Coronavirus disease (COVID-19) refers to a cluster of viral pneumonia cases that first occurred in Wuhan, a city in Hubei Province, China, beginning in December 2019. Etiology was unknown during the early stage of the outbreak until a novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was isolated on January 7, 2020, and the genome was sequenced (1).

During the initial outbreak, fever was the main symptom of COVID-19, and about one third of patients experienced acute respiratory distress syndrome. Approximately 16% of patients were in severe condition at admission, and the estimated mortality rate was 1.4% (2). Sustained human-to-human transmission was confirmed upon identification of cases clustering among families and transmission from patients to healthcare workers (3,4), which triggered China’s urgent public health actions and international concern.

As of February 28, 2020, a total of 78,824 COVID-19 cases had been diagnosed in mainland China and 2,788 persons had died. The disease had also spread to 50 other countries (5). The World Health Organization declared COVID-19 a pandemic in March 2020.

In Hong Kong, the first imported case was identified on January 23, 2020; the case-patient was a resident of mainland China who traveled to Hong Kong from Wuhan through Shenzhen by high-speed rail. The first local case with unknown source (i.e., patient who had no travel record during the 14-day incubation period) was reported on February 4, 2020 (6). By February 28, 2020, a total of 93 COVID-19 cases had been diagnosed in mainland China and 2,788 persons had died. The disease had also spread to 50 other countries (5). The World Health Organization declared COVID-19 a pandemic in March 2020.

In Hong Kong, the first imported case was identified on January 23, 2020; the case-patient was a resident of mainland China who traveled to Hong Kong from Wuhan through Shenzhen by high-speed rail. The first local case with unknown source (i.e., patient who had no travel record during the 14-day incubation period) was reported on February 4, 2020 (6).

By February 28, 2020, a total of 93 COVID-19 cases had been diagnosed in Hong Kong; ≥70 (75.3%) of those were local cases and those case-patients’ close contacts (6,7). Secondary and tertiary transmissions were observed in some case clusters. Because source of infection is unknown in most index cases in these clusters, hidden transmission chains were believed to be present in the community.
We report the demographic, clinical, and epidemiologic data of 50 hospitalized patients who accounted for 53.8% of COVID-19 cases in Hong Kong at the data cutoff point (February 28, 2020), including 3 imported cases and 6 transmission clusters of local infections. We characterized viral genomes in all these cases by using nanopore and Illumina sequencing. Phylogenetic and molecular evolutionary analyses were performed to determine the transmission link and the evolutionary rate of COVID-19 cases in the community.

Methods

Cases
For this retrospective, multicenter study, we enrolled case-patients with laboratory-confirmed COVID-19 from 4 public hospital clusters managed under the Hospital Authority of Hong Kong, namely Hong Kong East Cluster, Kowloon East Cluster, Kowloon West Cluster, and New Territories West Cluster, during January 26–February 28, 2020. Sputum specimens and throat swab specimens pooled with nasopharyngeal aspirates were collected from patients who fulfilled the reporting or enhanced surveillance criteria at hospital admission (8). Laboratory-confirmed infection was defined as the detection of SARS-CoV-2 by real-time reverse transcription PCR, which amplified the envelope (E) gene and RNA-dependent RNA polymerase (RdRp) gene (9).

We obtained demographic, clinical, and microbiologic data from patients’ medical records. Epidemiologic information was retrieved from the Centre for Health Protection of the Department of Health (6) and the website https://wars.vote4.hk (7). The definitions of clinical symptoms and complications are based on World Health Organization guidance (10). We adopted the Centre for Health Protection case numbering system, which is based on the date of case confirmation. This study was approved by the Institutional Review Boards of The Hong Kong Polytechnic University (approval no. RSA20021) and the public hospitals involved (HKECREC-20200014; KCC/KEC-20200070; KWC-20200040; NTWC-20200038).

Specimen Preparation
The respiratory specimens were centrifuged at 16,000 \( \times g \) for 2 minutes. Total nucleic acid was extracted from supernatant using MagNA Pure 96 System (Roche, https://lifescience.roche.com) or NucliSENS easyMAG (bioMérieux, https://www.biomerieux-nordic.com) according to the manufacturers’ instructions. DNase treatment was done by using TURBO DNA-free Kit (ThermoFisher Scientific, https://www.thermofisher.com) to remove residual host DNA.

