Received: 2 March 2022 | Revised: 7 September 2022 | Accepted: 13 September 2022

DOI: 10.1002/hsr2.856

ORIGINAL RESEARCH

Health Science Reports Open Access WILEY

# Molecular characterization of SARS-CoV-2 from the saliva of patients in the Republic of Korea in 2020

Se Hun Gu[1] | Dong Hyun Song[1] | Hyeongseok Yun[1] | Jung-Eun Kim[1] | Seung-Ho Lee[1] | Hyunjin Lee[1] | Tae Ho Lee[2] | Seol Muk Kang[3] | Yu Sub Jung[3] | Gyeunghaeng Hur[1] | Daesang Lee[1]

[1]Chem-Bio Technology Center, Agency for Defense Development, Daejeon, Republic of Korea

[2]Institute of Health and Environment, Daejeon, Republic of Korea

[3]Defense Test and Evaluation Research Institute, Agency for Defense Development, Taean, Republic of Korea

**Correspondence**

Daesang Lee, Chem-Bio Technology Center, Agency for Defense Development, Yuseong, P.O. Box 35, Daejeon 34186, Republic of Korea.
Email: dslee@add.re.kr

**Funding information**

Agency for Defense Development, Grant/Award Number: 211555-912821001

## Abstract

**Background and aims:** Despite global vaccination efforts, the number of confirmed cases of coronavirus disease 2019 (COVID-19) remains high. To overcome the crisis precipitated by the ongoing pandemic, characteristic studies such as virus diagnosis, isolation, and genome analysis of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) are necessary. Herein, we report the isolation and molecular characterization of SARS-CoV-2 from the saliva of patients who had tested positive for COVID-19 at Proving Ground in Taean County, Republic of Korea, in 2020.

**Methods:** We analyzed the whole-genome sequence of SARS-CoV-2 isolated from the saliva samples of patients through next-generation sequencing. We also successfully isolated SARS-CoV-2 from the saliva samples of two patients by using cell culture, which was used to study the cytopathic effects and viral replication in Vero E6 cells.

**Results:** Whole-genome sequences of the isolates, SARS-CoV-2 ADD-2 and ADD-4, obtained from saliva were identical, and phylogenetic analysis using Bayesian inference methods showed SARS-CoV-2 GH clade (B.1.497) genome-specific clustering. Typical coronavirus-like particles, with diameters of 70–120 nm, were observed in the SARS-CoV-2 infected Vero E6 cells using transmission electron microscopy.

**Conclusion:** In conclusion, this report provides insights into the molecular diagnosis, isolation, genetic characteristics, and diversity of SARS-CoV-2 isolated from the saliva of patients. Further studies are needed to explore and monitor the evolution and characteristics of SARS-CoV-2 variants.

**KEYWORDS**

isolation, Korea, next-generation sequencing, saliva, SARS-CoV-2

Se Hun Gu and Dong Hyun Song contributed equally to this study.

## 1 | INTRODUCTION

Coronaviruses are enveloped viruses containing a positive-sense single-stranded RNA genome and are transmitted among humans and animals. Most humans are frequently infected with common human coronaviruses, namely, 229E, NL63, OC43, and HKU1, which cause mild to lethal respiratory illness.[1] However, some human coronaviruses belonging to the genus Betacoronavirus, such as Middle East respiratory syndrome coronavirus (MERS-CoV), severe acute respiratory syndrome coronavirus (SARS-CoV), and SARS-CoV-2, are highly pathogenic.[1] Its genome contains six functional open reading frames (ORFs), encoding replicase (ORF1a/ORF1b), spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins, and seven putative ORFs encoding accessory proteins interspersed between the structural proteins.[1] SARS-CoV-2 shares 79% and 50% genome sequence identity with SARS-CoV and MERS-CoV, respectively.

SARS-CoV-2 causes an acute respiratory illness called coronavirus disease 2019 (COVID-19).[2] The COVID-19 outbreak began in December 2019 and was declared as a pandemic by the World Health Organization on March 11, 2020. As of May 30, 2022, the number of confirmed COVID-19 cases worldwide exceeded 526 million, with over 6.2 million related deaths.[3] In the Republic of Korea, 18,103,638 cases, including 24,176 deaths, were confirmed as of May 31, 2022.[4] Currently, several vaccines have been developed and approved by the Food and Drug Administration and European Medicines Agency.[5] Although more than 11.8 billion vaccine doses have been administered globally as of May 23, 2022,[3] herd immunity has not been achieved because of the insufficient supply of vaccines and the emergence of new virus variants. Therefore, epidemiological investigations and surveillance of novel variants of SARS-CoV-2 are crucial to managing the COVID-19 pandemic.

Over 11,135,020 genome sequences of SARS-CoV-2 from inpatients worldwide have been deposited at the global initiative on sharing all influenza data (GISAID) EpiCov public databases. These data have been subjected to extensive phylogenetic analyses to detect the emergence of variants and evaluate the geographical spread of the virus, population dynamics of transmission, and adaptation of the virus.[6] Epidemiological studies focusing on viral genomics have substantially contributed to the identification of novel SARS-CoV-2 variants such as B.1.1.7 (Alpha), B.1.351 (Beta), B.1.1.248/B1.1.28/P.1 (Gamma), B.1.617/B.1.617.2 (Delta) and B.1.1.529 (Omicron) in the United Kingdom, South Africa, Brazil, India, and South Africa, respectively. Among them, several variants have been classified as variants of concern, and are reportedly more infectious and pathogenic than wild-type SARS-CoV-2 strain.[7–10]

For the epidemiological and diagnostic study of newly emerging viruses, virus isolation through cell culture is the most important technique. SARS-CoV-2 has reportedly been isolated from various specimens such as nasopharyngeal swabs, sputum, and saliva samples of patients with COVID-19.[11,12] According to a recent study, saliva has been demonstrated as a fine specimen for the detection and isolation of SARS-CoV-2.[13,14] COVID-19 can be diagnosed using saliva without the need for specialized consumables and training for sample collection.[15] Saliva sampling has many advantages over nasopharyngeal and/or oropharyngeal swab sampling methods. Recently, detection of SARS-CoV-2 using saliva has been preferentially suggested over other methods owing to the noninvasiveness of the method, ease of self-administration, and reduction in the need for personal protective equipment to be used by healthcare workers while collecting samples.[16] Furthermore, compared to nasopharyngeal specimens, salivary specimens have shown comparable or higher sensitivities and are considered reliable for the diagnosis of both symptomatic and asymptomatic individuals.[15–18]

The COVID-19 pandemic has had a major effect on human health and society. Multiple SARS-CoV-2 variants have been reported to date, and the virus constantly changes through mutation. Major efforts are underway to identify emerging SARS-CoV-2 variants and characterize their virulence and transmission. Thus, extensive disease surveillance should include an analysis of genetic differences between isolates. Herein, we report the isolation, whole-genome sequencing, and phylogenetic and genome mutation analyses of SARS-CoV-2 (GH clade; B.1.497) from the saliva samples obtained from patients with a confirmed diagnosis of COVID-19 in Taean County, Chungcheongnam Province, Republic of Korea.

