Investigation of intra-hospital SARS-CoV-2 transmission using nanopore whole-genome sequencing

A.H. Løvestad [a,b], S.B. Jørgensen [b], N. Handal [b], O.H. Ambur [a,1], H.V. Aamot [b,c,*,1]

[a] Faculty of Health Sciences, Oslo Metropolitan University, Oslo, Norway
[b] Department of Microbiology and Infection Control, Akershus University Hospital, Lørenskog, Norway
[c] Department of Clinical Molecular Biology (Epigen), Akershus University Hospital and University of Oslo, Lørenskog, Norway

ARTICLE INFO

SUMMARY

Background: During the SARS-CoV-2 pandemic, healthcare workers (HCWs) are being exposed to infection both at work and in their communities. Determining where HCWs might have been infected is challenging based on epidemiological data alone. At Akershus University Hospital, Norway, several clusters of possible intra-hospital SARS-CoV-2 transmission were identified based on routine contact tracing.

Aim: To determine whether clusters of suspected intra-hospital SARS-CoV-2 transmission could be resolved by combining whole genome sequencing (WGS) of SARS-CoV-2 with contact tracing data.

Methods: Epidemiological data were collected during routine contact tracing of polymerase chain reaction-confirmed SARS-CoV-2-positive HCWs. Possible outbreaks were identified as wards with two or more infected HCWs defined as close contacts who tested positive for SARS-CoV-2 less than three weeks apart. Viral RNA from naso-/oropharyngeal samples underwent nanopore sequencing in direct compliance to the ARTIC Network protocol.

Findings: Five outbreaks were suspected from contact tracing. Viral consensus sequences from 24 HCWs, two patients, and seven anonymous samples were analysed. Two outbreaks were confirmed, one refuted, and two remained undetermined. One new potential outbreak was discovered.

Conclusion: Combined with epidemiological data, nanopore WGS was a useful tool for investigating intra-hospital SARS-CoV-2 transmission. WGS helped to resolve questions about possible outbreaks and to guide local infection prevention and control measures.

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

Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) pandemic has created a heavy strain on healthcare workers (HCWs) treating COVID-19 patients. In addition to the risk of burnout and psychological distress reported, HCWs may also be at risk of infection at work [1,2]. Furthermore, HCWs may also be a source of infection for

* Corresponding author. Address: Akershus University Hospital, Department of Microbiology and Infection Control, Box 1000, Lørenskog, 1478, Norway. Tel.: +47 93850682.
E-mail address: hege.vangstein.aamot@ahus.no (H.V. Aamot).
[1] These authors contributed equally.

patients and colleagues via asymptomatic carriage and transmissibility prior to the onset of symptoms. Due to the high mortality rate of COVID-19 among the elderly, a particularly difficult challenge has been to avoid virus entry to nursing homes and hospitals. Studies describing the risk and events of SARS-CoV-2 intra-hospital transmission are discrepant [3—8]. During an epidemic, when there is frequent viral transmission in the community, it is not always clear whether HCWs are infected at work or during their spare time. As the pandemic is constantly evolving, new awareness towards specific variants has soared from fear of strains more transmissible, pathogenic and likely to evade immunization efforts. As outbreak definitions vary and outbreak reports mainly depend on epidemiological data with SARS-CoV-2 test results, the true transmission patterns remain uncertain [9]. An aggregation of infected HCWs in a ward over some days or weeks does not necessarily imply intra-hospital transmission or a local outbreak.

High-throughput sequencing technology enables the investigation of microbial outbreaks and transmissions at high resolution, including those of SARS-CoV-2. With an aim to reduce time from sampling to interpretable epidemiological results in viral outbreaks, the ARTIC network was established in the UK and is now a global effort having partnered with the World Health Organization and other public health bodies worldwide (https://artic.network/ncov-2019). Through the employment of portable sequencing instruments and rigging an online integrative analysis platform, the protocols, primers, and bioinformatics tools devised by the ARTIC network allow for real-time epidemiology of the SARS-CoV-2 outbreak. Yet, only a few studies have been published in which whole-genome sequencing (WGS) has been combined with epidemiological data to trace possible transmission chains in healthcare settings [10—15].

In this cross-sectional study, the aim was to employ the ARTIC network protocol and to combine resulting SARS-CoV-2 whole-genome sequence data with contact tracing data to determine whether clusters of suspected intra-hospital SARS-CoV-2 transmission could be resolved.

Methods

Contact tracing and epidemiological data

Akershus University Hospital is a secondary emergency care hospital in Norway. It serves 640,000 people (12% of Norway's population) with approximately 1000 beds and 10,000 employees. Between March 5th, 2020 and July 1st, 2020, a total of 200 COVID-19 patients had been admitted to the hospital. The patients were treated in designated COVID-19 wards or in the intensive care unit in single or double rooms, including bathrooms. HCWs caring for COVID-19 patients used personal protective equipment (PPE) in the form of gloves, gowns, goggles and surgical face masks (respiratory masks if performing aerosol-generating procedures). Other infection prevention and control measures initiated in the hospital to contain the spread of SARS-CoV-2 included testing of patients and HCWs, isolation of SARS-CoV-2-infected patients, contact tracing around all SARS-CoV-2-infected patients and employees, quarantine of close contacts, visitors restrictions, and enhanced cleaning routines. Masks or other PPE were not worn

in contact with patients or colleagues without symptoms or suspected infection.