Reverse Transcription and Viral Genome Amplification Using Multiplex PCR
DNase-treated RNA was reverse-transcribed using random hexamers and SuperScript IV Reverse Transcriptase (ThermoFisher Scientific) as previously described (11). Viral cDNA was then amplified by using 2 PCRs containing tiled, multiplexed primers (Appendix 1 Table 1, https://wwwnc.cdc.gov/EID/article/27/1/20-1543-App1.xlsx) described in the ARTIC protocol (https://artic.network/ncov-2019) (12). Details of the multiplex PCR are provided in Appendix 2 (https://wwwnc.cdc.gov/EID/article/27/1/20-1543-App2.pdf).

Nanopore MinION Sequencing
Ligation-based 1D sequencing was carried out by using Litigation Sequencing Kit SQK-LSK109 (Oxford Nanopore Technologies, https://nanoporetech.com) according to manufacturer’s instructions. Multiplex PCR amplicons of each sample were normalized to 1 ng/µL before end-repair and native barcode ligation by using EXP-NBD104/114 (Oxford Nanopore Technologies). Barcoded samples were pooled and ligated to AMII sequencing adaptor. Sequencing was performed with Nanopore MinION device (Oxford Nanopore Technologies) by using R9.4.1 flow cell for 48 hours.

Illumina MiSeq Sequencing
Multiplex PCR amplicons were subjected to library preparation and dual-indexing by using KAPA HyperPrep Kit and Unique Dual-Indexed Adaptor Kit (Roche) according to manufacturer’s instructions. Ligated libraries were enriched by 6-cycle PCR amplification and purification and size selection by using AMPure XP beads (Beckman Coulter, https://www.beckmancouler.com/). The pooled library was sequenced with the MiSeq Reagent Kit V2 Nano on an Illumina MiSeq System (Illumina, https://www.illumina.com).

Bioinformatic Analysis
We analyzed nanopore sequencing data using modified Artic Network nCoV-2019 novel coronavirus bioinformatics protocol (Appendix 2) (13). Illumina sequencing reads were mapped with reference to respective consensus genome of each sample constructed from nanopore data. Variants were called by using freebayes version 1.0.0 (https://github.com/freebayes/
freebayes) with haploid decoding and minimum base quality set at Q30. Consensus genomes were constructed by GATK 4.1.4.1 based on the VCF file (14). SPAdes genome assembler 3.14.0 (https://cab.spbu.ru/software/spades) and minimap2 version 2.17 (https://anaconda.org/bioconda/minimap2) were used to combine nanopore and Illumina sequencing results for de novo assembly and to identify the sequence of the unmapped gap regions. The sequences have been submitted to GenBank (accession nos. MT232662711).

Genomic and Phylogenetic Analysis
To identify the amino acid change caused by each single-nucleotide polymorphism, we BLAST-searched the consensus genome of each specimen against the reference NC_045512.2 using BLASTX (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Nonsynonymous mutations were identified using custom Python script (https://github.com/kenssl/Blast_mismatch_search).

Consensus genomes were aligned by Clustal Omega 1.2.4 (https://www.ebi.ac.uk/Tools/msa/clustalo). Phylogenetic tree was constructed with PhyML 3.0 (http://www.atgc-montpellier.fr/phyml/) using the maximum-likelihood algorithm. Best-fitting substitution model was selected by Akaike information criteria, in which we selected the general time-reversible model with fixed proportion of invariable sites (15). Bootstrap replicates were set at 1,000×, and maximum-likelihood phylogenetic tree was rooted on the earliest published genome (accession no. NC_045512.2). Transmission clusters were defined by clear epidemiologic and onset-time relationship. Meanwhile, we downloaded an additional 478 SARS-CoV-2 genomes from the GISAID (https://www.gisaid.org) SARS-CoV-2 data hub (16) and analyzed the phylogenetic relationships by using maximum-likelihood with bootstrap value set at 500× and rooted on SARS-CoV-2 genome NC_045512.2.

