Short Report

Healthcare workers acquired COVID-19 disease from patients? An investigation by phylogenomics

R.C.W. Wong [a],[*], M.K.P. Lee [a], G.K.H. Siu [b], L.K. Lee [b], J.S.L. Leung [b], E.C.M. Leung [a], Y.I.I. Ho [a], R.W.M. Lai [a]

[a] Department of Microbiology, Prince of Wales Hospital, Hong Kong SAR, China
[b] Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hung Hom, Hong Kong SAR, China

ARTICLE INFO

SUMMARY

The increasing number of coronavirus disease 2019 (COVID-19) cases in the community has posed a significant epidemic pressure on healthcare settings. When healthcare workers (HCWs) acquire COVID-19, contact tracing and epidemiological investigation might not be adequate for determining the source of transmission. Here, we report a phylogenetic investigation involving two infected HCWs and nine patients to determine whether patient-to-HCW transmission had occurred in a hospital without a previous COVID-19 outbreak. This is the first study to apply phylogenomics to investigate suspected nosocomial transmission in a region with low prevalence of COVID-19. Our results do not support the occurrence of direct patient-to-HCW transmission.

© 2021 The Healthcare Infection Society. Published by Elsevier Ltd. All rights reserved.

Introduction

The current wave of community outbreak of coronavirus disease 2019 (COVID-19) in Hong Kong at the time of writing commenced in November 2020 and the total number of symptomatic and asymptomatic infected cases was greater than in previous waves. When there are ongoing outbreaks of COVID-19, healthcare workers (HCWs) may acquire the disease from communities or through nosocomial transmission such as patient-to-HCW and HCW-to-HCW [1—3]. To differentiate between community-acquired infection from hospital-

acquired infection in HCWs, traditional epidemiological study alone may be inadequate [2]. Few studies have applied phylogenomics to investigate whether COVID-19-infected HCWs have acquired the disease in hospital [3]. Recently, two HCWs working in one of our cluster hospitals were tested positive for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RNA. Both HCWs had direct contact with COVID-19 patients. They wore full personal protective equipment (PPE) and complied with the hospital infection control guidelines. No nosocomial COVID-19 outbreaks had been reported in this hospital since the pandemic began. A follow-up investigation to

differentiate between hospital-acquired and community-acquired infection in these two HCWs was therefore important. Whole-genome sequencing (WGS) had been applied to study the genomic epidemiology of SARS-CoV-2 in hospitals with known outbreaks; infected patients/HCWs with epidemiological links and identical viral genomes would indicate that intra-hospital transmission had occurred [2—6]. Hong Kong is a densely populated city with >7.5 million residents. As of January 18th, 2021, the number of COVID-19 cases reported by the Centre for Health Protection was 9665 and the estimated COVID-19 prevalence rate was 0.13% since the pandemic. Here we report a phylogenetic investigation on two suspected cases of hospital-acquired COVID-19 among HCWs in a regional hospital. In this study, we aimed to apply WGS to supplement epidemiological investigation to determine whether COVID-19 in these HCWs was acquired from the infected patients. To our knowledge, this is the first report to apply phylogenomic analysis in a region with low prevalence of COVID-19 to investigate whether nosocomial acquisition of COVID-19 among HCWs had occurred in a hospital which did not have an ongoing outbreak of COVID-19.

Methods

Epidemiological investigation for HCWs with suspected hospital-acquired COVID-19

In mid-January 2021, two independent HCWs had developed upper respiratory tract symptoms and were tested positive for SARS-CoV-2 RNA. These were a doctor working in an intensive care unit (ICU) and a nurse working in an isolation ward for COVID-19 patients. Both HCWs had a recent history of close contact with the same deteriorating COVID-19 patient who required ICU care. Contact tracing by the hospital infection control team had further identified a group of eight COVID-19 patients who had been managed by these two HCWs in the period between January 4th and 14th, 2021. From a few days to three weeks after the onset of symptoms, blood was collected from patients who were ready for discharge and clinically stable with non-detectable SARS-CoV-2 RNA or high cycle threshold (C_T) value in respiratory samples. Blood samples were tested for IgG antibodies to nucleocapsid protein with the use of Alinity i SARS-CoV-2 IgG qualitative assay (Abbott, Sligo, Ireland).

