

Original Article  
Preventive & Social Medicine



# Epidemiologic Linkage of COVID-19 Outbreaks at Two University-affiliated Hospitals in the Seoul Metropolitan Area in March 2020

Kuenyoul Park ,<sup>1\*</sup> Jaewoong Lee ,<sup>2\*</sup> Kihyun Lee ,<sup>3</sup> Jiwon Jung ,<sup>4</sup> Sung-Han Kim ,<sup>4</sup> Jina Lee ,<sup>5</sup> Mauricio Chalita ,<sup>3</sup> Seok-Hwan Yoon ,<sup>3</sup> Jongsik Chun ,<sup>3,6</sup> Kyu-Hwa Hur ,<sup>1</sup> Heungsung Sung ,<sup>1</sup> Mi-Na Kim ,<sup>1</sup> and Hae Kyung Lee <sup>2</sup>

OPEN ACCESS

Received: Nov 29, 2020

Accepted: Jan 12, 2021

Address for Correspondence:

Mi-Na Kim, MD

Department of Laboratory Medicine, Asan Medical Center, University of Ulsan College of Medicine, 88, Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea.

E-mail: mnkim@amc.seoul.kr

Heungsung Sung, MD

Department of Laboratory Medicine, Asan Medical Center, University of Ulsan College of Medicine, 88, Olympic-ro 43-gil, Songpa-gu, Seoul 05505, Korea.

E-mail: sung@amc.seoul.kr

\*Kuenyoul Park and Jaewoong Lee contributed equally to the writing of this article.

© 2021 The Korean Academy of Medical Sciences.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORCID iDs

Kuenyoul Park

<https://orcid.org/0000-0003-3950-4946>

Jaewoong Lee

<https://orcid.org/0000-0001-8318-050X>

Kihyun Lee

<https://orcid.org/0000-0001-5175-0029>

Jiwon Jung

<https://orcid.org/0000-0003-4333-3270>

<sup>1</sup>Department of Laboratory Medicine, University of Ulsan College of Medicine and Asan Medical Center, Seoul, Korea

<sup>2</sup>Department of Laboratory Medicine, Uijeongbu St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Uijeongbu, Korea

<sup>3</sup>ChunLab Inc., Seoul, Korea

<sup>4</sup>Department of infectious disease, University of Ulsan College of Medicine and Asan Medical Center, Seoul, Korea

<sup>5</sup>Department of Pediatrics, University of Ulsan College of Medicine and Asan Medical Center, Seoul, Korea

<sup>6</sup>Department of Biological Sciences, Seoul National University, Seoul, Korea

## ABSTRACT

**Background:** Coronavirus disease 2019 (COVID-19) outbreaks emerged at two university-affiliated hospitals in Seoul (hospital A) and Uijeongbu City (hospital S) in the metropolitan Seoul area in March 2020. The aim of this study was to investigate epidemiological links between the outbreaks using whole genome sequencing (WGS) of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).

**Methods:** Fifteen patients were enrolled in the study, including four non-outbreak (A1–A4) and three outbreak cases (A5–A7) in hospital A and eight cases (S1–S8) in hospital S. Patients' hospital stays, COVID-19 symptoms, and transfer history were reviewed. RNA samples were submitted for WGS and genome-wide single nucleotide variants and phylogenetic relationships were analyzed.

**Results:** The index patient (A5) in hospital A was transferred from hospital S on 26 March. Patients A6 and A7 were the family caregiver and sister, respectively, of the patient who shared a room with A5 for 4 days. Prior to transfer, A5 was at the next bed to S8 in the emergency room on 25 March. Patient S6, a professional caregiver, took care of the patient in the room next to S8's room for 5 days until 22 March and then S5 for another 3 days. WGS revealed that SARS-CoV-2 in A2, A3, and A4 belong to clades V/B.2, S/A, and G/B.1, respectively, whereas that of A5–A7 and S1–S5 are of the V/B.2.1 clade and closely clustered. In particular, SARS-CoV-2 in patients A5 and S5 showed perfect identity.

**Conclusion:** WGS is a useful tool to understand epidemiology of SARS-CoV-2. It is the first study to elucidate the role of patient transfer and caregivers as links of nosocomial outbreaks of COVID-19 in multiple hospitals.

**Keywords:** COVID-19; SARS-CoV-2; Nosocomial Outbreak; Whole Genome Sequencing; Clade; Lineage

Sung-Han Kim <https://orcid.org/0000-0002-6596-8253>Jina Lee <https://orcid.org/0000-0002-3435-251X>Mauricio Chalita <https://orcid.org/0000-0002-2173-6688>Seok-Hwan Yoon <https://orcid.org/0000-0002-4886-8170>Jongsik Chun <https://orcid.org/0000-0003-3385-5171>Kyu-Hwa Hur <https://orcid.org/0000-0002-2159-3592>Heungsup Sung <https://orcid.org/0000-0002-6062-4451>Mi-Na Kim <https://orcid.org/0000-0002-4624-6925>Hae Kyung Lee <https://orcid.org/0000-0001-8545-9272>

### Funding

This work was supported by a Bio and Medical Technology Development Program of the National Research Foundation (NRT) of Korea grant funded by the Korea government (grant number NRF-2016M3A9B6918716).

### Disclosure

The authors have no potential conflicts of interest to disclose.

### Author Contributions

Conceptualization: Kim MN, Sung H; Formal analysis: Park K, Lee J,<sup>1</sup> Lee K; Funding acquisition: Sung H; Methodology: Park K, Lee K, Chalita M, Yoon SH, Chun J; Project administration: Kim MN, Sung H; Resources: Hur KH, Lee J.<sup>1</sup>; Writing - original draft: Park K, Lee J,<sup>1</sup> Lee K; Writing - review & editing: Jung J, Kim SH, Lee J,<sup>2</sup> Sung H, Kim MN, Lee HK.

Lee J,<sup>1</sup> Jaewoong Lee; Lee J,<sup>2</sup> Jina Lee.