## 2 | MATERIALS AND METHODS

### 2.1 | Patients and samples

On December 2020, there were four suspected cases of COVID-19 in the Anheung Test Range Complex, Taean County; the patients had a fever accompanied by a headache, chills, sore throat, myalgia, and pneumonia. Nasopharyngeal swab and saliva samples were obtained from the patients during quarantine or hospitalization and stored at −70°C until processing at the Institute of Health and Environment, the Korea Disease Control and Prevention Agency (KDCA), and the Agency for Defense Development (ADD), respectively. Saliva samples were collected by directing patients to spit in sterile tissue bottles and transported to ADD for further investigation. This study was reviewed and approved by the Institutional Review Board of the Armed Forces Medical Command, Republic of Korea (AFMC-202104-HR-030-01).

### 2.2 | Virus isolation and propagation

Vero E6 cells (CRL 1586; American Type Culture Collection) were infected with the virus isolated from the saliva samples of patients confirmed to have COVID-19 as previously described.[19] Briefly, subconfluent monolayers of Vero E6 cells grown in T25 flasks were inoculated with filtered saliva samples in DMEM supplemented with 2% heat-inactivated fetal bovine serum (Thermo Fisher Scientific), 2 mM L-glutamine, and antibiotics at 37°C for 3 days in a 5% $CO_2$ incubator. The supernatants from the saliva-inoculated cultures were collected to assess virus replication and infectivity using quantitative

TaqMan real-time polymerase chain reaction (TaqMan qRT-PCR) and plaque assays, respectively.

## 2.3 | RNA extraction and TaqMan qRT-PCR

The total RNA was extracted from the saliva samples using a QIAamp Viral RNA Mini Kit (QIAGEN) according to the manufacturer's instructions and then analyzed using TaqMan qRT-PCR. The E gene of SARS-CoV-2 was detected by TaqMan qRT-PCR using the previously described primer and probe.[20] The probe was labeled with reporter dye FAM at the 5′-end and quencher dye BHQ-1 at the 3′-end. Each 20 µl of the reaction contained 1 µl of RNA, 5 µl of TaqPath™ 1-step Multiplex Master Mix (Thermo Fisher Scientific), 0.5 µl of each primer (36 µM), 0.5 µl of probe (10 µM), and 12.5 µl of UltraPure™ DNase/RNase-Free Distilled Water (Thermo Fisher Scientific). The reaction was performed at 50°C for 30 min and 95°C for 10 min, followed by 40 cycles at 95°C for 5 s, and 60°C for 30 s on a QuantStudio 6 Flex Real-Time PCR system (Thermo Fisher Scientific). RNA concentration was determined using a NanoDrop spectrophotometer (Thermo Fisher Scientific).[21]

## 2.4 | Plaque assay

To confirm virus isolation and quantify virus particles, a plaque assay was performed using Vero E6 cells in six-well culture plates. Briefly, monolayers of Vero E6 cells were inoculated with 10-fold diluted virus and incubated at 37°C for 1.5 h in a 5% $CO_2$ incubator. The supernatant was removed, and the cells were carefully overlaid with solution. After 3 days of incubation, the plates were fixed and inactivated using formaldehyde and stained with crystal violet.[22]

## 2.5 | Thin-section electron microscopy

SARS-CoV-2-infected Vero E6 cells were pelleted and fixed with 2% paraformaldehyde and 2.5% glutaraldehyde in 0.1 M phosphate buffer, pH 7.4, overnight. The cells were then washed twice in 0.1 M phosphate and then subjected to postfixation treatment with 2% osmium tetroxide in 0.1 M phosphate buffer for 2 h. The infected cells were dehydrated in a graded series of ethanol (60%, 70%, 80%, 90%, 95%, and 100%), washed with propylene oxide, and embedded in Epon-Araldite resin. Ultra-thin sections (60–70 nm) were obtained using a Leica EM UC7 ultramicrotome (Leica Microsystems) and double-stained with uranyl acetate and lead citrate. Grids were observed under a Hitachi H-7650 high-contrast/high-resolution transmission electron microscope operated at 100 kV (Hitachi).

## 2.6 | Next-generation sequencing (NGS)

Target capture-based enrichment was used for SARS-CoV-2 library preparation. The target capture custom probes were designed to cover the entire sequence of the reference strain SARS-CoV-2 (Wuhan-Hu-1; NC_045512). Overall, 745 biotinylated probes were designed. Each probe was 120 bp with ×3 tiling, and the total probe size was 29.9 kb (Celemics). TruSeq RNA Library Prep for Enrichment (Illumina) was used for library preparation. The reaction was performed with 10–100 ng of total RNA extracted from the saliva. All RNA samples were reverse-transcribed into cDNA, and adapters were ligated to the ends using the dual indices set (Illumina). The adapter-ligated libraries and amplified libraries were purified using AMPure XP beads (Beckman Coulter). The library quality and concentration were determined using TapeStation 4200 and D1000 ScreenTape (Agilent Technologies). The libraries were quantified using the KAPA Library Quantification kit (KAPA Biosystems) on a QuantStudio 6 Flex Real-time PCR system (Thermo Fisher Scientific). The final library concentration was 4–14 pM, with 1–10% PhiX control v3 (Illumina). NGS of the enriched samples was performed on a MiSeq and Nextseq. 500 benchtop platform (Illumina) using MiSeq Reagent v2 Kit 300 Cycle (2 × 150 bp) and Nextseq. 500/550 Mid Output Kit v2.5 (Illumina).