Sample collection, laboratory diagnosis and whole-genome sequencing

Eleven respiratory specimens, including eight posterior oropharyngeal saliva (also known as deep throat saliva) and one sputum, were collected from nine patients. One posterior oropharyngeal saliva and one pooled nasopharyngeal and throat swabs were collected from the two infected HCWs. Using the automated QIAcube system (Qiagen, Hilden, Germany), RNA was extracted from these samples with the QIAamp MinElute Virus Spin Kit (Qiagen) according to the manufacturer's recommendation and tested with the TIB-Molbiol LightMix® SarbecoV E-gene assay (TIB-MolBiol, Berlin, Germany) as described previously [7].

For WGS of SARS-CoV-2, human DNA was removed from the RNA elutes with the use of Turbo DNA-free Kit (ThermoFisher Scientific, Waltham, MA, USA) [8]. cDNA was synthesized by using LunaScript RT SuperMix Kit (New England BioLabs Inc. (NEB), Hitchin, UK) and was then amplified by multiplex PCR with Q5 Hot Start High-Fidelity 2X Master Mix Kit (NEB) and ARTIC nCoV-2019 V3 primer sets (Integrated DNA Technologies, Coralville, IA, USA). WGS was performed using Nanopore MinION MK1B with flow cell R9.4.1 (FLO-MIN106D) (Oxford Nanopore Technologies, Oxford, UK) according to the ARTIC nCoV-2019 sequencing protocol v3. The consensus sequences of the samples in this study were submitted to the NCBI GenBank under the accession numbers MW768952 to MW768962.

Bioinformatics, phylogenetic analysis and estimation of evolutionary rate

Sequencing data were analysed by using the pipelines modified from the ARTIC network nCoV-2019 novel coronavirus bioinformatics protocol [8,9]. In brief, filtered reads with length between 400 and 700 bp were mapped to the reference genome Wuhan-Hu-1 (GenBank no.: MN908947.3) through minimap2 version 2.17 (https://anaconda.org/bioconda/minimap2). Variants were called by using Medaka (https://github.com/nanoporetech/medaka) and Clair (https://github.com/HKU-BAL/Clair) pipelines. Nucleotide substitution was numbered with reference to the Wuhan-Hu-1 genome. Novel single nucleotide polymorphism (SNP) (also known as sample-unique variants) was defined as specific nucleotide substitution in viral genome present in the corresponding sample only. A similar approach has been used to determine intra-hospital transmission of SARS-CoV-2 [2,5]. Multiple sequence alignment for consensus genomes was performed by Clustal Omega 1.2.0 (http://www.clustal.org/omega/). PhyML 3.0 (https://github.com/stephaneguindon/phyml) was used to construct maximum-likelihood phylogenetic tree with smart model selection of GTR substitution model. 1000× bootstrap replications were performed and Mega X version 10.0 (https://megasoftware.net/) was used to plot phylogenetic tree and rooted on the reference genome Wuhan-Hu-1. Bayesian molecular clock analysis with Markov Chain Monte Carlo (MCMC) framework provided by BEAST version 2.6.2 was used to estimate the viral evolutionary rate in substitutions per site per year as published in our previous study [8].

Results and discussion

WGS revealed that all samples from the HCWs and patients were clustered into the same SARS-CoV-2 lineage (Pangolin lineage B.1.36.27/GISAID clade GH/Nextstrain clade 20A), which predominated in the recent wave of COVID-19 outbreak in Hong Kong. SARS-CoV-2 RNA in the sample collected from the doctor (HCW-D) was characterized by three novel SNPs, namely 14064T→C, 26832G→T (A104S in M gene) and 27739C→T (L116F in orf7a gene). These mutations were not found in the patients and the nurse (HCW-N) (Table I). On the other hand, sample from HCW-N harboured three other novel SNPs, namely 307 C→A, 4579T→A, and 22419C→T that were not found in the patients and HCW-D (Table I). Patient 1 (ICU patient) shared identical viral genome with patients 2 and 3 and did not have novel SNP. Other patients (4—9) carried one or two novel SNPs that were not present in the HCWs (Table I).