Patients were tested for SARS-CoV-2 upon admission to the hospital if they had any respiratory, gastrointestinal, or central nervous system symptoms of infection, fatigue, or myalgia. Patients who developed any of these symptoms during their stay were also tested. Strict testing criteria were applied for HCWs in March 2020 (fever, cough, or shortness of breath), but changed during April 2020 to include any symptoms of respiratory or gastrointestinal tract infections, headaches, myalgia or fatigue. Symptomatic HCWs were tested regardless of whether they had had any contact with known SARS-CoV-2-infected individuals, either at work or in the community. Close contacts of positive cases (whether patients or HCWs) were kept in quarantine, but not routinely tested unless they developed symptoms.

The hospital's infection control staff routinely recorded epidemiological data during concurrent contact tracing of each reverse transcription—polymerase chain reaction (RT—PCR)-confirmed SARS-CoV-2-infected HCW. A close contact was defined as a person who had had physical contact with the infected HCW without use of PPE, or who had been in close proximity (<2 m) without PPE for >15 min to the infected HCW, starting from 24 h (48 h from June 2020) before the onset of symptoms. All close contacts were quarantined for 14 days (10 days since May 2020).

For the period from March 10th, 2020 to July 1st, 2020, possible outbreaks were searched for by identifying wards with two or more infected HCWs who had had close contact as previously defined, and who tested positive for SARS-CoV-2 less than three weeks apart. If we had a suspected outbreak in a ward, all isolates from HCWs in those wards were included in the study, regardless of documentation of close contact between all the cases. All the suspected outbreaks in the somatic wards occurred in wards that were designated COVID-19-wards, and where the HCWs used PPE when caring for patients. Hence, the patients were not included as close contacts, unless there were reported or suspected breaches of infection control practices.

To assess the local diversity of SARS-CoV-2, we also included viral genomes from some HCWs who had no known connection to other cases in the hospital, and who worked in different units, and some viral genomes from anonymous patients in the hospital.

The numbers of eligible and included samples are presented in Supplementary Figure S1.

RNA isolation

RNA was isolated using an easyMAG extractor following the manufacturer's instructions for extraction of total nucleic acids from airways samples (bioMérieux, Marcy-l'Etoile, France). The qualitative RT—PCR detects the SARS-CoV-2 virus E-gene based on a method published by Corman *et al.* [16]. The eluate and samples of all positive RT—PCR are routinely stored at −80°C.

Library preparation and sequencing

Eluted RNA from 46 samples were reverse-transcribed and PCR-amplified using information provided by ARTIC Network (https://artic.network/ncov-2019). Briefly, the method uses

random hexamers for RT and multiplex PCR amplification of cDNA using a tiling amplicon scheme and the ARTIC nCoV-2019 version 3 primer set [17]. The annealing temperature of the PCR reaction was lowered to 63°C to increase amplification efficiency of problematic primer pairs. The PCR products were sequenced on a GridION sequencer (Oxford Nanopore Technologies, Oxford, UK). Five of the samples were sequenced twice to assess the reproducibility of the method.

Bioinformatic analysis

The COVID-19 bioinformatics Medaka-pipeline developed by the ARTIC network (https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html) was used to generate consensus sequences and call variant nucleotides relative to the reference sequence. Called variants were visualized in Geneious Prime (v2020.0.4) for validation using the BAM-files generated from the Artic pipeline. For the re-sequenced samples, the sample with the highest coverage was used for further analysis after determining the reproducibility of the method.

Phylogenetic analysis, Nextstrain clade assortment, and pangolin lineage assignment

To compare the study samples in broader context, published SARS-CoV-2 genomes were downloaded from GISAID (Supplementary Table S1) as follows: all from Norway ($N = 73$); international strains from European countries where contact tracing early in the pandemic had identified cases of SARS-CoV-2 importation to Norway ($N = 250$); and samples from China ($N = 6$) with collection dates up to 1 July 2020 [18]. A multiple sequence alignment (MSA) of the sequenced samples and downloaded SARS-CoV-2 genomes from GISAID was generated using MAFFT (v7.450) with the 1PAM scoring matrix. The MSA was then manually inspected to remove low-quality sequences. FastTree (v2.1.11) was used to generate phylogenetic trees, using GTR substitution model. The phylogenetic tree was further visualized and annotated using an in-house R-script with the ggtree package (v2.2.1) [19].

Samples were assorted to clades according to the Nextstrain nomenclature [20]. Clade assortment was carried out using a combination of phylogenetic placement of the samples and the presence of clade-specific signature mutations. In cases where samples had no coverage in areas of the genome with signature mutations, variants could in some cases be extrapolated from the presence of co-mutations.

Pangolin lineage assignment was done using the Pangolin COVID-19 Lineage Assigner online tool [21].

Outbreak assessment

The data generated by the nanopore sequencing were used to confirm or refute whether cases of close contacts were part of the same transmission chain. Whereas many variants make up the different SARS-CoV-2 clades, study-unique variants were weighted when assessing whether cases were the result of a suspected hospital transmission chain. Study-unique variants were defined as SARS-CoV-2 variants that met the following two criteria: (i) variants that showed no local geographic distribution and (ii) with two or more co-occurring mutations not found together in any other genome in the GISAID database.