## 2.7 | Extraction of viral genomic sequences

FASTQ adapter trimming and dual-index filtering were performed using in-house scripts. To acquire high-quality genomic sequences, low-quality reads (Phred quality score <Q30) were removed using FaQCs software (version 1.33). Human GRCh38, PhiX, and the NCBI Bacteria RefSeq database were used to remove host and bacterial reads from the FASTQ files. The qualified reads were aligned to the SARS-CoV-2 reference genome (GenBank No. NC_045512) using the Burrows-Wheeler Aligner (version 0.7.12) tool. Consensus sequences were generated using Samtools (version 0.1.19), vcftools, and CLC Genomics Workbench version 7.5.2 (CLC Bio). Positions lacking a minimum of ×20 read depth coverage were treated as missing (named as "N"). NUCmer (version 3.1), and in-house scripts were used to detect nucleotide mutations in consensus genome sequences.

## 2.8 | Sequence alignment and phylogenetic analysis

SARS-CoV-2 sequences were classified into different lineages using the PANGOLIN webtool (Pango v.4.0.6), Nextclade (v.2.2.0), and phylogenetic analysis. The viral genome sequences were aligned using the MAFFT program (v.7.450)[23] and trimmed using the software AliView (v.1.28).[24] The best-fit substitution model was generated by applying the maximum likelihood method and using MEGA 11.[25] The BEAST package (v.1.10.4) with Markov chain Monte Carlo (MCMC) analysis as the Bayesian inference method was implemented to generate phylogenetic trees.[26] The MCMC chain length was established to 100 million states by sampling every 50,000 states. The parameter results showed sufficient sample sizes (effective sample sizes [ESS] > 200). Network software (version

10.2.0.0, https://www.fluxusengineering.com) was used for phylogenetic network analysis. A maximum clade credibility tree was deduced using TreeAnnotator (V.1.10.4) and FigTree (V.1.4.0).

## 3 | RESULTS

### 3.1 | Epidemiological analysis

In late November 2020, the third wave of COVID-19 hit Korea, with outbreaks reportedly starting from pubs in the northern area of Daejeon Metropolitan City. At that time, patient DJ512 contracted SARS-CoV-2, likely in the local pub, and the infection spread to his family, including ADD-1. DJ512, the older brother of ADD-1, complained of fever on November 30 and tested positive for COVID-19 on December 2. All his family, including his mother, wife, and children, tested positive for COVID-19 on December 3. On December 2, KDCA classified and notified ADD-1 as a primary close contact because he had lunch with DJ512 for 30 min on November 29 in Daejeon Metropolitan City. ADD-1 tested positive for COVID-19 on December 3 at a public health center in Taean. The Anheung Test Range Complex staff immediately traced the contact history of ADD-1 with other workers after they were notified that ADD-1 had tested positive for COVID-19. More than 100 people who had come in direct or indirect contact with ADD-1 were screened for SARS-CoV-2 infection by qRT-PCR. Those who had close contact with ADD-1 were quarantined in their homes or dormitories for 14 days, and Proving Ground was closed for 3 days. ADD-2 and ADD-3, who were in contact with ADD-1 during several meetings, business trips, and dinners from November 30 to December 2, tested positive for COVID-19 on December 4 and 5, respectively. ADD-4, who had shared the same office as that of ADD-1 for ~8 h on November 30, was confirmed as being COVID-19-positive, 13 days after ADD-1 was diagnosed with COVID-19.

### 3.2 | Identification of SARS-CoV-2 isolated from patient saliva

Samples from four patients (ADD-1, ADD-2, ADD-3, and ADD-4) suspected to have contracted COVID-19 were analyzed using TaqMan qRT-PCR, and the initial cycle threshold (Ct) values were 34.1, 18.0, 36.0, and 23.0 for ADD-1, ADD-2, ADD-3, and ADD-4, respectively (Table 1).

To isolate SARS-CoV-2 strains (designated as SARS-CoV-2/human/KOR/ADD-1/2020, SARS-CoV-2/human/KOR/ADD-2/2020, SARS-CoV-2/human/KOR/ADD-3/2020, and SARS-CoV-2/human/KOR/ADD-4/2020), saliva samples from the four patients who had tested positive for COVID-19, as confirmed using TaqMan qRT-PCR, were inoculated into Vero E6 cell (CRL 1586) after filtration using a 0.22 μm syringe filter. Five days after inoculation, cytopathic effects (CPE) were observed in Vero E6 cells treated with saliva from ADD-2 and ADD-4. At 7 days after inoculation, we harvested the supernatants from ADD-2- and ADD-4-treated cultures. We then performed blind passage of the supernatants into new flasks with monolayers of fresh Vero E6 cells. The SARS-CoV-2 isolates from ADD-2 (ADD-2i) and ADD-4 (ADD-4i) induced CPE, including rounding and detachment of the cells 2 days after re-infection (Figure 1A,B). To confirm the isolation of the virus, we harvested the supernatants and cells after 4 days of infection with SARS-CoV-2 ADD-2i and ADD-4i (Figure 1A,B). The viral titers of ADD-2i and ADD-4i in the culture supernatants were $6.4 \times 10^5$ and $8.2 \times 10^5$ plaque-forming units/ml, respectively (Figure 1A,B). ADD-2i and ADD-4i particles were evaluated using thin-section transmission electron microscopy in the early postinfection time point. Typical coronavirus-like particles of diameters 70–120 nm were observed (Figure 1A,B). In general, SARS-CoV-2 has a spiky round crown-like structure with a diameter of ~100 nm.[1]

### 3.3 | Whole-genome sequencing of SARS-CoV-2 using NGS

Libraries were prepared using SARS-CoV-2-specific probes and RNA extracted from the saliva samples. We extracted SARS-CoV-2 RNA from the saliva samples (ADD-2 and ADD-4) and isolates (ADD-2i and ADD-4i) and obtained whole-genome sequences using NGS (Table 2). There was no difference in the viral sequences derived from the isolates and the original saliva samples. We designed 745 probes for target capture-based enrichment using the sequence of the SARS-CoV-2 Wuhan-Hu-1 strain as a template. Total raw reads were qualified using the Q30 score and trimmed for reference mapping, and consensus sequences were extracted using the Samtools, vcftools, and CLC Genomics Workbench (CLC Bio). The depth of coverage was calculated by the number of mapped reads. NGS data

**TABLE 1** Characteristics of four COVID-19 clinical samples collected

| ID | Date of onset | Date of collection | Symptoms at diagnosis | Source | Ct value | Virus isolation |
|---|---|---|---|---|---|---|
| ADD-1 | December 03, 2020 | December 16, 2020 | Fever, headache, chill | Saliva | 34.1 | - |
| ADD-2 | December 03, 2020 | December 04, 2020 | Fever, headache, sore throat, pneumonia | Saliva | 18.0 | + |
| ADD-3 | December 06, 2020 | December 16, 2020 | Fever, headache, chill, myalgia | Saliva | 36.0 | - |
| ADD-4 | December 14, 2020 | December 16, 2020 | Fever, headache, | Saliva | 23.0 | + |

Abbreviations: ADD, Agency for Defense Development; Ct, cycle threshold.