Table I
Summary of 11 COVID-positive healthcare workers and patients and their corresponding novel SNPs on SARS-CoV-2 genome

| Sample	Date of specimen collection	Date of symptom onset	Specimen type	C_T value of *E* gene	IgG Ab to SARS-CoV-2 detected on	Novel SNPs present in HCW-D only	Novel SNPs present in HCW-N only	Novel SNPs present in patients but not found in HCW-D or HCW-N
HCW								
D	18/1/2021	17/1/2021	NPS + TS	17.3	25/1/2021	T14064C (*orf1b* gene), G26832T (*M* gene), C27739T (*orf7a* gene)	—	—
N	20/1/2021	18/1/2021	DTS	23.4	—	—	C307A & T4579A (*orf1a* gene), C22419T (*S* gene)	—
Patient								
1	5/1/2021	3/1/2021	DTS	14.7	21/1/2021	—	—	—
2	31/12/2020	28/12/2020	Sputum	18.7	11/1/2021	—	—	—
3	1/1/2021	Asymptomatic	DTS	27.9	5/1/2021	—	—	—
4	31/12/2020	29/12/2020	DTS	32.3	19/1/2021	—	—	C18348T (*orf1b* gene), C21911T (*S* gene)
5	7/12/2020	30/11/2020	DTS	30.1	15/12/2020	—	—	A7066G and C8660T (*orf1a* gene)
6	8/1/2021	5/1/2021	DTS	30.4	19/1/2021	—	—	C27944T (*orf8* gene)
7	8/1/2021	31/12/2020	DTS	25.7	12/1/2021	—	—	C27944T (*orf8* gene)
8	8/1/2021	8/1/2021	DTS	27.5	18/1/2021	—	—	C27944T (*orf8* gene)
9	13/1/2021	7/1/2021	DTS	28.3	18/1/2021	—	—	C934T (*orf1a* gene)

NPS + TS, pooled nasopharyngeal and throat swabs; DTS, posterior oropharyngeal saliva; *E* gene, envelope gene; Ab, antibody; SNP, single nucleotide polymorphism, numbered with reference to nucleotide position of the Wuhan-Hu-1 genome (accession no.: NC_045512.2); *orf*, open reading frame; *M* gene, membrane gene; *S* gene, spike protein gene.
Shared (non-novel) SNPs were not elaborated in this table. Sample was labelled as the following: healthcare workers (HCW-D: doctor; HCW-N: nurse) and patients (patients 1–9) in which patient 1 was the intensive care unit patient.

The hierarchy of the phylogenetic tree suggested that patients 1−3 could be the most recent common ancestor of both HCWs and other patients (Figure 1). The novel mutations identified in the HCWs might have emerged over time within the new hosts [2]. However, the evolutionary rate of SARS-CoV-2 lineage B.1.36.27 was 7.74×10^{-4} substitutions per site per year (95% BCI: 6.74×10^{-4} to 8.89×10^{-4} substitutions per site per year), which is equivalent to 1.93 nucleotide changes per month. This estimation was supported by the observation that the viral genome of patient 1 remained identical to those of the locally acquired cases reported in October 2020. A similar evolutionary rate was also observed by the other studies [2,5,6]. As the interval of symptom onset between patients 1−3 and the HCWs was within one month (Table I), it is unlikely that the SARS-CoV-2 variant with this evolutionary rate accumulated three mutations in such short period. The genetic distance (three to five novel SNP differences) did not support direct transmission of patient-to-HCW, though the possibility of

nosocomial acquisition of COVID-19 cannot be completely excluded. The mutually exclusive novel SNPs in HCW-D and HCW-N provided phylogenetic evidence that there was no HCW-to-HCW transmission as they were infected by two different variants of SARS-CoV-2.