## INTRODUCTION

Coronavirus disease 2019 (COVID-19) was first described in December 2019, in Wuhan, Hubei Province, China.<sup>1</sup> As of December 21, 2020, there were 75,704,857 cases of COVID-19 confirmed in 222 countries,<sup>2</sup> and 45,475 cases were reported in South Korea since its introduction by a Chinese traveler on January 19, 2020.<sup>3,4</sup> COVID-19 cases are mainly found to be community-acquired; sporadic cases of unknown origin account for 16.0%.<sup>4</sup> A community outbreak involving 6,930 cases in the Daegu-Gyeongbuk area of South Korea emerged between mid-February and mid-March, accounting for the majority of COVID-19 cases to date nationwide.<sup>5,6</sup> Most community cases stem from numerous outbreaks of varying sizes in Korea.<sup>7,8</sup> COVID-19 is well known to be highly contagious in mild to asymptomatic cases, thus causing many community outbreaks and often become widespread very quickly in a community. Therefore, early detection based on rapid molecular diagnostics and thorough contact tracing are essential, basic tools to help prevent the spread of COVID-19 in a community.

Health care facilities house some of the most vulnerable and immunocompromised individuals and clearly have high morbidity and mortality related to COVID-19<sup>9</sup> and require strict infection control to prevent nosocomial outbreaks. From January to September 2020, several instances of nosocomial outbreaks of COVID-19 have been published.<sup>10,11</sup> Korea Centers for Disease Control and Prevention (KCDC) reviewed seven major COVID-19 nosocomial outbreaks in South Korea affecting 39 to 196 patients,<sup>5</sup> but no molecular epidemiological investigations were reported. Whole genome sequencing (WGS) is a powerful tool for epidemiological analysis of newly emerging viral disease outbreaks<sup>12,13</sup> and currently is an established technique to classify the clades and lineages of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) to investigate the evolution and epidemiology of the virus.<sup>14,15</sup> At the clade level, it is possible to match dominant clones to certain regions. According to information published on the GISAID Next hCoV-19 app (<https://www.gisaid.org/>), as of 8 October, SARS-CoV-2 clade L is the most ancestral clade from Wuhan, China, clade G is the second generation clade prevailing in the European Union, which then evolved to give rise to clade GH in the United States, and GR in Russia, Brazil, and the United Kingdom. Although initially the predominance of V clades is obvious due to a large clade V cluster outbreak in Daegu-Gyeongbuk, clades S (n = 24), V (n = 67), and G (n = 55) are found among 151 representative cases in Korea.<sup>5</sup> WGS is also remarkably successful in tracing clonality in community outbreaks<sup>16,17</sup> as well as nosocomial outbreaks.<sup>18</sup> However, to date, there has been no reports of WGS analysis of epidemiologic link between nosocomial outbreaks of COVID-19 in multiple hospitals.

In late March, an outbreak of three COVID-19 cases emerged in a pediatric ward in a tertiary care hospital in Seoul,<sup>19</sup> in which an index patient was transferred from a university-affiliated hospital in Uijeongbu City during an ongoing outbreak, comprising 60 patients.<sup>6</sup> The two hospitals are 36 km apart in the Seoul metropolitan area. This study was to further investigate the nosocomial outbreaks of COVID-19 in both hospitals using WGS analysis of SARS-CoV-2 to elucidate epidemiologic link between them.

## METHODS

### Patient's enrolment

A total of 15 patients, including 3 outbreak cases (A5–A7) and 4 non-outbreak cases (A1–A4) in Asan Medical Center (hospital A) and eight cases (S1–S8) in Uijeongbu St. Mary's hospital

(hospital S) were reviewed for their hospital stays, COVID-19 symptoms, and transfer history. For all patients, COVID-19 was confirmed by SARS-CoV-2 real-time reverse transcription polymerase chain reaction (rRT-PCR). We reviewed the electronic medical records for clinical symptoms and epidemiologic findings such as hospital stays and transfer history to generate an outbreak synopsis.

### WGS of SARS-CoV-2

Samples for RNA extraction were obtained by nasopharyngeal swabs or sputa. RNA was extracted using eMAG (bioMérieux, Marcy-l'Etoile, France) in hospital A and AdvanSure E3 (LG Chem, Seoul, Korea) in hospital S, and stored at  $-70^{\circ}\text{C}$  until further analysis. Whole metagenome sequencing was performed on library preparations prepared using TruSeq Stranded Total RNA Human/Mouse/Rat kit (Illumina, CA, USA) and sequenced using the NovaSeq 6000 (Illumina), following the scheme used by Wu et al.<sup>20</sup> WGS results were reconstructed by aligning the sequence reads to the reference genome, SARS-CoV-2 Wuhan-Hu-1 genome (NC\_045512.2).

### Analysis of single nucleotide variants (SNVs): clade and lineage assignment

SNVs were assigned based on the reference-based alignment of published sequences.<sup>21</sup> Clade assignment was performed according to the current scheme on the GISAID website<sup>14</sup>. The genome sequences were phylogenetically classified under the lineage classification scheme proposed by Rambaut et al. using Pangolin software version 1.1.14 and the lineage database version 2020-05-19.<sup>15</sup>

SARS-CoV-2 genome sequences, obtained from GISAID on June 1, 2020 ( $n = 34,806$ ), were aligned pairwise to the reference genome (NC\_045512.2) (Shu and McCauley, 2017). Following the sequence alignment, high-quality sequences with  $\geq 99\%$  completeness,  $\leq 30$  nucleotide differences relative to the reference genomes, and without frame-shift error in all open reading frames were selected ( $n = 24,684$ ). The high-quality sequences were de-replicated to 15,067 unique genome sequences. Genomic positions with  $< 99\%$  coverage by the non-redundant genomes were omitted from the single nucleotide polymorphism (SNP) analyses. The positions where the minor allele frequency was  $\geq 1\%$  in the non-redundant genome set were defined as SNP sites ( $n = 55$ ). The 104 GISAID-obtained genomes, representing the 104 unique allelic combinations in the genome-wide SNP sites, and the 57 GISAID-obtained genomes that had the fewest genome-wide single-nucleotide differences compared to the case genomes were included in phylogenetic analysis. From the whole genome alignment of 29,903 genomic positions, we removed the positions with more than a 1% gap or ambiguous characters, resulting in the clipped alignment of 28,371 positions. The ModelFinder function of IQ-Tree version 1.6.12 was run to determine the optimal substitution model for the data set. Using the selected model, a maximum-likelihood phylogenetic tree was generated with ultrafast bootstrap ( $n = 1,000$ ) for branch support estimation.