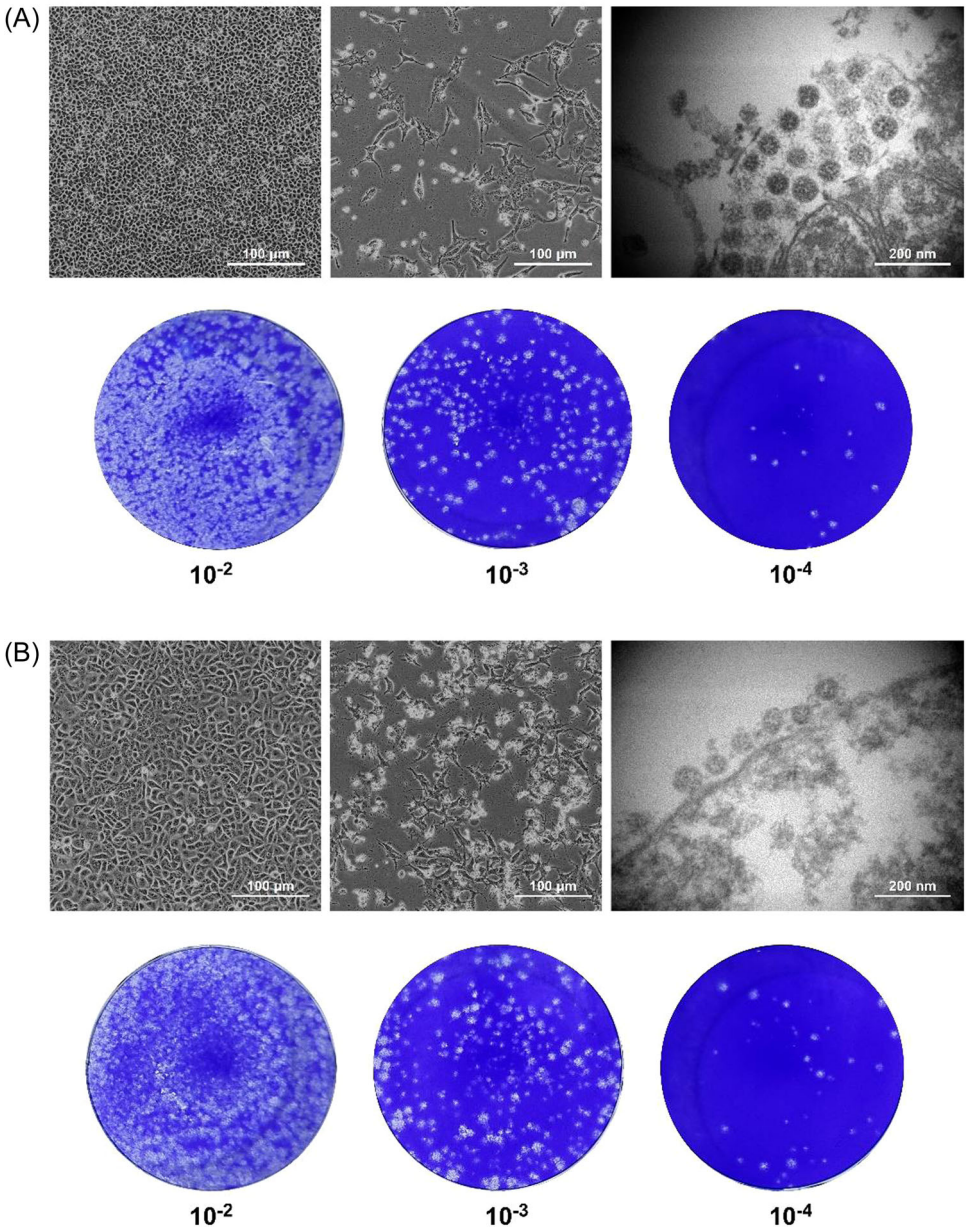


**FIGURE 1** Cytopathic effects (CPE) and thin-section electron micrograph of SARS-CoV-2 strains ADD-2i (A) and ADD-4i (B) in Vero E6 cells. CPE was observed 4 days after infection with ADD-2i (A, upper middle) and ADD-4i (B, upper middle). Normal Vero E6 cells showed no evident CPE (A, upper left and B, upper left). Thin-section transmission electron micrograph, showing crown-like particles in Vero E6 cells inoculated with SARS-CoV-2 isolate ADD-2i (A, upper right) and ADD-4i (B, upper right). Magnification bars are shown on the micrographs. Plaque assay shows infectious SARS-CoV-2 strain ADD-2i (A, lower) and ADD-4i (B, lower) inoculated on Vero E6 cells. Each well represents the virus quantitation of infectious plaque at serial dilutions from $10^{-2}$ to $10^{-4}$ of the viral isolate stocks. ADD, Agency for Defense Development; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

obtained using NextSeq from the saliva sample (ADD-2) generated 6,040,442 mapped reads (base coverage: 99.94%); the viral isolate (ADD-2i) showed 100% base coverage with the reference strain. SARS-CoV-2 from ADD-4 and ADD-4i using MiSeq showed 99.88% and 100% genome coverage with the reference strain, respectively (Table 2). The genomic sequences determined in this study have been deposited in GenBank under accession numbers MZ054402 (ADD-1), MZ004104 (ADD-2), MZ004105 (ADD-2i), MZ099651 (ADD-3), MZ004106 (ADD-4), MZ004107 (ADD-4i), and MZ004152 (DJ512).