SARS-CoV-2 has relatively low genetic diversity and therefore in-depth genomic study with single nucleotide resolution is necessary in order to resolve suspected hospital-acquired transmission. The application of SNP analysis in viral genome is a stringent approach to supplement the routine contact tracing. By using the same typing method, a recent study revealed that a SARS-CoV-2 variant that harboured T470N amino acid substitution ($22971C \rightarrow A$) in the spike (*S*) gene was disseminated in the catchment area of another hospital in Hong Kong in mid-December 2020 and eventually led to a nosocomial outbreak [10]. It should be noted that the genomic investigation must be integrated with the clinical and epidemiological information. Patients/HCWs having epidemiological
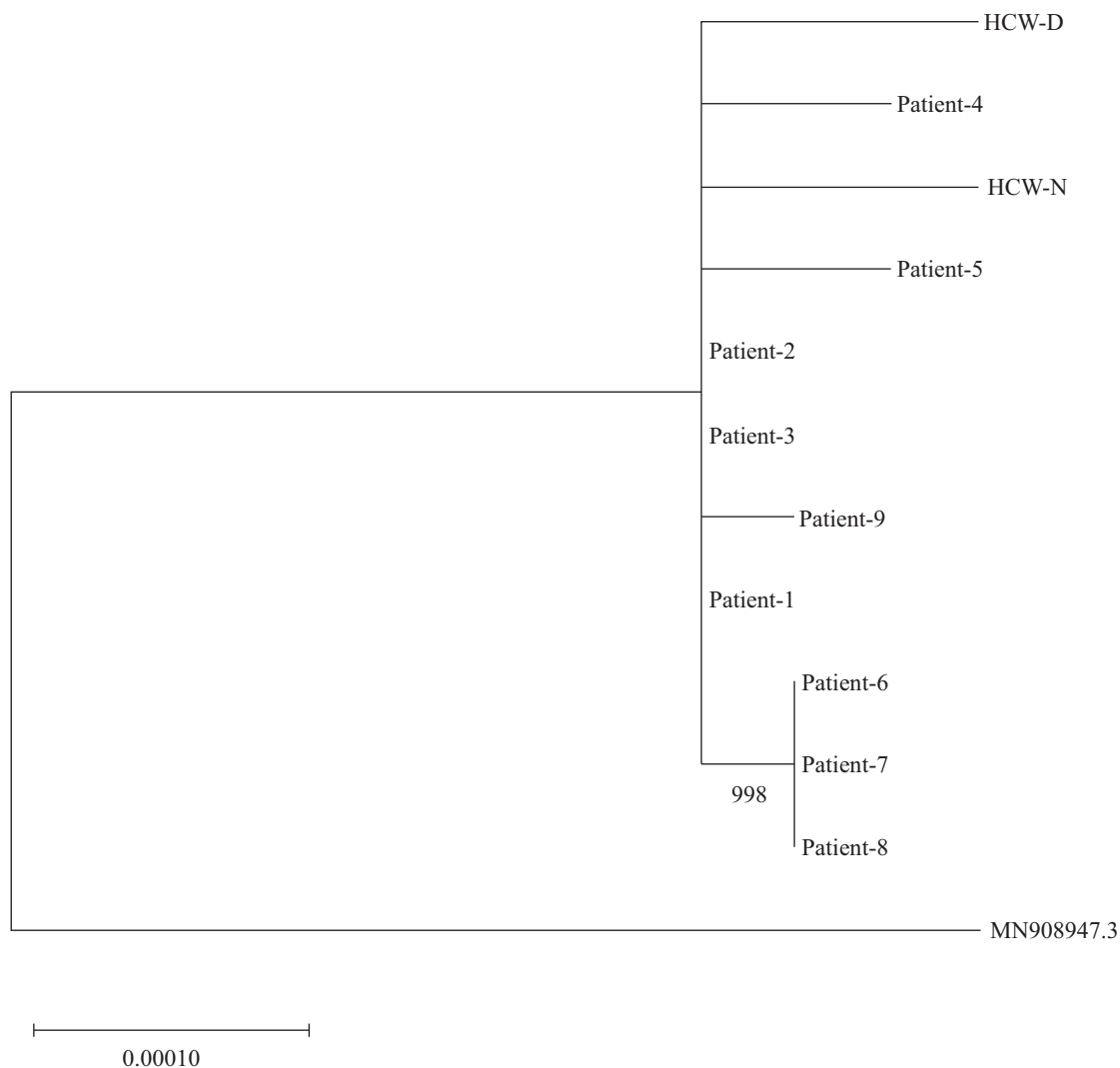
Figure 1. Phylogenetic tree of 11 SARS-CoV-2 genomes collected from two healthcare workers and nine patients. It was constructed by Mega X and PhyML using maximum likelihood with bootstrap value set at $1000 \times$. The tree was rooted on the earliest reference genome Wuhan-Hu-1 (GenBank no.: MN908947.3). The genome name of sample was labelled as the following: healthcare workers (HCW-D, doctor; HCW-N, nurse) and patients (patients 1−9) in which patient 1 was the ICU patient.

links with no SNP difference in viral genome strongly indicated the existence of nosocomial transmission of COVID-19 [3]. However, as spontaneous mutations might emerge over time within a newly infected host, the presence of one or two nucleotide differences might not be sufficient to rule out the transmission linkage among patients/HCWs. The evolutionary rate of the circulating SARS-CoV-2 lineage could be considered as a basis to determine whether the virus is likely to accumulate a certain number of mutations over the transmission period. Large numbers of SNP difference among the cases suggested that the infections would more likely have been acquired from different sources.

In our study period (January 4th to 14th, 2021), the total number of infected cases in this city was 450. Despite the low prevalence, it was concerning that nosocomial infection might have occurred as two HCWs became infected. Both cases would have been identified as patient-to-HCWs transmission if based on routine epidemiological investigation alone. WGS, however, provided high-resolution information to rule out this route of transmission. Our data also reassured that the current guidance for infection control measures and use of PPE protected the HCWs from contracting infection from patients and from other HCWs. WGS is therefore an indispensable technique in the investigation of suspected nosocomial transmission.