Ethical approval

The study was approved by Akershus University Hospital's Data Protection Official (2020_62). The data were recorded as part of the hospital's routine for outbreak investigations, as authorized by the institutional infection control programme and the Norwegian regulation of infection control in the healthcare service (FOR-2005-06-17-610).

Results

Identification of transmission clusters based on routine contact tracing

During the study period, 68 HCWs from 38 wards tested positive for SARS-CoV-2. Based on routine contact tracings it appeared that the majority of the HCWs had been infected abroad or had a household/close social contact with SARS-CoV-2 infection that preceded their own illness.

Data from 24 HCWs and two patients from 11 wards were analysed (Table I), as well as seven anonymous patient samples from our hospital. Five of the wards had two or more HCWs who tested positive for SARS-CoV-2 less than three weeks apart. In one of the wards, there had been close contact between the positive HCWs. Hence, these cases had originally not been regarded as part of the same outbreak. In the four other wards, there were five different clusters of cases in which direct transmission was suspected among some of the HCWs due to close contact or work on the same shift (Table I, outbreaks A, C, and D). In addition, there was a possible link between one HCW who reported a breach in infection control procedures during treatment of a SARS-CoV-2-infected patient (Table I, outbreak B), and a probable link between five HCWs from two different wards who all displayed COVID-19 symptoms a few days after treating the same SARS-CoV-2-positive patient (Table I, outbreak E). The remaining samples were singletons with no epidemiological links to other cases. In Table I, we list the cases by date and ward, and illustrate which cases were linked by contact tracing information, and how WGS helped us refute or confirm some of the suspected outbreaks.

Sequencing results

In total, 46 samples were sequenced on the GridION. The average genome coverage for all the samples was 84.6%. However, by removing samples with coverage <80% ($N = 9$), the coverage of the analysed samples increased to 95.5%. Thirty-three samples were chosen for downstream analysis after filtering out samples with <80% coverage and replicates (Table I and Supplementary Figure S1).

Variants analysed

In total, 273 variants were called relative to the reference genome (MN908947.3) over 62 sites. The lowest number of variants in any sample was five (HCW19 and HCW22) and the highest was 13 (Anonymous 5). The average number of variants per sample was 8.3. The reported variants were identical for all the re-sequenced samples where they shared coverage.

Table I
Assessment of intra-hospital SARS-CoV-2 transmission

Case	Sample date	Ward	Outbreak as defined by contact tracing	Outbreak as defined by WGS	Nextstrain clade	Pangolin lineage	Study-unique variants	Interpretation
HCW1	Mar 29th	1	A	Singleton	20C	B.1	C20762T	HCW1, HCW2: outbreak refuted, the two
HCW2	Apr 6th	1	A	Singleton	20B	B.1.1.64	G21724T	close contacts had virus from different clades.
HCW3	Apr 13th	1	No positive close contacts	Singleton	20A/ 20268G	B.1.35	G6419A G15438T C23481T T27384C	No close contacts included in the study, but several cases from the same ward.
HCW4	Apr 14th	1	No positive close contacts	Singleton	20C	B.1	G15380T[a]	No close contacts included in the study, but several cases from the same ward. Shares one study-unique variant with Anonymous 1.
Patient 1	Apr 19th	1	B	B	20A	B.1	G4300T[a] G7975A[a]	Patient 1, HCW5: outbreak confirmed, including two close contacts as hypothesized. Two study-unique
HCW5	Apr 27th	1	B	B	20A	B.1	G4300T[a] G7975A[a] C23185T C29095T	variants are also shared with Anonymous samples 4.
HCW6	Apr 20th	1	C	C‡	20C/24368T	B.1	T24304C	HCW6, HCW7: outbreak cannot be refuted
HCW7	Apr 22nd	1	C	C‡	20C/24368T	B.1	G21624T	or confirmed. Same clade, but there are no shared study-unique variants. Two HCWs who worked together on the same shift, but with no close contact.
HCW8	May 11th	1	D	D‡	20A	B.1		HCW8, HCW9: outbreak cannot be refuted or
HCW9	May 12th	1	D	D‡	20A	B.1	C21114T A25442G	confirmed. Close contacts from the same ward, but with no shared study-unique variants.
HCW10	Apr 10th	2	No positive close contacts	Singleton	20C/24368T	B.1	T6178C	No close contacts included in the study.
HCW11	Apr 14th	2	No positive close contacts	Singleton	20A	B.1	G17347T C23895T	No close contacts included in the study.
HCW12	Apr 14th	3	No positive close contacts	F	20A	B.1	C6706T[a]	HCW12, HCW13: new outbreak detected by WGS in two HCWs from the same ward, but with no record of
HCW13	Apr 13th	3	No positive close contacts	F	20A	B.1	C6706T[a]	close contact.
Patient 2	Jun 10th	4/5	E	E	20C/24368T	B.1	G5036A[a] G6986A[a]	Patient 2, HCW14—18: outbreak including five HCWs and one patient confirmed as likely despite use of
HCW14	Jun 19th	4	E	E	20C/24368T	B.1	G5036A[a] G6986A[a]	PPE.
HCW15	Jun 19th	4	E	E	20C/24368T	B.1	G5036A[a] G6986A[a]	
HCW16	Jun 22th	5	E	E	20C/24368T	B.1	G5036A[a] G6986A[a]	