### 3.3.1 | Genome mutation and phylogenetic analyses

We identified mutations in the consensus genome sequences (nucleotide changes supported by >50% of the reads) and compared them with the genome of the reference strain SARS-CoV-2 (NC_045512). The genomes of ADD-2, ADD-2i, ADD-4, ADD-4i, and DJ512 belong to the GH clade (B.1.497), with two specific nucleotide mutations (A23403G and G25563T) resulting in amino acid mutations of the spike protein (D614G) and NS3 (ORF3a)

**TABLE 2** Summary of next-generation sequencing results

| ID | Sample source | NGS platform | Raw reads | Total reads after host removal and filter | Mapped reads to reference NC_045512 | Base coverage (%) | Average depth (x) | Genome length (in base pairs) |
|---|---|---|---|---|---|---|---|---|
| ADD-2 | Saliva | NextSeq | 11,228,886 | 6,109,600 | 6,040,442 | 99.94 | 22752.84 | 29,887 |
| ADD-2i | Isolate | NextSeq | 11,503,554 | 7,892,673 | 7,650,559 | 100 | 16037.29 | 29,903 |
| ADD-4 | Saliva | MiSeq | 1,214,036 | 298,062 | 178,362 | 99.88 | 472.74 | 29,868 |
| ADD-4i | Isolate | MiSeq | 1,440,502 | 1,194,811 | 1,046,286 | 100 | 2997.29 | 29,903 |
| DJ512 | Nasopharyngeal swab | MiSeq | 2,693,840 | 2,467,734 | 113,506 | 99.81 | 4074.79 | 29,855 |

Abbreviations: ADD, Agency for Defense Development; NGS, Next-generation sequencing.

**TABLE 3** Nucleotide mutations of SARS-CoV-2 genomes obtained in this study

| ID | Position | Reference codon | Substitution codon | AA in ref. | AA in sub. | Synonymous | Product (A.A. mutation) |
|---|---|---|---|---|---|---|---|
| DJ512, ADD-2, ADD-2i, ADD-4, ADD-4i | 241 | C | T | | | | Intergenic region |
| | 1059 | ACC | ATC | T | I | No | GU280_gp01:orf1a polyprotein (T265I) |
| | 3037 | TTC | TTT | F | F | Yes | GU280_gp01:orf1a polyprotein |
| | 7086 | ACT | ATT | T | I | No | GU280_gp01:orf1a polyprotein (T2274I) |
| | 11,916 | TCA | TTA | S | L | No | GU280_gp01:orf1a polyprotein (S3884L) |
| | 14,408 | CCT | CTT | P | L | No | GU280_gp01:orf1ab polyprotein (P314L) |
| | 16,650 | CTC | CTT | L | L | Yes | GU280_gp01:orf1ab polyprotein |
| | 18,027 | GTG | GTT | V | V | Yes | GU280_gp01:orf1ab polyprotein |
| | 20,675 | CAA | CTA | Q | L | No | GU280_gp01:orf1ab polyprotein (Q2403L) |
| | 20,679 | CCG | CCT | P | P | Yes | GU280_gp01:orf1ab polyprotein |
| | 23,403 | GAT | GGT | D | G | No | GU280_gp02:Spike protein (D614G) |
| | 25,563 | CAG | CAT | Q | H | No | GU280_gp03:ORF3a protein (Q57H) |
| | 28,606 | TAC | TAT | Y | Y | Yes | GU280_gp10:nucleocapsid phosphoprotein |
| | 29,179 | CCG | CCT | P | P | Yes | GU280_gp10:nucleocapsid phosphoprotein |
| | 29,745 | G | T | | | | Intergenic region |
| | 29,755 | G | T | | | | Intergenic region |
| | 29,779 | G | T | | | | Intergenic region |
| ADD-2, ADD-2i, ADD-4, ADD-4i | 1623 | ATA | ACA | I | T | No | GU280_gp01:orf1a polyprotein (I453T) |
| | 1916 | ACT | GCT | T | A | No | GU280_gp01:orf1a polyprotein (T551A) |
| | 2144 | GTC | ATC | V | I | No | GU280_gp01:orf1a polyprotein (V627I) |
| | 2701 | ACC | ACT | T | T | Yes | GU280_gp01:orf1a polyprotein |
| | 29,095 | TTC | TTT | F | F | Yes | GU280_gp10:nucleocapsid phosphoprotein |

Abbreviations: ADD, Agency for Defense Development; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

protein (Q57H), respectively. The DJ512 strain harbored 17 genetic mutations, sharing identical mutations with ADD-2, ADD-2i, ADD-4, and ADD-4i. ADD-2, ADD-2i, ADD-4, and ADD-4i harbored the same 22 mutations, including five more mutations than the DJ512 strain (Table 3). The phylogenetic analyses of full-length genome sequences of SARS-CoV-2 stains from GISAID (Supporting Information: Table S1), including ADD-2, ADD-2i, ADD-4, ADD-4i, and DJ512, with MCMC analysis as the Bayesian inference method, showed that the GH clade (B.1.497) shared a common ancestor (Figure 2). The SARS-CoV-2 strains ADD-2 and -4 from the saliva and the isolates (ADD-2i and ADD-4i) showed identical sequences. Partial genomic sequences were obtained from the low-titer samples,