Estimation of Evolutionary Rate and Divergence
Time of Transmission
To reconstruct the evolutionary model of COVID-19 cases using the viral genomes obtained in Hong Kong, we implemented Bayesian inference through Markov Chain Monte Carlo (MCMC) framework in BEAST version 2.6.2 (17). Death rate δ (which refers to the time needed for a case-patient to become non-contagious) was determined as the lag time between date of symptom onset and date of hospital admission, because the transmission link in Hong Kong was practically stopped once the patient was hospitalized. Bayesian phylodynamic analysis was performed using strict clock and relaxed clock models with coalescent exponential growth tree priors. We ran MCMC chains for 10⁴ generations and sampled every 500 steps. Bayesian output was analyzed after the results were visualized by Tracer version 1.7.1 (18). All parameters had an effective sample size of >200, indicating sufficient sampling.

Results
Our investigation included 50 COVID-19 patients; 54.0% were women, and the mean age was 55.2 (range 22–96) years (Table 1). Of the case-patients, we categorized 3 cases as imported because the patients stayed in Wuhan before traveling to Hong Kong in mid-January. Four patients traveled to Japan and other provinces of China in mid-January and were hospitalized after they returned to Hong Kong, but active community transmission of COVID-19 was not officially reported in these areas during the study period, so these cases were considered possible local infections. The other 43 patients’ cases were categorized as local infection because of no recent travel history.

Eighteen (36.0%) patients had chronic illnesses, of which cardiovascular and cerebrovascular diseases were the most common (Table 1). In total, 74.0% of the patients were experiencing cough at admission. Fever occurred in 58.0% of patients at time of admission, but that rate gradually increased to 64.0% during the course of hospitalization. Other less common symptoms were muscle aches (25.0%), sore throat

### Table 1. Demographics, travel record, and baseline medical history of 50 coronavirus disease patients, Hong Kong, February 2020*

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Case-patients, n = 50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age group, y</td>
<td></td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>55.2 (19.5)</td>
</tr>
<tr>
<td>Range</td>
<td>22–96</td>
</tr>
<tr>
<td>&lt;30</td>
<td>8 (16.0)</td>
</tr>
<tr>
<td>31–40</td>
<td>5 (10.0)</td>
</tr>
<tr>
<td>41–50</td>
<td>6 (12.0)</td>
</tr>
<tr>
<td>51–60</td>
<td>11 (22.0)</td>
</tr>
<tr>
<td>61–70</td>
<td>10 (20.0)</td>
</tr>
<tr>
<td>&gt;71</td>
<td>10 (20.0)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>27 (54.0)</td>
</tr>
<tr>
<td>M</td>
<td>23 (46.0)</td>
</tr>
<tr>
<td>Travel record ≤14 d before symptom onset</td>
<td>7 (14.0)</td>
</tr>
<tr>
<td>Wuhan, Hubei Province, China</td>
<td>3 (6.0)</td>
</tr>
<tr>
<td>Other regions in mainland China</td>
<td>1 (2.0)</td>
</tr>
<tr>
<td>Regions outside mainland China</td>
<td>3 (6.0)</td>
</tr>
<tr>
<td>No travel record</td>
<td>43 (86.0)</td>
</tr>
<tr>
<td>Chronic medical illnesses</td>
<td>18 (36.0)</td>
</tr>
<tr>
<td>Cardiovascular and cerebrovascular</td>
<td>14 (28.0)</td>
</tr>
<tr>
<td>diseases</td>
<td></td>
</tr>
<tr>
<td>Endocrine system diseases</td>
<td>11 (22.0)</td>
</tr>
<tr>
<td>Nervous system diseases</td>
<td>5 (10.0)</td>
</tr>
<tr>
<td>Digestive system diseases</td>
<td>4 (8.0)</td>
</tr>
<tr>
<td>Malignant tumor</td>
<td>1 (2.0)</td>
</tr>
</tbody>
</table>

*Values are no. (%) except as indicated.
(24.0%), shortness of breath (24.0%), and diarrhea (14.3%) (Table 2). Two persons (4.0%) were asymptomatic throughout the study period. On radiologic examination, 27 (54.0%) had bilateral pneumonia, 11 (22.0%) had unilateral pneumonia, and 17 (34.7%) showed multiple areas of mottling and ground-glass opacity. None of the patients were co-infected with other respiratory viruses or fungi.