### Data availability

The raw reads from the total RNA sequencing of nasopharyngeal samples ( $n = 15$ ) were deposited in the NCBI SRA as BioProject number PRJNA645035. The SARS-CoV-2 genome sequences recovered in this study ( $n = 11$ ) were deposited in GISAID with accession numbers EPI\_ISL\_485001 - EPI\_ISL\_485002 and EPI\_ISL\_485388 - EPI\_ISL\_485396.

**Ethical statement**

This study was reviewed and approved by the Institutional Review Board of Asan Medical Center (S2019-1169-0001) and the Institutional Review Board of the Catholic Medical Center (UC20SID10057). Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements.

**RESULTS**

**Outbreak synopsis**

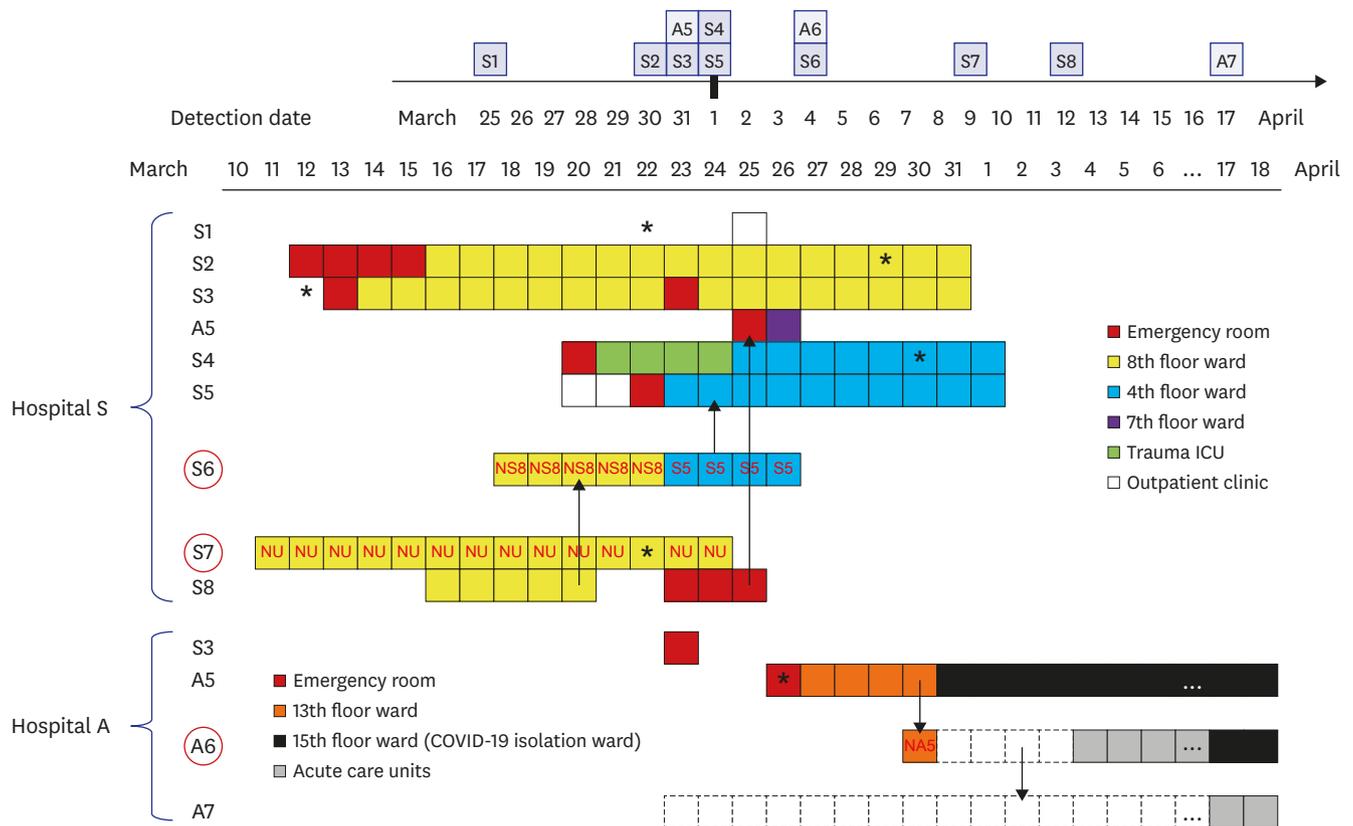
Clinical and epidemiological features of all fifteen COVID-19 patients are presented in **Table 1**, and ward transfer history of outbreak-related patients is detailed in **Fig. 1**. Patient A5 was the index patient in hospital A, who was transferred from hospital S on 26 March; A6 and A7 were the mother and elder sister, respectively, of the non-infected patient (NA5) who stayed in the bed across from A5 in a six-bed hospital room. A6 stayed overnight in the hospital room as a family caregiver of NA5 on 30 March and then went home and stayed with another daughter (A7) at home for 4 days, and then returned to provide care for NA5. A5 was negative for SARS-CoV-2 by rRT-PCR testing when she was tested in the ER upon transfer to hospital A on 26 March. Patients were sequentially diagnosed on 31 March (A5), 4 April (A6), and 17 April (A7). S1 visited the outpatient clinic in hospital S and was diagnosed with COVID-19 on 25 March. S2–S5 and S8 were inpatients in the outbreak wards located on the 8th or 4th floor in hospital S and S6 and S7 were professional caregivers in these wards. S1–S8 were diagnosed with COVID-19 between March 25 and April 12 (**Table 1**). A5 was originally admitted into the bed next to S8 in the emergency room (ER) in hospital S, 1 hour after S8 had departed. S3 was transferred to the ER in hospital A on 23 March but returned to the ER in hospital S on the same day. S6 was the caregiver for the non-infected patient (NS8) in a room next to S8 for 5 days during S8’s stay in the 8th floor ward and then took care of S5, who was on the 4th floor.