Sample	Date	No.	PPE	Type	Clade	Lineage	Mutations	Notes
HCW17	Jun 25th	5	E	E	20C/24368T	B.1	T21965A, A25817G, G5036A[a], G6986A[a], C11824T, C23127T	
HCW18	Jul 1st	5	E	E	20C/24368T	B.1	G5036A[a,*], G6986A[a], T16515C	
HCW19	Mar 12th	6	No positive close contacts	Singleton	19A	B.2	A136C, T17247C	No epidemiological links to other cases.
HCW20	Mar 24th	7	No positive close contacts	Singleton	20C	B.1.114	C13426T, G28368A	No epidemiological links to other cases.
HCW21	May 4th	8	No positive close contacts	Singleton	20A	B.1		No epidemiological links to other cases.
HCW22	Apr 6th	9	No positive close contacts	Singleton	19A	B.2	C18086T	No epidemiological links to other cases.
HCW23	Mar 24th	10	No positive close contacts	Singleton	20A/20268G	B.1.5	G5230T, G15846T	No epidemiological links to other cases.
HCW24	Apr 6th	11	No positive close contacts	Singleton	20A/20268G	B.1.5.6	C3096T, G15438T	No epidemiological links to other cases.
Anonymous 1	Apr 21st	—	—	—	20C	B.1	C8772T, G15380T[a]	Anonymous sample, but shares one study-unique variant with HCW4.
Anonymous 2	Apr 21st	—	—	—	20C/24368T	B.1		—
Anonymous 3	Apr 24th	—	—	—	20C/24368T	B.1	C4234T	—
Anonymous 4	Apr 22nd	—	—		20A	B.1	G4300T[a], G7975A[a]	Anonymous sample, but shares study-unique variants with patient 1 and HCW5.
Anonymous 5	Apr 23rd	—	—	—	20C/24368T	B.1	C6633T, G11083T, T14257C, G20433C, C22482T, G26439T	—
Anonymous 6	Apr 23rd	—	—	—	20A	B.1	G1820A, T29026C	—
Anonymous 7	Apr 23rd	—	—	—	20A	B.1		—

HCW, healthcare worker; WGS, whole-genome sequencing; PPE, personal protective equipment.

Sample HCW3 did not have coverage at position A20268G.

Sample HCW7 did not have coverage at position C3037T.

[a] Study-unique variants found in more than one sample in the dataset. Mutations shown according to Nextstrain classification system.

* Denotes samples with no coverage in the given region with variant inferred from co-mutations.

‡ Outbreaks that cannot be confirmed or refuted by WGS.

Phylogenetic analysis, Nextstrain clade assortment, and pangolin lineage assignment

The results from the phylogenetic analysis and clade assortment showed that the samples mainly clustered into two large clades based on shared mutation profiles (Figure 1, Supplementary Table S1). Eleven of the samples were classified as clade 20A, with three samples clustering within the Nextstrain emerging clade 20A/20268G. Sixteen of the samples clustered within 20C and these samples clustered within two distinct groups. The largest group consisted of 12 samples that shared the G24368T mutation causing the amino acid change D936Y in the heptad repeat 1 (HR1) domain of the spike protein. The mutation profile shared between these samples has a high frequency in other Nordic countries [22]. Therefore, the name 20C/24368T is used when referring to this group to distinguish them from the rest. Furthermore, two samples were classified as clade 19A and one as 20B. No samples were classified as clade 19B.

For the pangolin lineage assignment, the samples were assigned to lineage B.1 ($N = 26$), B.1.1.64 ($N = 1$), B.1.114 ($N = 1$), B.1.35 ($N = 1$), B.1.5 ($N = 1$), B.1.5.6 ($N = 1$), and B.2 ($N = 2$) (Table I).

*Resolving outbreaks by contact tracing, sampling times, and phylogenetic relationships*

In total, five possible outbreaks were identified based on routine contact tracing and one additional outbreak was identified based on WGS data (Table I). Groups A, C, D, and F all consist of HCWs with close contact or simultaneous work on the same ward, while outbreaks B and E consist of samples from both HCWs and patients.

Group A
Virus from HCW1 was classified as clade 20B and virus from HCW2 as 20C. Thus, they were classified as two different genetic clades and direct transmission was ruled out.