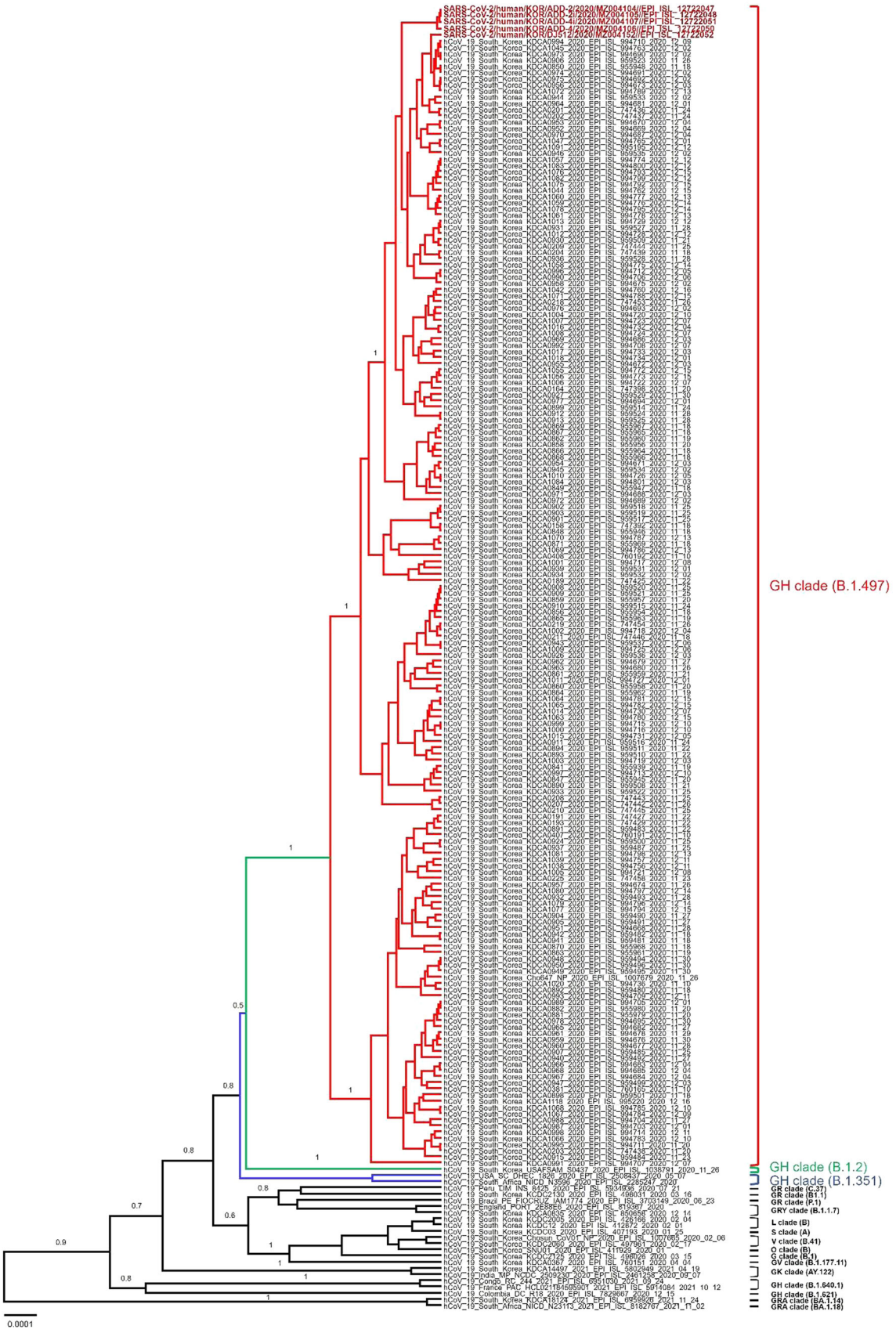


**FIGURE 2** The phylogenetic tree constructed using 210 SARS-CoV-2 complete genome sequences from GISAID. Phylogenetic inferences were performed using BEAST (1.10.4 version) with default priors and assuming homochromous tips. The Markov chain Monte Carlo method was used until effective sample sizes (ESS > 200) were obtained. The maximum clade credibility tree was summarized using a 10% burn-in by TreeAnnotator (1.10.4 version). GISAID, global initiative on sharing all influenza data; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

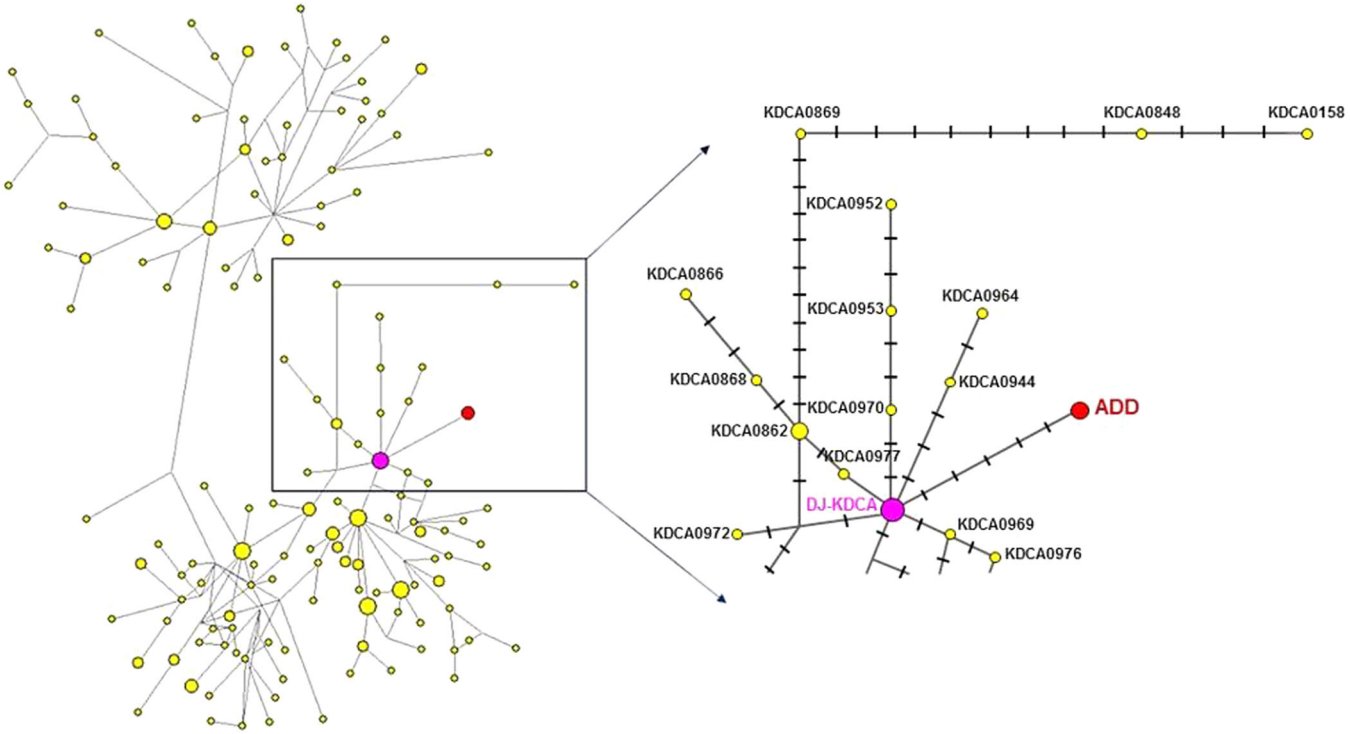


**FIGURE 3** Phylogenetic network of 188 SARS-CoV-2 genomes. The pink circle cluster, named DJ-KDCA, contained five members (DJ512, KDCA0850, KDCA0956, KDCA0974, and KDCA0975). The red circle cluster named ADD consisted of four members (ADD-2, ADD-2i, ADD-4, and ADD-4i). DJ-KDCA was the parent type of the ADD cluster. The ADD cluster had five more mutations than the DJ-KDCA cluster and was directly related to DJ-KDCA. A mutated nucleotide position is represented by each notch on the link, and the circle area is proportional to the number of taxa. ADD, Agency for Defense Development; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2.