The major limitation of this study was that we did not include all the community-acquired cases over the study period (14 days before date of symptom onset of HCW cases) for phylogenetic analysis. However, all the patients were epidemiologically linked with known local community cases. In order to determine whether the B.1.36.27 lineage is the prevalent lineage in our locality under the same time-period, a total of nine samples collected from known community-acquired COVID-19 patients (epidemiologically unrelated to our study) had been sequenced and were found to be B.1.36.27 lineage. Therefore, our samples may represent the local circulating variants of SARS-CoV-2 in the community with low prevalence of COVID-19. Moreover, we only provided evidence to rule out direct patient-to-HCW transmission. The definite source of their infection, however, cannot be identified. It was likely that both HCWs acquired the disease in the community.

Acknowledgements

Author contributions
R.C.W. Wong: laboratory work, acquisition of data, analysis and interpretation of data, drafting, reviewing and editing the manuscript; M.K.P. Lee: conceived the study, clinical input, supervision, reviewing and editing manuscript; G.K.H. Siu: supervision, reviewing and editing manuscript, secured research funding; L.K. Lee: analysis and interpretation of data, bioinformatics; J.S.L. Leung: bioinformatics; E.C.M. Leung: sample retrieval, reviewing and editing manuscript; Y.I.I. Ho Ho: reviewing and editing manuscript; R.W.M. Lai: project administration, supervision; all authors read and approved the final manuscript.

Conflict of interest statement

Funding sources

References

[1] Höring S, Fussen R, Neusser J, Kleines M, Laurentius T, Bollheimer LC, et al. Management of a hospital-wide COVID-19 outbreak affecting patients and healthcare workers. SN Compr Clin Med 2020:1–6. https://doi.org/10.1007/s42399-020-00597-2.

[2] Løvestad AH, Jørgensen SB, Handal N, Ambur OH, Aamot HV. Investigation of intra-hospital SARS-CoV-2 transmission using nanopore whole genome sequencing. J Hosp Infect 2021. https://doi.org/10.1016/j.jhin.2021.02.022. S0195-6701(21)00081-5.

[3] 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;20:1263–71. https://doi.org/10.1016/S1473-3099(20)30562-4.

[4] Paltansing S, Sikkema RS, de Man SJ, Koopmans MPG, Oude Munnink BB, de Man P. Transmission of SARS-CoV-2 among healthcare workers and patients in a teaching hospital in the Netherlands confirmed by whole-genome sequencing. J Hosp Infect 2021;110:178–83. https://doi.org/10.1016/j.jhin.2021.02.005.

[5] Takenouchi T, Iwasaki YW, Harada S, Ishizu H, Uwamino Y, Uno S, et al. Clinical utility of SARS-CoV-2 whole genome sequencing in deciphering source of infection. J Hosp Infect 2021;107:40–4. https://doi.org/10.1016/j.jhin.2020.10.014.

[6] Bandoy DJDR, Weimer BC. Analysis of SARS-CoV-2 genomic epidemiology reveals disease transmission coupled to variant emergence and allelic variation. Sci Rep 2021;11:7380. https://doi.org/10.1038/s41598-021-86265-4.

[7] Wong RC, Wong AH, Ho YI, Leung EC, Lai RW. Evaluation on testing of deep throat saliva and lower respiratory tract specimens with Xpert Xpress SARS-CoV-2 assay. J Clin Virol 2020;131:104593. https://doi.org/10.1016/j.jcv.2020.104593.

[8] Leung KS, Ng TT, Wu AK, Yau MC, Lao HY, Choi MP, et al. Territorywide study of early coronavirus disease outbreak, Hong Kong, China. Emerg Infect Dis 2021;27:196–204. https://doi.org/10.3201/eid2701.201543.

[9] Loman N, Rambaut A. nCoV-2019 novel coronavirus bioinformatics environment setup. Available from: https://artic.network/ncov-2019/ncov2019-it-setup.html [last accessed June 2021].

[10] Cheng VC, Fung KS, Siu GK, Wong SC, Cheng LS, Wong MS, et al. Nosocomial outbreak of COVID-19 by possible airborne transmission leading to a superspreading event. Clin Infect Dis 2021 Apr 14. https://doi.org/10.1093/cid/ciab313 [Epub ahead of print].