**Table 1.** Clinical and epidemiological features of fifteen coronavirus disease 2019 patients

Case	Age/Sex	Comorbidity	Symptoms at diagnosis	Pneumonia on chest X-ray	Acquisition	Resident region
A1	73/F	Arrhythmia	Fever, cough	Pneumonia, consolidation, and GGO	Community	Daegu-Gyeongbuk
A2	61/F	TB history	Fever, cough	Pneumonia, consolidation, and GGO	Community	Daegu-Gyeongbuk
A3	62/F	Dyslipidemia	Cough, headache	Pneumonia, patchy GGO	Community	Seoul
A4	16/F	VSD	Fever	Pneumonia, peribronchial infiltration	Community	United Kingdom
A5	9/F	ICH	Fever <sup>a</sup>	No active lung lesions	Nosocomial	Northern Gyeonggi
A6	40/F	None	Asymptomatic	Pneumonia, patchy increased opacities	Nosocomial	Southern Gyeonggi
A7	2/F	None	Asymptomatic	No active lung lesions	Community	Southern Gyeonggi
S1	76/M	None	Fever, cough, sore throat, rhinorrhea	Not tested	Community	Northern Gyeonggi
S2	82/F	Tuberculosis	Fever	Pneumonia, increased opacities in LLLF	Nosocomial	Northern Gyeonggi
S3	56/M	Liver abscess	Fever	Subsegmental atelectasis	Nosocomial	Northern Gyeonggi
S4	83/M	ICH	Fever	Subsegmental atelectasis	Nosocomial	Northern Gyeonggi
S5	52/M	Spine fracture	Asymptomatic	No active lung lesions	Nosocomial	Northern Gyeonggi
S6	64/F	Hypertension	Asymptomatic	Pneumonia, consolidation, and GGO	Nosocomial	Northern Gyeonggi
S7	59/F	None	Asymptomatic	No active lung lesions	Nosocomial	Northern Gyeonggi
S8	66/F	DLBCL	Headache	No active lung lesions	Nosocomial	Northern Gyeonggi

DLBCL = diffuse large B-cell lymphoma, GGO = ground-glass opacification, ICH = intracranial hemorrhage, LLLF = left lower lung field, TB = tuberculosis, VSD = ventricular septal defect.

<sup>a</sup>The patient A5 initially had fever in the emergency room and no more fever until discharge with continuing antipyretics under impression of ICH-associated fever; <sup>b</sup>They were caregivers without comorbidity except hypertension.



**Fig. 1.** Hospital stay and transfer history of 11 COVID-19 patients and caregivers associated with the outbreaks in hospitals A and S. Detection dates were denoted on the calendar in the upper corner. The wards in which patient A5, A6, and A7 from hospital A and all eight patients from hospital S (S1–S8) stayed were denoted using the color-coded squares. The red circles represents caregivers and the red letters inside squares indicated the patients cared by the caregivers. Asterisk denoted the onset of symptoms related to COVID-19 and dotted squares indicated the family home in which A6 and A7 lived. Possible transmission directions between outbreak patients were indicated by arrows.

**WGS determination of clades and lineage**

RNA samples from four patients were negative for SARS-CoV-2 reads, but the SARS-CoV-2 genome sequences from eleven samples were successfully constructed (Table 2).

Among the eleven SARS-CoV-2 genomes, SNVs were detected at 25 positions, 9 of which overlap with SNP positions in the global SARS-CoV-2 pandemic population (Fig. 2A). The single nucleotide variant (SNV) positions were distributed throughout the entire length of the genome including most of the SARS-CoV-2 protein-coding genes (Fig. 2A, bottom panel). Nine out of twenty-five SNVs occurred among the global SNP sites, whereas the other sixteen were presented in positions where the global frequency of variants was quite low. Three to six alleles were found to be missing due to the incompleteness of the A4, S2, S3, and S4 genomes at 25 SNV positions. Fifteen different SNV sites were observed among the eleven genomes (Supplementary Table 1). However, the eight outbreak strains in the study (A5–A7 and S1–S5) showed up to 6 SNVs. A5–A7 and S5 were the least variable with 3 SNVs. Mutation events among outbreak strains, as defined based on ancestral reconstruction, included eight genomic positions that were well separated across the genomes without a linkage to any particular protein-coding gene (Fig. 2B).

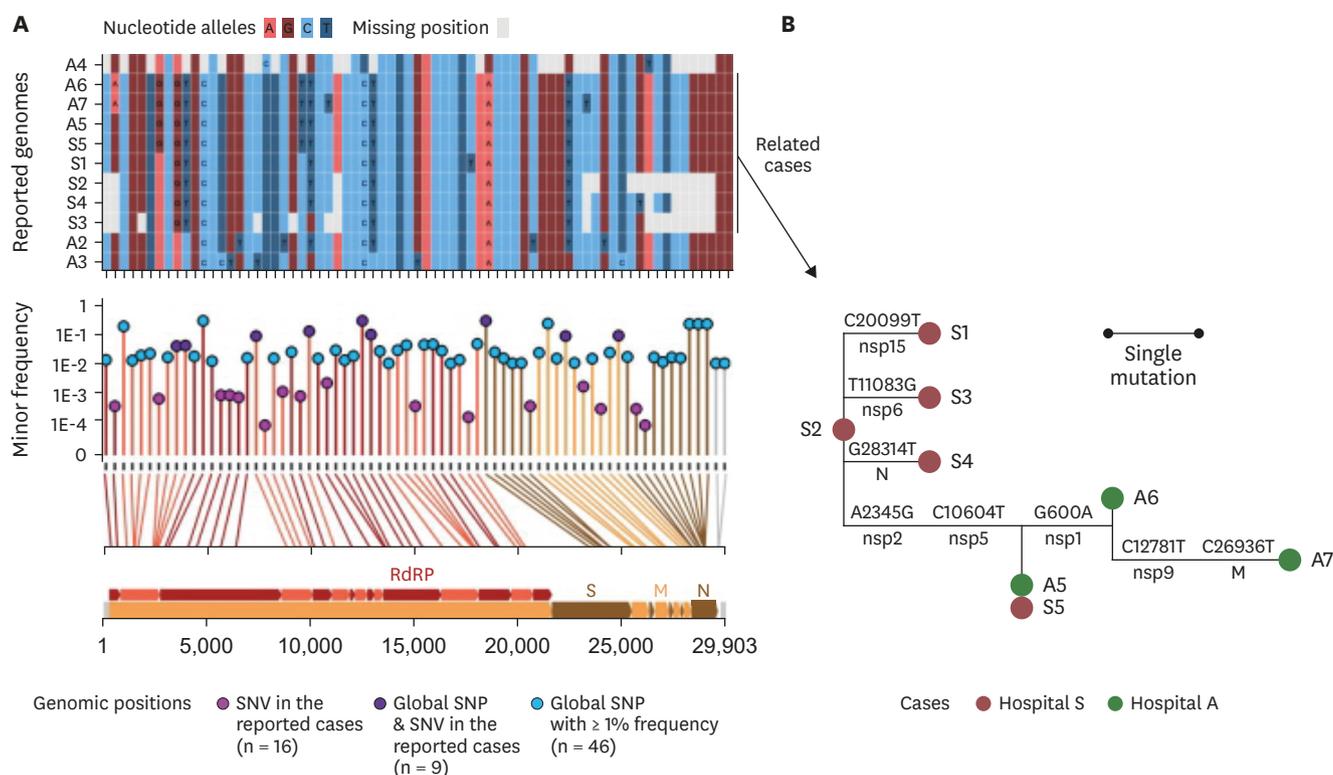
According to the popular clade classification scheme of SARS-CoV-2 genomic variants, as presented on the GISAID website, seven out of the eight outbreak genomes were unambiguously classified as the V clade, while the remaining one case (S3) showed an