ADD-1 and ADD-3. In addition, to understand SARS-CoV-2 transmission routes in Proving Ground, we constructed a phylogenetic network with 188 SARS-CoV-2 genome sequences, including 183 genomes extracted from GISAID (Supporting Information: Table S1) and 5 genomes from the current study. DJ512 was clustered within the pink circle, labeled as DJ-KDCA, which contained four additional viral genomes, namely, KDCA0850, KDCA0956, KDCA0974, and KDCA0975 (Figure 3). Network analysis indicated that the DJ-KDCA cluster was the parent type of ADD. Its members KDCA0970, KDCA0977, KDCA0944, and KDCA0969 were also related to another outbreak path. Although we failed to obtain the whole-genome sequences of ADD-1 and ADD-3, phylogenetic analysis suggested that the ADD-2 and ADD-4 variants originated from the DJ512 variant.

## 4 | DISCUSSION

SARS-CoV-2 can be diagnosed in specimens such as nasopharyngeal swabs, sputum, and saliva.[11,12] Among them, nasopharyngeal swabbing is the most commonly used method; however, it has recently been reported that diagnosis using saliva is also effective.[15,27] The ability to detect SARS-CoV-2 in saliva over a 7-day period presents saliva testing as an initial screening test in point-of-care testing and self-diagnosis.[11,15] SARS-CoV-2 has been successfully isolated from saliva, but analysis of the virus characteristics was not reported.[11] In this study, we demonstrated that saliva from patients with a confirmed diagnosis of COVID-19 is suitable clinical samples for molecular characterization, including disease diagnosis, virus isolation, genetic characteristic determination, and SARS-CoV-2 variant identification. In addition, we detected viral RNA in the saliva of four patients and isolated SARS-CoV-2 from two saliva samples. However, it is difficult to isolate the virus from low-titer saliva samples (Ct values of 34 or higher).

Currently, epidemiological information and whole-genome sequences of COVID-19 are being aggregated worldwide.[28] Continuous collections of patient epidemiological information and devising strategies for whole-genome sequencing are important for epidemiological surveys. In general, correct epidemiological surveys of viral infections should be based on whole-genome sequences. Ultra-low-copy samples limit the possibility of using NGS to obtain whole-genome sequences. However, various sample preparation methods have been developed to overcome this limitation. The whole-genome sequences can be obtained using a probe-based target-enrichment method. Furthermore, whole-genome sequences of the isolates can be acquired using the well-known sequence-independent single-primer amplification method.[29] Particularly, in the case of clinical samples, the target enrichment method can be employed to obtain

data corresponding to an average coverage of 99% and depth of ×1000 or more.

The SARS-CoV-2 strains harboring the spike D614G mutation are known to be more infectious and pathogenic.[30] The Q57H substitution in ORF3a introduces a premature stop codon in ORF3b, resulting in a truncated ORF3b with a loss of function.[31] The prevalence of strains from the GH clade in patients with severe disease or is deceased patients is higher than that of strains from the G clade.[32] The T265I and Q57H substitutions in ORF1a and ORF3a, respectively, have also been reported, and they are known to occur in variants with D614G mutations. T265I in ORF1a, D614G, and NS3 Q57H are the most prevalent mutations and are associated with the removal of the protein stability-related -OH group.[33] There were five additional mutations in the SARS-CoV-2 strains ADD-2/2i and ADD-4/4i, namely three nonsynonymous mutations in the ORF1a protein (I453T, T551A, and V627I) and two synonymous mutations in the ORF1a polyprotein and nucleocapsid phosphoprotein. ORF1a/b encodes a replicase polypeptide that gives rise to 16 mature nonstructural proteins (nsp 1–16) required for viral RNA replication and transcription.[34] Thus, the additional mutations from ORF1a may affect viral replication, which should be further evaluated. The co-occurrence of these mutations may have functional effects on molecular interactions and on infectivity, survivability, transmissibility, and infection pathogenesis. SARS-CoV-2 evolves at a rate of ~$1.1 \times 10^{-3}$ substitutions/site/year, that is, one mutation can occur per genome every 11 days.[35]

## 5 | CONCLUSION

In summary, we performed comprehensive genetic and epidemiological analyses of SARS-CoV-2 isolated from saliva samples of patients at Proving Ground in the Taean Area. Further epidemiological surveillance and analysis of SARS-CoV-2 during community transmission are needed to identify viral phenotypes and monitor emerging variants. Now, effective SARS-CoV-2 counter-measures, including social distancing, face mask use, and personal hygiene are essential to mitigate the COVID-19 pandemic. Furthermore, early diagnosis, adequate clinical treatment, quarantine, and detailed epidemiological surveys such as contact tracing, genome sequencing, and variant surveillance are imperative to limit SARS-CoV-2 spread until effective therapeutics are developed and herd immunity is established. Molecular characteristic analyses of pathogenic viruses including whole-genome sequencing, phylogenetic analysis, and virus isolation are urgently needed to gain more information about the epidemiology, develop vaccines, and equip societies to overcome the next pandemic.

## AUTHOR CONTRIBUTIONS

**Se Hun Gu**: Conceptualization; data curation; formal analysis; investigation; writing–original draft. **Dong Hyun Song**: Conceptualization; data curation; formal analysis; funding acquisition; investigation; writing–original draft. **Hyeongseok Yun**: Formal analysis; investigation. **Jung-Eun Kim**: Writing–original draft. **Seung-Ho Lee**: Data curation; formal analysis. **Hyunjin Lee**: Investigation. **Tae Ho Lee**: Investigation. **Seol Muk Kang**: Investigation. **Yu Sub Jung**: Investigation. **Gyeunghaeng Hur**: Funding acquisition. **Daesang Lee**: Conceptualization; data curation; formal analysis; funding acquisition; supervision; writing–original draft; writing–review and editing.

## ACKNOWLEDGMENTS

This research was funded by the Agency for Defense Development, Republic of Korea (grant number 211555-912821001). We gratefully acknowledge all data contributor, that is, the authors and their originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based. We would like to thank Editage (www.editage.co.kr) for English language editing.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## DATA AVAILABILITY STATEMENT

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

## TRANSPARENCY STATEMENT

The lead author Daesang Lee affirms that this manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

## ETHICAL APPROVAL

All procedures involving human participants were in accordance with ethical standards.