Group B
Patient 1 and HCW5 both had viruses with two variants that were neither shared with any other virus nor found with high frequency in the GISAID database (G4300T, G7975A). However, the virus from HCW5 had two additional variants. HCW5 was tested nine days after the patient. Anonymous 4 was sampled three days after the patient and shared G4300T, G7975A
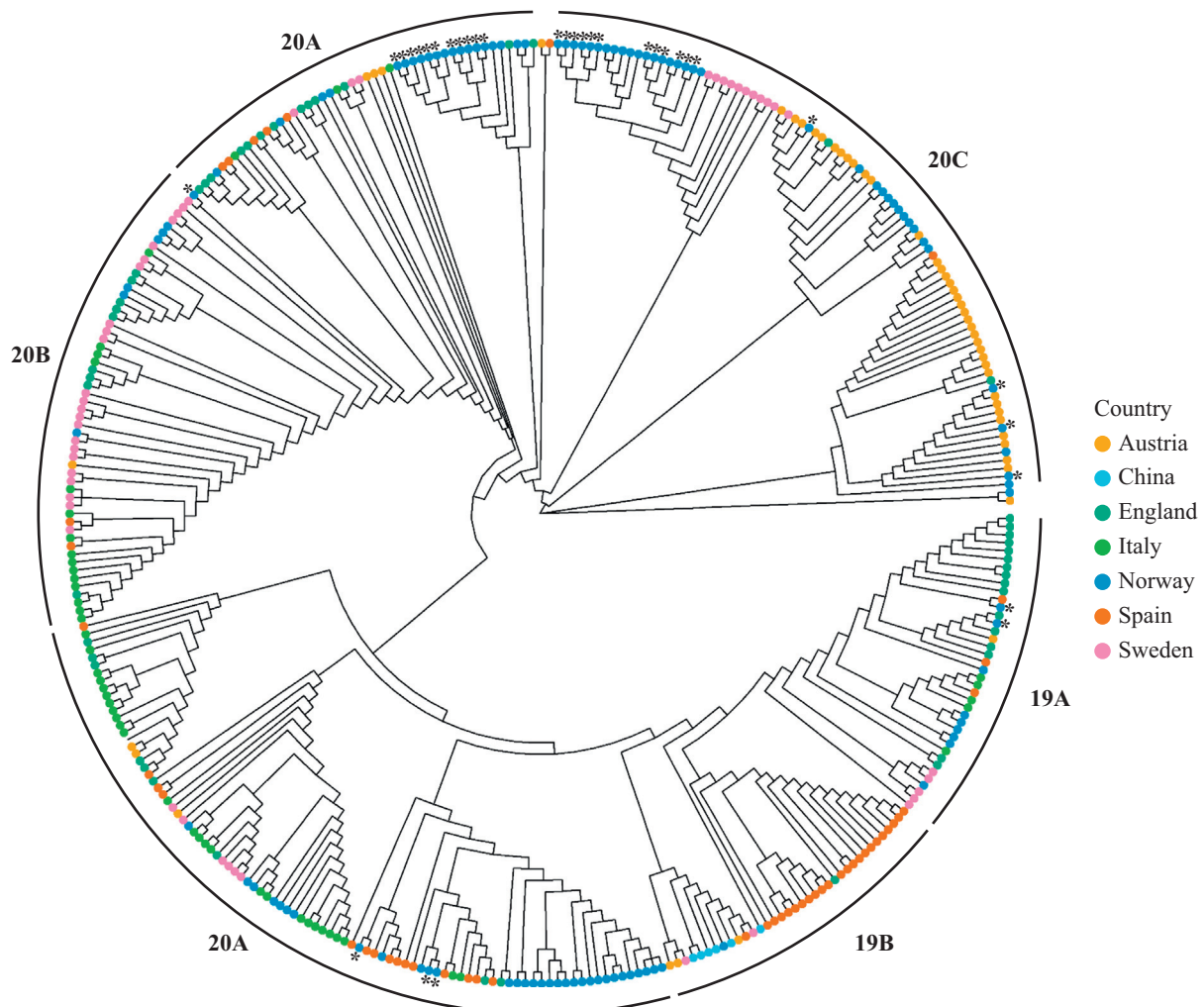
Figure 1. Phylogenetic tree of whole-genome-sequenced SARS-CoV-2 virus from Akershus University Hospital, Norway compared to all national and a selection of international viral genomes collected up until July 1st and published in the GISAID database. *Samples from this study.

without additional variants. HCW5 reported that the patient suffered from violent cough attacks, and that the PPE had felt insufficient during treatment of this patient.

Group C

Viruses from HCW6 and HCW7 were classified as 20C/24368T. However, as they both had one additional variant not shared by the other and did not share any study-unique variants, direct transmission between HCW6 and HCW7 was interpreted as uncertain.

Group D

The two viruses were classified as clade 20A. However, whereas virus from HCW8 had no additional variants, virus from HCW9 had two (C21114T, A25442G). These HCWs were close contacts and worked on the same ward for several shifts at a time when there was very low transmission activity in the community [23]. It is likely that they were linked in a transmission chain within the ward, but since the viruses did not share any unique variants this cannot be certain.

Group E

This group consists of primary case (Patient 2) and five samples from HCWs (HCW14—18) known to have interacted with them. Contact tracing indicated that HCW14—18 were all infected during the same shift. The viruses in this group shared the clade-defining G24368T variant and two study-unique variants (G5036A, G6986A). During this study a Norwegian sample was submitted to the GISAID database (Norway/2829/2020) harbouring the same three co-mutations (G5036A, G6986A, G24368T), leaving our set of variants in outbreak E not strictly study-unique according to the defined criteria. The Norway/2829/2020 sample was taken on June 29th, 2020, towards the end of this outbreak investigation, predating only the sample obtained from HCW18 (July 1st, 2020). The viruses from the patient and the two HCWs at ward 4 (HCW14, HCW15) were identical. HCW14 and HCW15 were tested on the same day and shortly after their only contact with the patient — nine days after the patient had been tested (June 10th, 2020). The viruses from the three HCWs from ward 5, where the patient was later transferred (HCW16—18), each had one or two additional non-shared variants. These samples were taken 12, 15, and 21 days after the patient's sample. The associations between contact tracing, individual sample timelines, and viral genotypes suggest a common source of infection in outbreak E. In addition, the low incidence of COVID-19 in the region at the time suggests that a common source of infection was to be found at the hospital and not in the community [23]. We elaborate on the appearance of non-shared variants in the Discussion.