**Table 2.** Quality parameters and typing results of SARS-CoV-2 genome sequences

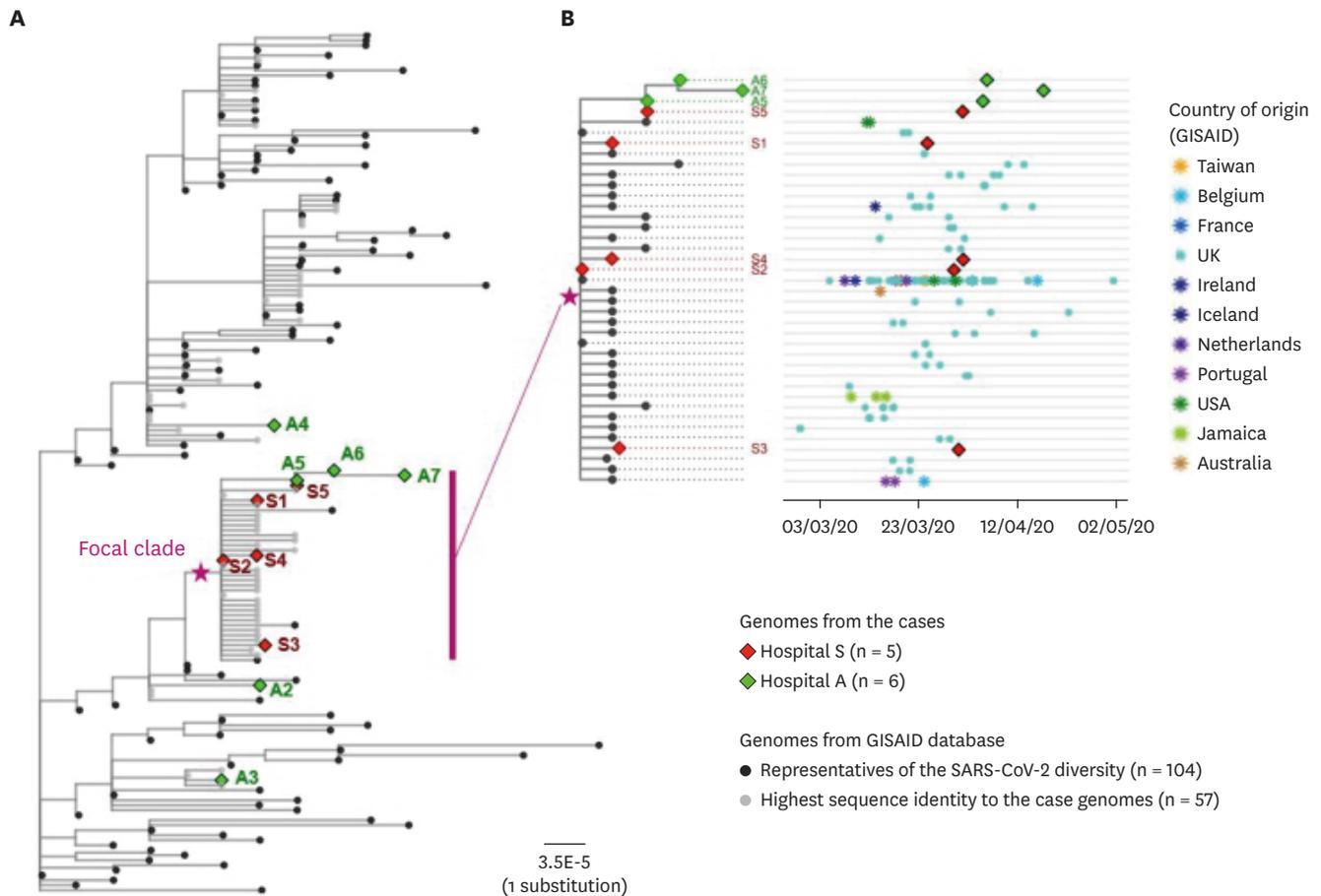
Case	Sequencing run		SARS-CoV-2 coverage			Typing results	
	Reads	Bases (Mbp)	Covered positions (bp)	Coverage breadth (%)	Mean depth	Clade	Lineages <sup>a</sup>
A1	6.4 × 10 <sup>6</sup>	967	0	0	0	NA	NA
A2	8.8 × 10 <sup>6</sup>	1,328	29,869	99.9	87.3	V	B.2
A3	7.2 × 10 <sup>6</sup>	1,086	29,896	100	172	S	A
A4	6.2 × 10 <sup>6</sup>	932	19,384	64.8	2.73	G	B.1
A5	8.1 × 10 <sup>6</sup>	1,216	29,715	99.4	14.5	V	B.2.1
A6	7.7 × 10 <sup>6</sup>	1,159	29,901	100	4,867	V	B.2.1
A7	1.4 × 10 <sup>7</sup>	2,130	29,862	99.9	27.8	V	B.2.1
S1	7.1 × 10 <sup>6</sup>	1,072	29,873	99.9	70	V	B.2.1
S2	6.0 × 10 <sup>6</sup>	908	27,696	92.6	27.7	V	B.2.1
S3	1.4 × 10 <sup>7</sup>	2,094	26,853	89.8	11.9	V <sup>b</sup>	B.2.1
S4	6.9 × 10 <sup>6</sup>	1,047	28,241	94.4	23.3	V	B.2.1
S5	8.3 × 10 <sup>6</sup>	1,260	29,903	100	7,199	V	B.2.1
S6	7.6 × 10 <sup>6</sup>	1,141	0	0	0	NA	NA
S7	6.0 × 10 <sup>6</sup>	904	0	0	0	NA	NA
S8	4.6 × 10 <sup>7</sup>	7,006	0	0	0	NA	NA

NA = not available.