Intensive care unit admission was relatively uncommon (3/50, 6.0%) in our cohort compared with admission rates in previous studies (19–22). This difference could be attributed to underdiagnosis of milder cases during the initial COVID-19 outbreak in China. One patient’s sputum specimen was culture-positive for *Klebsiella aerogenes* bacteria; a second patient’s specimen was culture-positive for *Ralstonia pickettii* bacteria. Both had acute respiratory distress syndrome and acute respiratory injury accompanied by septic shock or acute renal injury and required admission to the intensive care unit.

Of the 50 case-patients, 42 (84.0%) could be clustered based on their epidemiologic links (Figure 1, https://wwwnc.cdc.gov/EID/article/27/1/20-1543-F1.htm). We identified 6 transmission clusters (clusters 1–6). Cluster 1 involved a 4-member family. The father, who traveled to Guangdong, China, in late January 2020, was believed to have infected his wife and subsequently their daughter and son-in-law at a family gathering. Clusters 2 and 3 were family clusters of local infection and unknown source. Both clusters involved 3 household members with no recent travel history. Cluster 4 was attributed to a superspreading event (SSE): a barbecue and hotpot party involving 19 family members in late January. Symptom onset in these patients occurred during days 2–13 after the party. A colleague of 1 infected person, who did not attend the party, also tested positive for SARS-CoV-2. Cluster 5 initiated from a resident of a public housing estate, in whom COVID-19 was diagnosed on January 30. Eleven days later, diagnoses were made in 3 members of a household that resided in the same building (10 stories below) the index case-patient. Two household members subsequently attended a family gathering of 29 persons at a Chinese restaurant during the incubation period. COVID-19 was diagnosed consecutively in 3 persons ≈2 weeks after the gathering. In addition, a Filipino domestic aide of 1 infected family member, who did not attend the family gathering, also tested positive. For cluster 6, the first reported case was in a 70-year-old woman who visited a Buddhist worship hall during the Chinese New Year. A further 8 persons who visited the same Buddhist worship hall during this period later tested positive for SARS-CoV-2. At the data cutoff point, ≥4 other household members who had never been to the worship hall also tested positive. Details of the demographic and epidemiologic information on the cases and clusters are provided in Appendix 1 Tables 2–8.

Consensus genomes of all 50 cases were constructed based on nanopore sequencing and refined by Illumina sequencing. On average, 62,387 reads/genome were obtained with 550× coverage for nanopore, and 18,747 reads/genome were obtained with 132× coverage for Illumina platform. The consensus genome size was ≈29.9

<table>
<thead>
<tr>
<th>Table 2. Clinical characteristics and outcomes of 50 coronavirus disease patients, Hong Kong, February 2020*</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Characteristic</strong></td>
</tr>
<tr>
<td><strong>Signs and symptoms</strong></td>
</tr>
<tr>
<td>Fever at admission</td>
</tr>
<tr>
<td>Fever during hospitalization</td>
</tr>
<tr>
<td>Cough</td>
</tr>
<tr>
<td>Sore throat</td>
</tr>
<tr>
<td>Shortness of breath</td>
</tr>
<tr>
<td>Muscle ache</td>
</tr>
<tr>
<td>Diarrhea</td>
</tr>
<tr>
<td>Rhinorrhea</td>
</tr>
<tr>
<td>Nausea and vomiting</td>
</tr>
<tr>
<td>Confusion</td>
</tr>
<tr>
<td>&gt;1 sign or symptom</td>
</tr>
<tr>
<td>Asymptomatic</td>
</tr>
<tr>
<td><strong>Complications§</strong></td>
</tr>
<tr>
<td>Acute respiratory distress syndrome</td>
</tr>
<tr>
<td>Acute respiratory injury</td>
</tr>
<tr>
<td>Acute renal injury</td>
</tr>
<tr>
<td>Septic shock</td>
</tr>
<tr>
<td>&gt;1 complication</td>
</tr>
<tr>
<td>No complications</td>
</tr>
<tr>
<td><strong>Radiological findings</strong></td>
</tr>
<tr>
<td>Unilateral pneumonia</td>
</tr>
<tr>
<td>Bilateral pneumonia</td>
</tr>
<tr>
<td>Multiple areas of mottling and ground-glass opacity</td>
</tr>
<tr>
<td>No abnormality</td>
</tr>
<tr>
<td><strong>Coinfection</strong></td>
</tr>
<tr>
<td>Other viruses</td>
</tr>
<tr>
<td>Bacteria</td>
</tr>
<tr>
<td>Fungi</td>
</tr>
<tr>
<td><strong>Clinical outcome#</strong></td>
</tr>
<tr>
<td>In serious condition, ICU admission</td>
</tr>
<tr>
<td>Hospitalized, in stable condition</td>
</tr>
<tr>
<td>Discharged</td>
</tr>
<tr>
<td><strong>Interval from symptom onset to hospital admission, d</strong></td>
</tr>
<tr>
<td>Average (SD)</td>
</tr>
<tr>
<td>Range</td>
</tr>
<tr>
<td><strong>Length of hospital stay, d††</strong></td>
</tr>
<tr>
<td>Average (SD)</td>
</tr>
<tr>
<td>Range</td>
</tr>
</tbody>
</table>

