducted with whole-genome sequencing using the Ion AmpliSeq SARS-CoV-2 Research Panel with the Ion Torrent S5 platform. For NGS data analysis, Torrent Suite Server 5.12 was used to check the quality parameters, summary and other primary analysis such as coverage analysis and variant calling. Our samples showed over 70% Ion sphere particle (ISP) density loading across different runs. In the coverage analysis, minimum mapped reads were around 500,000 reads which resulted over 3000× Mean Depth and ≥96% on Target Score for all the samples. Consistent with the results of the in-house designed primers and Sanger sequencing assay, whole-genome sequencing indicated that of 278 samples, 123 samples contained the alpha variant, 78 samples contained the beta variant, no samples contained the gamma variant and 13 samples contained the delta. For the four samples that failed, when sequenced by the Sanger approach, one sample contained the alpha variant and three samples contained a few mutations that were not of concern.

Given the global decreased availability of commercial SARS-CoV-2 sequencing kits, we designed, synthesized and validated SARS-CoV-2 spike gene primer sets to identify current VOCs in an efficient, cost-effective and timely manner. We specifically designed and validated these primer sets to make them available for use in other laboratories worldwide. This is a low-cost, easily performed protocol, and the reagents, laboratory supplies and equipment are currently available as well as already found in most pathology or diagnostic laboratories.

The COVID-19 pandemic has underscored the importance of genomic surveillance as a valuable tool that quickly and effectively informs governments regarding public health decisions. However, these surveillance efforts are scattered globally owing to limited supplies and lack of sequencing infrastructure and trained personnel, especially in low- and middle-income countries [10]. Indeed, a lack of accurate and fast SARS-CoV-2 genomic results may prolong the pandemic. According to the US Centers for Disease Control and Prevention, genomic sequencing is an important approach that allows scientists to identify SARS-CoV-2 and its variants and monitor how they change over time into new variants, to understand how these changes affect the characteristics of the virus and to use this information to better understand how it might impact public health [11]. According to the Global Initiative on Sharing Avian Influenza Data, most countries are falling short on the global repository of SARS-CoV-2 genomic data. Of 180 million confirmed COVID-19 cases, approximately 2 million have been submitted [12]. Thus, our sequencing protocol will help overcome the challenges of limited supplies and provide in-house sequencing primers and protocol that can be performed in many laboratories.

Our whole-genome sequencing results showed high concordance with the PCR/Sanger sequencing results. Both platforms were successful in detecting the alpha, beta and delta VOCs. Recent studies have explored the possibility of using in-house primers and Sanger sequencing to identify SARS-CoV-2 variants. Testing with PCR has been considered the gold standard for viral gene detection, and it is widely used because of the ease in assay design and its relatively low setup and running costs. During the pandemic, several other laboratories have leveraged PCR technology for detection or sequencing to diagnose COVID-19, with infected individuals identified by the successful amplification of the viral genome obtained using nasopharyngeal swabs. Some facilities, especially those in resource-constrained settings, may lack NGS equipment, but most facilities have Sanger technology. Overall, the cost of Sanger sequencing, as a target sequencing, is cheaper in comparison to the whole-genome sequencing. Indeed, Shaibu *et al.* in Nigeria sequenced the whole SARS-CoV-2 genome using Sanger sequencing technology [13]. When researchers in China mapped the genome of the coronavirus isolated from one of the first patients in the Wuhan outbreak in December 2019 and made it available to scientists worldwide, they enabled the global development of diagnostic tests, vaccines and drugs. Using this information, we previously developed in-house detection assay that was approved for emergency use by the Saudi Food and Drug Authority [9].

In addition, when Public Health England announced in December 2020 that their national testing laboratories using a TaqPath detection assay were showing marked increases in spike gene target failures as of late November 2020 [14], we went back to analyze our previous TaqPath assay results that were conducted from June to August 2020 at KFSHRC using the same TaqPath assay. Out of 20,000 tests, we never found negative spike gene results (data not shown). This finding provides supporting evidence for the absence of the alpha variant in our community during that time. Since then, we established our epidemiology and genomic surveillance system in our hospital to monitor the emerging SARS-CoV-2 variants.

A major strength of this study was that by focusing on detecting only the spike gene, the turnaround time was faster than that for a whole-genome sequencing platform, with less than 24 h needed to provide the results for 94 samples. However, there is a major limitation of using conventional PCR and Sanger sequencing over NGS. In contrast to the Sanger method, which can only sequence one DNA fragment at a time, NGS can sequence millions of DNA fragments at once. This procedure results in the simultaneous sequencing of hundreds to thousands of genes. Nonetheless, because Sanger sequencing can produce high-quality DNA sequencing data for sections up to 1000 bases, it is still the method of choice, particularly for the examination of low-volume DNA sequences.

## Conclusion

In conclusion, the ongoing COVID-19 pandemic poses one of the greatest global threats in modern history and has triggered severe human, social and economic costs. Efficient, rapid and cost-effective sequencing protocols will aid in genomic surveillance. Thus, we successfully developed such an assay for detecting SARS-CoV-2 VOCs that can be used by laboratories that cannot afford or cannot obtain commercial high-throughput sequencing kits. We found ≥98% identity between the full genome sequence generated by NGS

and by our developed Sanger sequencing platform. This in-house assay offers a viable alternative approach to increase sequencing capacity. KFSHRC will provide these primers free of charge to research laboratories in Saudi Arabia. Moreover, the developed method may be beneficial beyond the COVID-19 pandemic because it can be easily adapted for use with any emerging microbe.