Group F

HCW12 and HCW13 had no close contact according to definitions used in the contact tracing, but they worked in the same ward during the same week. Their viral samples had identical sequences and they shared the study-unique variant C6706T. Hence, they were most likely part of the same transmission cluster. This potential outbreak within the hospital would have gone undetected without the use of WGS data.

Discussion

By adding WGS of SARS-CoV-2 virus to routine contact tracing in investigations of hospital outbreaks, this study shows both the potential power and challenges with high-resolution genotyping in local outbreak settings. Of the five suspected outbreaks, two were confirmed, two remain undetermined and one was refuted. In addition, one new possible transmission was detected, previously unidentified by routine contact tracing. Based on high-resolution genomic data, the timely implementation of SARS-CoV-2 WGS can guide local infection prevention and control measures. With the emergence of novel variants in the second and third waves of the COVID-19 pandemic with feared capabilities, the importance of swiftly obtaining high-resolution genomic SARS-CoV-2 data cannot be overstated. Different protective measures from PPE and personal behaviour recommendations to regional and national lockdowns and curfews are now guided by case-counts prioritized with data at the virus variant level. The rapid detection of new and potentially more transmissible strains in hospitals can raise the alert and devise even higher safety measures including HCW routines and staff rotation.

So far, data from the GISAID database has been useful for detecting potential structural changes in the virus, monitoring large-scale transmission dynamics, potential antigenic drift and SARS-CoV-2 evolution [24—27]. However, until now, there have been few attempts to use WGS in real-time outbreak investigations. In a retrospective cross-sectional Dutch study, genomes from three different hospitals were compared to genomes previously entered in GISAID, allowing the researchers to conclude that nosocomial transmission was probably not a common source of infection among the HCWs studied [11]. A British prospective surveillance study found possible transmission links involving patients and symptomatic HCWs, although it was not reported whether the HCWs were index cases [10].

The confirmed outbreaks in our study contained samples that all shared study-unique variants. By emphasizing the presence of study-unique variants instead of using a predetermined cut-off of maximum allowed differences in variants to determine intra-hospital transmissions, we lean into a more stringent confirmation criterion than other studies. This approach was chosen because SARS-CoV-2 is a novel human virus with low genetic diversity, and there were few SARS-CoV-2 genomes from Norway available online for comparison at the time of analysis [26].

With limited data on the genetic background of virus circulating in the community, and few available genomes from hospital patients, we cannot confidently conclude that all our seemingly linked cases by contact tracing were in fact intra-hospital transmissions. In the one suspected outbreak that was refuted, the samples belonged to different Nextstrain major clades and pangolin lineages, with several different variants reported. These cases are the easiest to resolve using WGS data, as the number of variants that distinguish them makes the probability for linked transmission during a short timeframe infinitesimally small. Hence, this method is, for the time being, a stronger tool for refuting outbreaks than for confirming outbreaks when used on its own. When suspected outbreak genomes fall into different clades, they are not from the same intra-hospital transmission chain.

However, the real challenge is that of cases that do not share any study-unique variants, but which belong to the same clade and are genetically very similar. It is difficult to determine whether the few variants they do not share are the result

of genetic variability in the viral genomes from a common source, de-novo mutations that have developed over time within the study participants, or due to infection from different sources. There is still little research into intra-host variation and the effects of transmission bottlenecks of SARS-CoV-2, but more knowledge in this field may help us interpret outbreaks at finer resolution [28–30]. In-hospital studies such as this may be helpful in studying these anticipated effects on transmission dynamics and genetic variability, since the environment and SARS-CoV-2 infections therein are tightly monitored and controlled.

Mutations found in one case but not in others from the same outbreak may be due to mutations that arise in the new host *de novo*. The association between sampling times and new variants supports the notion that new variants are generated in HCWs during outbreak B and E in the periods between suspected transmission events and sampling. Late acquired (>8 days) samples in both confirmed outbreaks carry individual variants ($N = 7$) in a total of four HCWs not found in the primary cases (Patients 1 and 2). Community acquisition of these unique variants is considered highly unlikely for HCWs who were under strict regimens to avoid SARS-CoV-2 infections at work and in their spare time as Norway was in lockdown (March 12th to July 15th, 2020). The SARS-CoV-2 mutation rate is estimated to result in about two mutations per month and it is possible that the consensus genome differs by one, or even two, nucleotides from one case to the next, especially if the date of sampling differs by about seven days, as is the situation in some of our cases [31]. Further studies are required to determine intra-person mutation rates and minor viral allelic diversities (i.e. signs of de-novo generation of mutations) in the context of COVID-19 disease severity and SARS-CoV-2 infectivity.

This study also discovered a potential outbreak using WGS data that would otherwise go unnoticed (outbreak F). This, again, shows the advantages of incorporating information from WGS technology to guide local infection prevention and control measures.

Oxford Nanopore sequencers have been used to investigate the global spread of SARS-CoV-2 from its origin in China and to follow transmissions between and within countries. This genomic information has been valuable in identifying local clusters of transmission and for evaluating the effect of preventive measures, as shown in studies from Iceland, China, and USA [32–34]. Several studies have used the Oxford Nanopore sequencing platform to generate whole genome sequences of SARS-CoV-2 and the technology produces highly accurate consensus-level results [35].