*Values are no. (%) except as indicated. ICU, intensive care unit.†Data were missing for 2 patients.‡Data were missing for 1 patient.§Definitions of complications are provided in Appendix 2 (https://wwwnc.cdc.gov/EID/article/27/1/20-1543-App2.pdf).¶Moderate growth of *Klebsiella aerogenes* and *Ralstonia pickettii* bacteria were obtained from sputum specimens collected from case 38 and 70.#Data cutoff of the clinical outcome of patients was February 28, 2020.**Data from symptomatic patients were excluded.††Data were missing for 1 patient.
Emerging Infectious Diseases • www.cdc.gov/eid • Vol. 27, No. 1, January 2021

Discussion

This study provides a territorywide overview of the early COVID-19 outbreak in Hong Kong, an international city with borders connecting to mainland China, by integrating demographic, clinical, epidemiologic, phylogenomic, and phylodynamic data. In Hong Kong, most cases recorded in January 2020 were imported cases. After February 1, most were local cases and close contacts of those case-patients, indicating local community transmissions. Transmission in closed settings, especially during family and religious gatherings, is a hallmark of recent cases recorded in Hong Kong. Among 6 clusters identified on the basis of epidemiologic links, 3 (clusters 4–6; Figure 1) were considered SSEs because of the larger number of persons involved (n = 8–13). We performed whole-genome sequencing on all 50 cases to investigate phylogenetic relationship and transmission link.

The SARS-CoV-2 samples in Hong Kong had 99.98% identity to the reference genome (GenBank accession no. NC_045512.2) and demonstrated no apparent major genome modification since the initial COVID-19 outbreak in Wuhan. As shown in global phylogeny, SARS-CoV-2 genomes isolated in Hong Kong could be segregated into 2 lineages. Lineage 1 was phylogenetically related to the strains isolated from China and was the cause of the cases in Cluster 1. Lineage 2 was more closely related to strains from France and South Korea. It also harbored a common mutation at Orf3a-G251V, which accounted for 88.0% of cases in this study.

Regarding the local phylogenetic analysis, clustering of samples was highly concordant to the epidemiologic link, despite the marginally supportive bootstrap value of the nodes because of the limited genetic variability. Cluster 1 demonstrated the closest genetic distance to the reference genome among all cases reported in Hong Kong (Figures 1, 2). The index case of cluster 1 (case 66) was initially defined as possible local infection because the patient traveled to Guangdong Province, which was not considered to have active community transmission at that time. However, our sequencing result demonstrated that the genome of case 66 was 100% identical to the first published SARS-CoV-2 genome, and all cases in cluster 1 did not harbor Orf3a-G251V, which was recognized as a hallmark of local cases with unknown source in our community. Therefore, instead of possible local infections, cluster 1 was more likely imported from mainland China through index case-patient 66.