<sup>a</sup>Lineage classification scheme proposed by Rambaut et al. (14) using Pagolin software version 1.1.14 and the lineage database version 2020-05-19; <sup>b</sup>S3 strain had the G26144T variant but not the G11083T variant.



**Fig. 2.** SNVs among the genomes reconstructed from eleven RNA samples. **(A)** Allele states of the eleven genomes at the SNV sites and global SNP sites. The upper panel outlined the allele states of the 11 genomes at the 71 positions that had either SNVs from the cases in this study or from GISAID sequences. The lower panel described the minor allele frequency of the 71 SNV/SNP positions. The vertical lines connecting the allele frequency point to the bottom x-axis were colored according to the color of the genomic feature in the genomic map that was inserted in the bottom of the figure. The genomic coordinates of each of the 71 positions were mapped by line to the genomic map at the bottom. RdRp, S, M, and N stood for regions encoding the RNA-dependent RNA polymerase, the spike protein, the membrane protein, and the nucleocapsid, respectively. **(B)** Reconstructed mutation events among the eight epidemiologically related cases. The SNVs were phylogenetically analyzed using the maximum-likelihood tree method. Mutations on the branches were defined based on ancestral reconstruction of the allele states. Labels above the branches denoted the nucleotide changes at a given genomic coordinate; those below the branches denoted the protein-level regions where the mutated loci resided. SNV = single nucleotide variant, SNP = single nucleotide polymorphism, nsp = nonstructural protein, N = nucleocapsid protein, M = membrane protein.



**Fig. 3.** Phylogenetic analysis of the sequenced SARS-CoV-2 genomes with the global reference genomes. **(A)** Maximum-likelihood phylogenetic tree of the genomes recovered from the reported cases (n = 11) and the selected genomes from the GISAID database (n = 161). Epidemiological information on the patient samples, when available, were noted on the right side of the tip labels. **(B)** Subtree of the focal clade containing eight outbreak-related cases and the public genomes with highly similar sequences ( $\leq 3$  nucleotide differences). To the right was the corresponding scatter plot of the collection date and geographic origin of the genomes belonging to each tip in the subtree. Each tip in the phylogenetic tree represented a cluster of identical genomes of multiple samples.

L-clade-associated allele (G) at position 11,083 and a V-clade-associated allele (T) at the other marker position 26,144 (Table 2). Based on the placement of the S3 GS in the whole-genome phylogenetic tree (Fig. 3) and the lineage assignment (to be described below), S3 could be robustly assigned to a V clade variant (Fig. 2B) with guanine at position 26,144 like an L clade variant. Non-outbreak cases were classified as clade S/lineage A, clade G/lineage B.1, and clade V/lineage B.2 for each of A3, A4, and A2 genomes, respectively. The A4 genome had V614G variant of spike protein but could not be further classified to the GH or GR clade due to its low coverage. According to the phylogenetic classification scheme, the genomes from outbreak-associated patients were all classified as lineage B.2.1 (n = 8).

### Whole genome phylogenetic analysis

A maximum-likelihood whole genome phylogeny containing the genomes representing the reported cases, the global SNP types, and those publicly available that had the highest identity to the case-derived genomes, was inspected to cross-check the epidemiological relationship for the cases. The genomes belonging to clade V and lineage B.2.1 (n = 8) appeared to be a focal clade in which the close epidemiological relationships were plausible (Fig. 3A). Genomes assigned to clade G, clade S, and clade V- lineage B.2 were distant from the focal clade on the phylogenetic tree. Furthermore, the phylogenetic tree constructed using the WGS data of all

nosocomial outbreak strains from both hospitals showed that all were almost identical, which could be deduced as inter-hospital transmission mediated by a single patient (A5).

We further investigated the date of collection and the geographic origins of the genomes belonging to the focal clade containing 8 outbreak-related strains. The outbreak strains of this study revealed high similarity (< 6 nucleotide difference) with the strains sampled from March to April 2020 in Europe, United States, Jamaica, Australia, and Taiwan. Also, S2 was identical to the representative strain of the clade. Other patients belonging to S hospital outbreak had one base pair difference from patient S2, with the exception of S5, who had a difference of two base pairs. Sequence obtained from A3 was identical to that of S5, and had one base pair difference from A6 and two base pair difference from A7 (**Fig. 3B**).

## DISCUSSION

Phylogenetic analysis revealed that the SARS-CoV-2 genomes of A5–A7 clustered with those of S1–S5 and were almost identical to S5, suggesting transmission of the virus from A5 to A6 and A7 and a link between the outbreaks in the two hospitals. These findings were consistent with the finding that patient A5 tested negative for SARS-CoV-2 by initial rRT-PCR performed in the ER in hospital A, but tested positive 5 days later in hospital A. Since S6 could explain the transmission of the same clone between S8 and S5, and A5 was identical to S5, it was most likely that A5 acquired the virus from S8 in the ER of hospital S. Therefore, the caregivers certainly played an important role in the nosocomial spread of SARS-CoV-2 between and within the two hospitals. This situation is likely explained by the fact that caregivers resided in the patient's room, that increased the density of number of humans in the room. In addition, caregivers taking care of patients professionally have more risk to facilitate the spread of SARS-CoV-2 more extensively. Thus, transmission of the virus can occur not only among hospital inpatients but also between other people in the hospital and those in the community. Patient transfer between hospitals could be an added risk that causes consecutive outbreaks between hospitals. These findings are not surprising given that numerous healthcare-associated outbreaks of COVID-19 have been reported<sup>22</sup> and healthcare workers in hospitals and long-term care facilities are well-known as high-risk group to acquire COVID-19.<sup>23</sup>

Only a few nosocomial outbreaks of COVID-19 in pediatric patients have been reported.<sup>11,19</sup> Children are often unnoticed cases in household outbreaks of COVID-19 since they are usually asymptomatic or have only mild to moderate symptoms.<sup>24,25</sup> It is uncertain whether they are victims or silence spreaders, as observed in this study. Pediatric inpatients require caregivers so the risk of nosocomial transmission could be doubled. High transmissibility of SARS-CoV-2 is well known and a part of the reason is the role that asymptomatic patients play as undetected carriers of the virus.<sup>26</sup>