## ORCID

*Se Hun Gu* http://orcid.org/0000-0003-0430-1348
*Daesang Lee* http://orcid.org/0000-0001-8487-1302

## REFERENCES

1. Hu B, Guo H, Zhou P, Shi ZL. Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*. 2021;19:141-154.
2. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet*. 2020;395:497-506.
3. WHO COVID-19 Dashboard. World Health Organization. 2020 (last cited February 18, 2022). http://covid19.who.int/
4. Coronavirus (COVID-19). Republic of Korea (last cited February 21, 2022). http://ncov.mohw.go.kr/en/
5. Kyriakidis NC, López-Cortés A, González EV, Grimaldos AB, Prado EO. SARS-CoV-2 vaccines strategies: a comprehensive review of phase 3 candidates. *NPJ Vaccines*. 2021;6:28.
6. Martin MA, VanInsberghe D, Koelle K. Insights from SARS-CoV-2 sequences. *Science*. 2021;371:466-467.
7. Volz E, Mishra S, Chand M, et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*. 2021;593:266-269.

8. Tegally H, Wilkinson E, Giovanetti M, et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature*. 2021;592:438-443.
9. Faria NR, Mellan TA, Whittaker C, et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science*. 2021;372:815-821.
10. Singh J, Rahman SA, Ehtesham NZ, Hira S, Hasnain SE. SARS-CoV-2 variants of concern are emerging in India. *Nat Med*. 2021;27:1131-1133.
11. Jeong HW, Kim S-M, Kim H-S, et al. Viable SARS-CoV-2 in various specimens from COVID-19 patients. *Clin Microbiol Infect*. 2020;26:1520-1524.
12. Kim J-M, Chung Y-S, Jo HJ, et al. Identification of coronavirus isolated from a patient in Korea with COVID-19. *Osong Public Health Res Perspect*. 2020;11:3-7.
13. Williams E, Bond K, Zhang B, Putland M, Williamson DA. Saliva as a noninvasive specimen for detection of SARS-CoV-2. *J Clin Microbial*. 2020;58:e00776-00720.
14. To KK, Tsang OT, Leung W-S, et al. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. *Lancet Infect Dis*. 2020;20:565-574.
15. Williams E, Isles N, Chong B, et al. Detection of SARS-CoV-2 in saliva: implications for specimen transport and storage. *J Med Microbiol*. 2021;70:001285.
16. Wyllie AL, Fournier J, Casanovas-Massana A, et al. Saliva or Nasopharyngeal swab specimens for detection of SARS-CoV-2. *N Engl J Med*. 2020;383:1283-1286. doi:10.1056/NEJMc2016359
17. Fernández-González M, Agulló V, de la Rica A, et al. Performance of saliva specimens for the molecular detection of SARS-CoV-2 in the community setting: does sample collection method matter? *J Clin Microbiol*. 2021;59:e03033.
18. Ota K, Yanagihara K, Sasaki D, et al. Detection of SARS-CoV-2 using qRT-PCR in saliva obtained from asymptomatic or mild COVID-19 patients, comparative analysis with matched nasopharyngeal samples. *PLoS One*. 2021;16:e0252964.
19. Zhu N, Zhang D, Wang W, et al. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med*. 2020;382:727-733.
20. Corman VM, Landt O, Kaiser M. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill*. 25:2000045.
21. Kim JM, Park SY, Lee D, et al. Genomic investigation of the coronavirus disease-2019 outbreak in the Republic of Korea. *Sci Rep*. 2020;11:6009. doi:10.1038/s41598-021-85623-6
22. Gu SH, Yu CH, Song Y, et al. A small interfering RNA lead targeting RNA-dependent RNA-polymerase effectively inhibit the SARS-CoV-2 infection in Golden Syrian hamster and Rhesus macaque. *bioRxiv*. 2020. doi:10.1101/2020.07.07.190967
23. Katoh K, Misawa K, Kuma KI, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. 2002;30:3059-3066.
24. Larsson A. AliView: a fast and lightweight alignment viewer and editor for large datasets. *Bioinformatics*. 2014;30:3276-3278.
25. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol*. 2021;38:3022-3027.
26. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. Bayesian phylogenetic and phylo-dynamic data integration using BEAST 1.10. *Virus Evol*. 2018;4:vey016.
27. McCormick-Baw C, Morgan K, Gaffney D, et al. Saliva as an alternate specimen source for detection of SARS-CoV-2 in symptomatic patients using Cepheid Xpert Xpress SARS-CoV-2. *J Clin Microbiol*. 2020;58:e01109-e01120.
28. Yuelong S, McCauley J, et al. GISAID: global initiative on sharing all influenza data—from vision to reality. *Eurosurveillance*. 2017;22:30494.
29. Song DH, Kim WK, Gu SH, et al. Sequence-independent, single-primer amplification next-generation sequencing of hantaan virus cell culture-based isolates. *Am J Trop Med Hyg*. 2017;96:389-394.
30. Plante JA, Liu Y, Liu J, et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*. 2021;592:116-121.
31. Lam JY, Yuen CK, Ip JD, et al. Loss of orf3b in the circulating SARS-CoV-2 strains. *Emerg Microbes Infect*. 2020;9:2685-2696.
32. Hamed SM, Elkhatib WF, Khairalla AS, Noreddin AM. Global dynamics of SARS-CoV-2 clades and their relation to COVID-19 epidemiology. *Sci Rep*. 2021;11:8435.
33. Laha S, Chakraborty J, Das S, Manna SK, Biswas S, Chatterjee R. Characterizations of SARS-CoV-2 mutational profile, spike protein stability and viral transmission. *Infect Genet Evol*. 2020;85:104445.
34. Benedetti F, Snyder GA, Giovanetti M, et al. Emerging of a SARS-CoV-2 viral strain with a deletion in nsp1. *J Transl Med*. 2020;18:329.
35. Duchene S, Featherstone L, Haritopoulou-Sinanidou M, Rambaut A, Lemey P, Baele G. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol*. 2020;6:veaa061.


## SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

**How to cite this article:** Gu SH, Song DH, Yun H, et al. Molecular characterization of SARS-CoV-2 from the saliva of patients in the Republic of Korea in 2020. *Health Sci Rep*. 2022;5:e856. doi:10.1002/hsr2.856