Cluster 5 originated within a public housing estate, in which a family of 3 members (cases 42, 48, and
Early COVID-19 Outbreak, Hong Kong, China

49) were suspected to have been infected through a confirmed case-patient (case 12) who lived 10 stories above them in the same building, through a potentially faulty sewage pipe setup or other environmental exposure. Based on phylogenetic analysis, viral genomes in cluster 5 shared a similar genetic distance from the reference genome and were assigned to the same branch of the tree. This finding supports a potential transmission link among these cases.

Cluster 4 was a family gathering–associated SSE during Chinese New Year. In concert with epidemiologic information, all 11 cases from cluster 4 shared 3 common missense mutations, namely S-L8V, Orf1ab-H3233Y, and Orf3a-G251V; 7 cases shared identical

Figure 2. Maximum-likelihood phylogenetic tree of 50 coronavirus disease cases, Hong Kong, February 2020. The tree was rooted on the earliest published genome of severe acute respiratory syndrome coronavirus 2 (GenBank accession no. NC_045512.2). Bootstrap value was set at 1,000× and nodes with bootstrap value >50% were shown. Branch lengths were measured in number of substitutions per site. Samples are color-coded by epidemiologic link. Cases 84 and 102 were asymptomatic at the time of sample collection and are marked with asterisks. Each case is identified by case number used by the Centre of Health Protection, Department of Health, Hong Kong, and date of symptom onset. SSE, superspreading event.

Cluster 1: Imported cluster
Cluster 2: Local cluster
Cluster 3: Local cluster
Cluster 4: Hotel–related SSE
Cluster 5: Hotspot party–related SSE
Cluster 6: Public housing estate–related SSE
Cluster 7: Public housing estate–related SSE
Cluster 8: Public housing estate–related SSE
Cluster 9: Public housing estate–related SSE
Cluster 10: Public housing estate–related SSE
Cluster 11: Public housing estate–related SSE
Cluster 12: Public housing estate–related SSE
Cluster 13: Public housing estate–related SSE
Cluster 14: Public housing estate–related SSE
Cluster 15: Public housing estate–related SSE
Cluster 16: Public housing estate–related SSE
Cluster 17: Public housing estate–related SSE
Cluster 18: Public housing estate–related SSE
Cluster 19: Public housing estate–related SSE
Cluster 20: Public housing estate–related SSE
Cluster 21: Public housing estate–related SSE
Cluster 22: Public housing estate–related SSE
Cluster 23: Public housing estate–related SSE
Cluster 24: Public housing estate–related SSE
Cluster 25: Public housing estate–related SSE
Cluster 26: Public housing estate–related SSE
Cluster 27: Public housing estate–related SSE
Cluster 28: Public housing estate–related SSE
Cluster 29: Public housing estate–related SSE
Cluster 30: Public housing estate–related SSE
Cluster 31: Public housing estate–related SSE
Cluster 32: Public housing estate–related SSE
Cluster 33: Public housing estate–related SSE
Cluster 34: Public housing estate–related SSE
Cluster 35: Public housing estate–related SSE
Cluster 36: Public housing estate–related SSE
Cluster 37: Public housing estate–related SSE
Cluster 38: Public housing estate–related SSE
Cluster 39: Public housing estate–related SSE
Cluster 40: Public housing estate–related SSE
Cluster 41: Public housing estate–related SSE
Cluster 42: Public housing estate–related SSE
Cluster 43: Public housing estate–related SSE
Cluster 44: Public housing estate–related SSE
Cluster 45: Public housing estate–related SSE
Cluster 46: Public housing estate–related SSE
Cluster 47: Public housing estate–related SSE
Cluster 48: Public housing estate–related SSE
Cluster 49: Public housing estate–related SSE
Cluster 50: Public housing estate–related SSE
genomes. Considering the fast-evolving property of RNA viruses, the presence of identical genetic sequences among the strains implies that transmission occurred over a short period or even in a single event. Meanwhile, 2 isolated cases (case 23 and case 43) and 3 cases from another local cluster (case 38, case 39, and case 40) shared highly similar genomes to those of cluster 4 (Figures 1, 2). Although no apparent epidemiologic links were observed, the high degree of genomic similarity suggests that these cases might have originated in a single source. That speculation was further supported by the geographic distribution of case-patients who lived near one another and whose social circles might have overlapped (Figure 3). Our results demonstrate that the integration of epidemiologic and phylogenetic data is critical for providing more accurate information about transmission patterns.