## Future perspective

The in-house primers and Sanger sequencing protocol provides rapid and cost-effective solutions to detect and track variants. By establishing and validating this workflow, it will be useful to detect and track emerging and reemerging pathogens. The challenge remains in accessing the genomic data in due course. We foresee a continuous threat by SARS-CoV-2 and its variants. Therefore, validating these assays using a target-sequencing approach is the optimal solution for countries' preparedness.

## Author contributions

F Alhamlan: supervision and writing – original draft. D Bakheet: primer synthesis. M Bohol, M Alsanea, B Alahideb, F Alhadeq, F Alsuwairi, M Al-Abdulkareem and M Asiri: molecular experiments. R Almaghrabi: clinical validation. S Altamimi: sample collection. M Mutabagani and S Althawadi: resources and validation processes. AA Alqahtani: conceptualization and project administration. All authors have read and approved the final draft of the article and have approved its submission for publication.

## Acknowledgments

We would like to extend our gratitude to the Sequencing Core Facility, Center for Genomic Medicine, King Faisal Specialist Hospital and Research Center. The support of the Research Center administration at King Faisal Specialist Hospital and Research Center is highly appreciated. We thank the Integrated Gulf BioSystems LLC team led by S Zaem, B Mani and GK Udayaraja who helped with the next-generation sequencing experiments.

## Financial & competing interests disclosure

This study was funded by the King Faisal Specialist Hospital and Research Center COVID-19 grant fund (Research Advisory Committee no. 2200021). The funder had no role in the design of the study and collection, analysis and interpretation of data and in writing the manuscript. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

## Ethical conduct of research

This study was performed in compliance with all applicable national and international ethical guidelines for conducting research on human participants, including in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki), and was approved by the institutional review board at King Faisal Specialist Hospital and Research Center (Institutional Review Board no. 220 0021). This board also granted a waiver for obtaining informed consent owing to the use of de-identified samples for this study.

## Availability of data & materials

Sequences generated in this study are uploaded to the Global Initiative on Sharing Avian Influenza Data.

## Open access

This work is licensed under the Attribution-NonCommercial-NoDerivatives 4.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

## References

1. WHO. Coronavirus disease (COVID-19) pandemic (2020). [www.who.int/emergencies/diseases/novel-coronavirus-2019](http://www.who.int/emergencies/diseases/novel-coronavirus-2019)
2. WHO. Weekly epidemiological update on COVID-19 – 13 July 2021 (2021). <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19---13-july-2021>
3. Chan J, Kok K, Zhu Z *et al.* Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.* 9(1), 221–236 (2020).
4. Huang Y, Chan Y, Xin feng X *et al.* Structural and functional properties of SARS-CoV-2 spike protein: potential antiviral drug development for COVID-19. *Acta Pharmacol. Sin.* 41(9), 1141–1149 (2020).
5. Heald-Sargent T, Gallagher T. Ready, set, fuse! The coronavirus spike protein and acquisition of fusion competence. *Viruses* 4(4), 557–580 (2012).
6. Qiang XL, Xu P, Fang G *et al.* Using the spike protein feature to predict infection risk and monitor the evolutionary dynamic of coronavirus. *Infect. Dis. Poverty.* 9(1), 33 (2020).
7. Slatko BE, Gardner AF, Ausubel FM. Overview of next-generation sequencing technologies. *Curr. Protoc. Mol. Biol.* 122(1), e59 (2018).
8. Global Initiative on Sharing Avian Influenza Data. [www.epicov.org/epi3/cfrendend#lightbox-814829872](http://www.epicov.org/epi3/cfrendend#lightbox-814829872)
9. Alhamlan FA, A-Q A, Bakheet DM *et al.* Development and validation of an in-house, low-cost SARS-CoV-2 detection assay. *J. Infect. Public Health* 14(9), 1139–1143 (2021).
10. Cyranoski D. Alarming COVID variants show vital role of genomic surveillance. *Nature* 589(7842), 337–338 (2021).
11. Centers for Disease Control and Prevention. What is genomic surveillance? (2021). [www.cdc.gov/coronavirus/2019-ncov/variants/genomic-surveillance.html#:~:text=Scientists%20use%20genomic%20sequencing%20to,a%20process%20called%20genomic%20surveillance](http://www.cdc.gov/coronavirus/2019-ncov/variants/genomic-surveillance.html#:~:text=Scientists%20use%20genomic%20sequencing%20to,a%20process%20called%20genomic%20surveillance)
12. Global Initiative on Sharing Avian Influenza Data. 2,385,336 sequence entries with complete collection date information shared via GISAID by July 26 2021 (2021). <https://gisaid.org/databases/features/covsurver-mutations-app/>

13. Shaibu JO, Onwuamah CK, James AB et al. Full length genomic sanger sequencing and phylogenetic analysis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in Nigeria. *PLoS ONE* 16(1), e0243271 (2021).
14. Public Health England. Investigation of novel SARS-COV-2 variant. Variant of concern 202012/01 (2020). [www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201](http://www.gov.uk/government/publications/investigation-of-novel-sars-cov-2-variant-variant-of-concern-20201201)