Among the 15 cases in this study, less than 50% displayed pneumonia or respiratory symptoms. As demonstrated by A5, asymptomatic or paucisymptomatic patients may not be suspected of having COVID-19.<sup>27,28</sup> Due to the COVID-19 pandemic, both the university-affiliated hospitals adopted a strict policy to screen and test patients with suspected symptoms or epidemiological relevance for COVID-19 (S1 and A5 when they visited the outpatient clinic and ER, respectively). However, the outbreaks in the two hospitals clearly showed that the policies in place were not sufficient to prevent the spread of COVID-19 in patients in the acute care hospitals. Inpatients are especially vulnerable to nosocomial

outbreaks due to the high-risk of close contact among patients as well as caregivers who share the same wards or even the same multiple-occupancy room with each other. The outbreak in hospital S eventually comprises 60 cases of COVID-19 from March to April 2020,<sup>6</sup> whereas the outbreak in hospital A was limited to 3 cases. The difference between them was likely due to how quickly introductions of new COVID-19 infections were detected in hospital A. Universal testing of SARS-CoV-2 in pregnant women for childbirth has proven to be effective in controlling spread of COVID-19 infections.<sup>29</sup> Although the prevalence is not high in the community, universal screening of all inpatients upon admission and resident caregivers may be an effective strategy to prevent outbreaks in a tertiary care hospital such as hospital A, which has a high influx of patients from other hospitals. In fact, two patients in this study had a transfer history to hospital A among the five inpatients who acquired COVID-19 in hospital S. The infection control policy for COVID-19, such as mandated use of masks and frequent use of alcohol-based hand sanitizers is currently being employed in all Korean hospitals, but it is difficult to enforce these practices among children. Thus, universal screening of inpatients and their caregivers is needed and even more so in pediatric wards in which caregivers were always needed for all inpatients.

When genotypes of the SARS-CoV-2 genomes were analyzed, it was noted that all the patients of hospital S and A5 were residents of the northern part of Gyeonggi-do in the Seoul metropolitan area, in which hospital S located. Therefore, it is possible that A5 may have acquired the virus in the community before admission to hospital S. The outbreak of COVID-19 in the community of northern Gyeonggi-do may have preceded the outbreak in hospital S since S1, the first COVID-19 patient in hospital S, was diagnosed with COVID-19 in an outpatient clinic. The SARS-CoV-2 sequences of all COVID-19 patients from hospital S including S1 belong to the V clade/B.2.1 lineage, although no epidemiological link was identified among them by contact tracing conducted by KCDC officials. The V clade/B.2.1 lineage is found globally and first identified on February 23, 2020 in Korea according to public database (<https://www.gisaid.org/>), and this clone is close to Daegu-Gyeongbuk clone, V clade/B.2. Although G614 variant of the spike protein, well-known for its high infectivity, is currently dominant variant,<sup>30</sup> the outbreak strains in this study did not harbor it. Patient S1 did not have any history of hospital visits. Therefore, a probable regional spread of this lineage in the northern part of Gyeonggi-do occurred and later, a nosocomial outbreak of this strain took place in a secondary care hospital in that region. Only patients A5–A7 in hospital A belonged to the same lineage, whereas three non-outbreak strains were assigned to different clades or lineages. However, A5 more likely acquired the virus from S8 and transmitted the virus to A6. Subsequently A6 transmitted the infection to A7, considering that the S5 genome had no SNPs, and the A6 and A7 genomes had 2 SNPs and 3 SNPs, respectively, in alignment with the A5 genome. Phylogenetic analysis showed that clonal evolution of each of the index strains, S2 and A3, of nosocomial outbreaks in two hospitals proceeded during virus propagation. Considering that the accumulation of SNPs is time-dependent, outbreaks in the two hospitals spanning almost one month likely originated from a single strain and evolved enough to trace the linkage between the individuals and hospitals. Genotyping using SNVs is a powerful tool for epidemiological analysis.<sup>15,31</sup>

However, there were some limitations of this study. First, virus genomes were not detected in four samples including those from S8. Therefore, the linkage between the A5 genome and the S8 and S5 genomes was not fully proven by WGS. Second, S7 and S8 were diagnosed too late after being discharged from hospital S to obtain a RNA sample with high viral load or clinical analysis of COVID-19. Therefore, genotyping and clinical data were incomplete. Last, the

distribution of the uploaded genome sequences on GISAID could be unbalanced. This could lead to inadequate geographical inference of transmission sources of SARS-CoV-2.

In conclusion, WGS is a useful tool to find linkages of epidemiologically-related outbreaks. Caregivers may facilitate virus transmission within a hospital as well as between community and hospital, and patient transfer poses a risk of inter-hospital spread of COVID-19 outbreaks. Therefore, professional caregivers and patients transferred from other hospitals should be considered a high-risk group and subjected to better screening with proactive diagnostic testing to prevent nosocomial outbreaks of COVID-19, especially in tertiary care hospitals.

## ACKNOWLEDGMENTS

We would like to thank Moon Jung Kim M.D. and Jaeyun Lim M.D. (Department of Laboratory Medicine, Myungji Hospital) for kindly providing RNA samples and Editage ([www.editage.co.kr](http://www.editage.co.kr)) for editing and reviewing this manuscript for English language.

## SUPPLEMENTARY MATERIAL

### Supplementary Table 1

The fifteen sites that are variable among the case genomes and not missing in any case genome

[Click here to view](#)

## REFERENCES

- Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020;395(10223):497-506.  
[PUBMED](#) | [CROSSREF](#)
- Weekly operational update on COVID-19 - 21 December 2020. <https://www.who.int/publications/m/item/weekly-operational-update-on-covid-19---21-december-2020>. Updated 2020. Accessed December 22, 2020.
- Kim JY, Choe PG, Oh Y, Oh KJ, Kim J, Park SJ, et al. The first case of 2019 novel coronavirus pneumonia imported into Korea from Wuhan, China: implication for infection prevention and control measures. *J Korean Med Sci* 2020;35(5):e61.  
[PUBMED](#) | [CROSSREF](#)
- The updates on COVID-19 in Korea as of 21 December. <https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030>. Updated 2020. Accessed December 22, 2020.
- The updates on COVID-19 in Korea as of 6 July. <http://www.cdc.kr/board/board.es?mid=a30402000000&bid=0030>. Updated 2020. Accessed July 6, 2020.
- COVID-19 National Emergency Response Center Epidemiology and Case Management Team Korea Centers for Disease Control and Prevention. Coronavirus disease-19: the first 7,755 cases in the Republic of Korea. *Osong Public Health Res Perspect* 2020;11(2):85-90.  
[PUBMED](#) | [CROSSREF](#)
- The updates on COVID-19 in Korea as of 8 July. <https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030>. Updated 2020. Accessed July 8, 2020.
- Park SY, Kim YM, Yi S, Lee S, Na BJ, Kim CB, et al. Coronavirus disease outbreak in call center, South Korea. *Emerg Infect Dis* 2020;26(8):1666-70.  
[PUBMED](#) | [CROSSREF](#)
- Guo W, Li M, Dong Y, Zhou H, Zhang Z, Tian C, et al. Diabetes is a risk factor for the progression and prognosis of COVID-19. *Diabetes Metab Res Rev* 2020;36(7):e3319.  
[PUBMED](#) | [CROSSREF](#)