Regarding the reproducibility of the method, all re-sequenced samples in our set called the same variants relative to the reference genome. While nanopore sequencing has been shown to have a high per-read error rate, the strategy of generating consensus sequences from samples sequenced with enough depth overcomes this problem [36]. There were some differences in the coverage between the replicates; however, this is attributed to stochastic processes in the PCR reaction from primer performances and not the sequencing step. Using nanopore sequencing in real-time surveillance and outbreak investigation would help with better identification and demarcation of outbreaks and limit further spread by aiding the implementation of targeted measures. However, our results show that the analysis is dependent on samples with C_T-

value <33 for consistent amplification efficiency and consequently high genome coverage (Supplementary Figure S2). Due to the low start-up cost, portability, in addition to the short time from sampling to interpretable and actionable results, nanopore sequencing is also well suited for 'lab-in-a-suitcase' initiatives where sequencing core facilities are missing.

Because we did not have the resources to sequence the viral genomes from all the patients who had been cared for by the infected HCWs, our study could not be used to investigate possible transmissions between HCWs with PPE and their patients in general. However, our sample includes one outbreak (outbreak E) in which the HCWs all wore PPE as recommended by the Norwegian Institute for Public Health and where no breach in infection control procedures was reported. The patient was transferred from a regular ward where the staff used surgical masks, eye protection, coats, and gloves to an intensive care unit where the staff wore the same equipment but with FFP2 or FFP3 respirators instead of surgical masks. All five HCWs who cared for the patient that one night were infected regardless of which mask was used. Hence, this is an example of a super-spreader event where a single person infected several other individuals within only a few hours. The patient had severe cough and respiratory failure and was treated with an oxygen mask with a flow of 12 L/min before transfer to the ICU.

In terms of patient safety and for the protection of HCWs, it is important to monitor and examine any possible SARS-CoV-2 outbreaks in healthcare settings. Our results show that nanopore WGS was a useful tool for investigating intra-hospital SARS-CoV-2 transmission in combination with epidemiological data. Epidemiological tracing alone falsely identified one hospital outbreak and overlooked one outbreak. WGS can provide a better understanding of nosocomial transmission pathways and allow for necessary and timely adaptations of local infection prevention and control routines.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhin.2021.02.022.

References

[1] Matsuo T, Kobayashi D, Taki F, Sakamoto F, Uehara Y, Mori N, et al. Prevalence of health care worker burnout during the coronavirus disease 2019 (COVID-19) pandemic in Japan. JAMA Netw Open 2020;3:e2017271. https://doi.org/10.1001/jamanetworkopen.2020.17271.

[2] Shechter A, Diaz F, Moise N, Anstey DE, Ye S, Agarwal S, et al. Psychological distress, coping behaviors, and preferences for support among New York healthcare workers during the COVID-19 pandemic. Gen Hosp Psychiatry 2020;66:1—8. https://doi.org/10.1016/j.genhosppsych.2020.06.007.

[3] McMichael TM, Currie DW, Clark S, Pogosjans S, Kay M, Schwartz NG, et al. Epidemiology of covid-19 in a long-term care facility in King County, Washington. N Engl J Med 2020;382:2008—11. https://doi.org/10.1056/NEJMoa2005412.

[4] Arons MM, Hatfield KM, Reddy SC, Kimball A, James A, Jacobs JR, et al. Presymptomatic SARS-CoV-2 infections and transmission in a skilled nursing facility. N Engl J Med 2020;382:2081—90. https://doi.org/10.1056/NEJMoa2008457.

[5] Basso T, Nordbø SA, Sundqvist E, Martinsen TC, Witsø E, Wik TS. Transmission of infection from non-isolated patients with COVID-19 to health care workers. J Hosp Infect 2020;106:639—42. https://doi.org/10.1016/j.jhin.2020.08.015.

[6] Folgueira MD, Munoz-Ruiperez C, Alonso-Lopez MA, Delgado R. SARS-CoV-2 infection in health care workers in a large public hospital in Madrid, Spain, during March 2020. Nat Commun 2020;11:3500. https://doi.org/10.1038/s41467-020-17318-x.

[7] Liu M, Cheng SZ, Xu KW, Yang Y, Zhu QT, Zhang H, et al. Use of personal protective equipment against coronavirus disease 2019 by healthcare professionals in Wuhan, China: cross sectional study. BMJ 2020;369:m2195. https://doi.org/10.1136/bmj.m2195.

[8] Lemieux J, Siddle KJ, Shaw BM, Loreth C, Schaffner S, Gladden-Young A, et al. Phylogenetic analysis of SARS-CoV-2 in the Boston area highlights the role of recurrent importation and super-spreading events. MedRxiv 2020;2020. https://doi.org/10.1101/2020.08.23.20178236. 08.23.20178236.

[9] O'Neil EA, Naumova EN. Defining outbreak: breaking out of confusion. J Public Health Policy 2007;28:442—55. https://doi.org/10.1057/palgrave.jphp.3200140.

[10] 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.

[11] Sikkema RS, Pas SD, Nieuwenhuijse DF, O'Toole Á, Verweij J, van der Linden A, et al. COVID-19 in health-care workers in three hospitals in the south of the Netherlands: a cross-sectional study. Lancet Infect Dis 2020;20:1273—80. https://doi.org/10.1016/S1473-3099(20)30527-2.

[12] Paltansing S, Sikkema RS, Man SJ de, Koopmans MPG, Munnink BBO, Man P de. 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 Feb 8. https://doi.org/10.1016/j.jhin.2021.02.005. S0195-6701(21)00051-7 [online ahead of print].

[13] Olmos C, Campaña G, Monreal V, Pidal P, Sanchez N, Airola C, et al. SARS-CoV-2 infection in asymptomatic healthcare workers at a clinic in Chile. PLoS One 2021;16:e0245913. https://doi.org/10.1371/journal.pone.0245913.

[14] Safdar N, Moreno GK, Braun KM, Friedrich TC, O'Connor DH. Using virus sequencing to determine source of SARS-CoV-2 transmission for healthcare worker. Emerg Infect Dis 2020;26:2489—91. https://doi.org/10.3201/eid2610.202322.

[15] Lucey M, Macori G, Mullane N, Sutton-Fitzpatrick U, Gonzalez G, Coughlan S, et al. Whole-genome sequencing to track severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission in nosocomial outbreaks. Clin Infect Dis 2020. https://doi.org/10.1093/cid/ciaa1433. ciaa1433 [online ahead of print].

[16] Corman VM, Landt O, Kaiser M, Molenkamp R, Meijer A, Chu DKW, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT—PCR. Eurosurveillance 2020;25. https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045.

[17] Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc 2017;12:1261—6. https://doi.org/10.1038/nprot.2017.066.

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

[19] Yu G. Using ggtree to visualize data on tree-like structures. Curr Protoc Bioinforma 2020;69. https://doi.org/10.1002/cpbi.96.

[20] Hodcroft EB, Hadfield J, Neher RA, Bedford T. Year-letter genetic clade naming for SARS-CoV-2 on nextstrain.org. Nextstrain 2020;2020.

[21] COG-UK, n.d. https://pangolin.cog-uk.io/[last accessed November 2020].

[22] Ling J, Hickman RA, Li J, Lu X, Lindahl JF, Lundkvist Å, et al. Spatio-temporal mutational profile appearances of Swedish SARS-CoV-2 during the early pandemic. Viruses 2020;12:1026. https://doi.org/10.3390/v12091026.

[23] ukesrapport FHI. Oppsummering uke 2020;20.

[24] Jaimes JA, André NM, Chappie JS, Millet JK, Whittaker GR. Phylogenetic analysis and structural modeling of SARS-CoV-2 spike protein reveals an evolutionary distinct and proteolytically sensitive activation loop. J Mol Biol 2020;432:3309—25. https://doi.org/10.1016/j.jmb.2020.04.009.

[25] Benvenuto D, Angeletti S, Giovanetti M, Bianchi M, Pascarella S, Cauda R, et al. Evolutionary analysis of SARS-CoV-2: how mutation of Non-Structural Protein 6 (NSP6) could affect viral autophagy. J Infect 2020;81:e24—7. https://doi.org/10.1016/j.jinf.2020.03.058.

[26] Dearlove B, Lewitus E, Bai H, Li Y, Reeves DB, Joyce MG, et al. A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. Proc Natl Acad Sci USA 2020;117:23652—62. https://doi.org/10.1073/pnas.2008281117.

[27] 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;182:812—27. https://doi.org/10.1016/j.cell.2020.06.043. e19.

[28] Shen Z, Xiao Y, Kang L, Ma W, Shi L, Zhang L, et al. Genomic diversity of severe acute respiratory syndrome-coronavirus 2 in patients with coronavirus disease 2019. Clin Infect Dis 2020;71:713—20. https://doi.org/10.1093/cid/ciaa203.

[29] Wang D, Wang Y, Sun W, Zhang L, Ji J, Zhang Z, et al. Population bottlenecks and intra-host evolution during human-to-human transmission of SARS-CoV-2. BioRxiv 2020. https://doi.org/10.1101/2020.06.26.173203. 2020.06.26.173203.

[30] Pfefferle S, Guenther T, Kobbe R, Czech-Sioli M, Noerz D, Santer R, et al. Low and high infection dose transmission of SARS-CoV-2 in the first COVID-19 clusters in Northern Germany. MedRxiv 2020 Jun 16. https://doi.org/10.1101/2020.06.11.20127332.

[31] 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(2):veaa061. https://doi.org/10.1093/ve/veaa061.

[32] Lu J, du Plessis L, Liu Z, Hill V, Kang M, Lin H, et al. Genomic epidemiology of SARS-CoV-2 in Guangdong Province, China. Cell 2020;181:997—1003.e9. https://doi.org/10.1016/j.cell.2020.04.023.

[33] Gonzalez-Reiche A, Hernandez M, Sullivan M, Ciferri B, Alshammary H, Obla A, et al. Introductions and early spread of SARS-CoV-2 in the New York City area 2020. Science 2020;369(6501):297—301. https://doi.org/10.1126/science.abc1917.

[34] Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL, et al. Spread of SARS-CoV-2 in the Icelandic population. N Engl J Med 2020;382:2302—15. https://doi.org/10.1056/NEJMoa2006100.

[35] Bull RA, Adikari TN, Ferguson JM, Hammond JM, Stevanovski I, Beukers AG, et al. Analytical validity of nanopore sequencing for rapid SARS-CoV-2 genome analysis. Nat Commun 2020;11:1–8. https://doi.org/10.1038/s41467-020-20075-6.

[36] Bowden R, Davies RW, Heger A, Pagnamenta AT, de Cesare M, Oikkonen LE, et al. Sequencing of human genomes with nanopore technology. Nat Commun 2019;10. https://doi.org/10.1038/s41467-019-09637-5.