10. Wang X, Zhou Q, He Y, Liu L, Ma X, Wei X, et al. Nosocomial outbreak of COVID-19 pneumonia in Wuhan, China. *Eur Respir J* 2020;55(6):2000544.  
[PUBMED](#) | [CROSSREF](#)
11. Schwierzeck V, König JC, Kühn J, Mellmann A, Correa-Martínez CL, Omran H, et al. First reported nosocomial outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in a pediatric dialysis unit. *Clin Infect Dis* 2020;ciaa491.  
[PUBMED](#) | [CROSSREF](#)
12. Dudas G, Bedford T. The ability of single genes vs full genomes to resolve time and space in outbreak analysis. *BMC Evol Biol* 2019;19(1):232.  
[PUBMED](#) | [CROSSREF](#)
13. Grubaugh ND, Faria NR, Andersen KG, Pybus OG. genomic insights into zika virus emergence and spread. *Cell* 2018;172(6):1160-2.  
[PUBMED](#) | [CROSSREF](#)
14. Clade and lineage nomenclature, July 4, 2020. <https://www.gisaid.org/references/statements-clarifications/clade-and-lineage-nomenclature-aids-in-genomic-epidemiology-of-active-hcov-19-viruses/>. Updated 2020. Accessed July 7, 2020.
15. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* 2020;5(11):1403-7.  
[PUBMED](#) | [CROSSREF](#)
16. Böhmer MM, Buchholz U, Corman VM, Hoch M, Katz K, Marosevic DV, et al. Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. *Lancet Infect Dis* 2020;20(8):920-8.  
[PUBMED](#) | [CROSSREF](#)
17. Sekizuka T, Itokawa K, Kageyama T, Saito S, Takayama I, Asanuma H, et al. Haplotype networks of SARS-CoV-2 infections in the Diamond Princess cruise ship outbreak. *Proc Natl Acad Sci U S A* 2020;117(33):20198-201.  
[PUBMED](#) | [CROSSREF](#)
18. Meredith LW, Hamilton WL, Warne B, Houldcroft CJ, Hosmillo M, Jahun AS, et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect Dis* 2020;20(11):1263-72.  
[PUBMED](#) | [CROSSREF](#)
19. Jung J, Hong MJ, Kim EO, Lee J, Kim MN, Kim SH. Investigation of a nosocomial outbreak of coronavirus disease 2019 in a paediatric ward in South Korea: successful control by early detection and extensive contact tracing with testing. *Clin Microbiol Infect* 2020;26(11):1574-5.  
[PUBMED](#) | [CROSSREF](#)
20. Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. *Nature* 2020;579(7798):265-9.  
[PUBMED](#) | [CROSSREF](#)
21. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* 2017;22(13):30494.  
[PUBMED](#) | [CROSSREF](#)
22. Rhee C, Baker M, Vaidya V, Tucker R, Resnick A, Morris CA, et al. Incidence of nosocomial COVID-19 in patients hospitalized at a large US academic medical center. *JAMA Netw Open* 2020;3(9):e2020498.  
[PUBMED](#) | [CROSSREF](#)
23. Ran L, Chen X, Wang Y, Wu W, Zhang L, Tan X. Risk factors of healthcare workers with corona virus disease 2019: a retrospective cohort study in a designated hospital of Wuhan in China. *Clin Infect Dis* 2020;71(16):2218-21.  
[CROSSREF](#)
24. Dong Y, Mo X, Hu Y, Qi X, Jiang F, Jiang Z, et al. Epidemiology of COVID-19 among children in China. *Pediatrics* 2020;145(6):e20200702.  
[PUBMED](#) | [CROSSREF](#)
25. Qiu H, Wu J, Hong L, Luo Y, Song Q, Chen D. Clinical and epidemiological features of 36 children with coronavirus disease 2019 (COVID-19) in Zhejiang, China: an observational cohort study. *Lancet Infect Dis* 2020;20(6):689-96.  
[PUBMED](#) | [CROSSREF](#)
26. Bai Y, Yao L, Wei T, Tian F, Jin DY, Chen L, et al. Presumed asymptomatic carrier transmission of COVID-19. *JAMA* 2020;323(14):1406-7.  
[PUBMED](#) | [CROSSREF](#)
27. Yamagishi T, Kamiya H, Kakimoto K, Suzuki M, Wakita T. Descriptive study of COVID-19 outbreak among passengers and crew on Diamond Princess cruise ship, Yokohama Port, Japan, 20 January to 9 February 2020. *Euro Surveill* 2020;25(23):28-35.  
[PUBMED](#) | [CROSSREF](#)

28. Xu XW, Wu XX, Jiang XG, Xu KJ, Ying LJ, Ma CL, et al. Clinical findings in a group of patients infected with the 2019 novel coronavirus (SARS-Cov-2) outside of Wuhan, China: retrospective case series. *BMJ* 2020;368:m606.  
[PUBMED](#) | [CROSSREF](#)
29. Sutton D, Fuchs K, D'Alton M, Goffman D. Universal screening for SARS-CoV-2 in women admitted for delivery. *N Engl J Med* 2020;382(22):2163-4.  
[PUBMED](#) | [CROSSREF](#)
30. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* 2020;182(4):812-827.e19.  
[PUBMED](#) | [CROSSREF](#)
31. Yin C. Genotyping coronavirus SARS-CoV-2: methods and implications. *Genomics* 2020;112(5):3588-96.  
[PUBMED](#) | [CROSSREF](#)