Cluster 6 was an SSE occurring in a Buddhist worship hall. Two missense mutations, Orf1ab-G295V and Orf1ab-L3606F, were unique to this cluster. Epidemiologic investigation identified a 43-year-old monk (case 102; Figure 1), who was the abbot of the worship hall and had traveled to mainland China in early January. He was sent to a quarantine center in late February after being linked to a series of confirmed cases connected to the worship hall. He was asymptomatic throughout the study period. Phylogenetic analysis showed that this case was closest to the root of the cluster (Figure 2), suggesting that case 102 could be the index patient of cluster 6. By the time of data cutoff, the cluster involved 13 patients and spread was ongoing. This pattern demonstrates the possibility of a hidden spreader as a source of COVID-19 community outbreak. That likelihood also highlights the importance of rapid quarantine of close contacts of confirmed case-patients, regardless of the presence of symptoms, to halt community spread.

In the evolutionary clock study, the reproduction number of COVID-19 within Hong Kong as of February 28, 2020, was estimated at 1.84 (95% BCI 1.37–2.35). That value strongly indicated that the
outbreak in Hong Kong was ongoing, but it was smaller than the estimated reproduction number of 2.6 in Wuhan (23,24). The smaller value is a combined outcome of reduced growth rate and increased $\delta$. The reduced growth rate is attributed to strong public health awareness among the general public, which resulted in greatly reduced social activities and strong compliance with mask-wearing during this period (25,26). The increased $\delta$ can be attributed to robust laboratory surveillance and fast quarantine time. In addition, time of most recent common ancestor for the cases in Hong Kong was determined to be December 24, 2019, $\approx$25 days before the first patient in our cohort (Case 2) demonstrated symptoms on January 18, 2020 (Figure 1).

Our study has several limitations. Although we included 53.8% of the cases reported in Hong Kong as of February 28, another 43 cases, including 2 fatal cases, were not analyzed in this study. Moreover, incubation periods of cases in which the source of infection is unknown might vary widely. Studies have demonstrated that incubation periods can vary from 4.5 to 15.8 days (24) and can be longer for patients experiencing mild symptoms. However, because patients might already be infectious during the incubation period, the reproductive number in this study could be underestimated. Furthermore, our calculations were based solely on phylodynamic analysis, which could differ from calculations on the basis of epidemiologic models. Finally, ambiguous bases were observed in some of our consensus genomes. This ambiguity is mainly because whole-genome sequencing was performed on respiratory specimens instead of viral culture, in which viral load plays a critical role in the subsequent genome quality as reflected by the cycle threshold of each specimen. The paucity of viral load in specimens could affect the yield of sequencing libraries. In our study, specimens with cycle threshold $<28$ were usually free of ambiguous bases. Nevertheless, the uncovered area only accounted for $\approx$1–3% of the entire viral genome, although the remaining mapped regions had an average coverage of $>100\times$, which should provide sufficient and accurate information for subsequent analyses (Appendix 1).

In conclusion, phylogenomic data were consistent with epidemiologic findings that transmission in closed settings, especially during family and religious gatherings, is a hallmark of COVID-19 outbreak in Hong Kong. Social distancing and vigilant infection control measures, such as rapid isolation of suspected or confirmed case-patients and their close contacts, are crucial for containing COVID-19 in the community.

Acknowledgments

We appreciate Oxford Nanopore Technologies Limited, especially Hai Wang and Eva Yu, for their supportive service in the delivery of sequencing flow cells and reagents, and the provision of technical advices.

This study was supported by the Faculty of Health and Social Science and the Department of Health Technology and Informatics of The Hong Kong Polytechnic University. It was also supported by the Health and Medical Research Fund Commissioned Research on COVID-19 (COVID190204) and the Innovation and Technology Fund Innovation and Technology Support Programme’s Public Sector Trial Scheme (SST/050/20G).

K.S.-S.L. and G.K.-H.S. designed the study, collected, analyzed, and interpreted the data, conducted literature search, and drafted and critically reviewed the manuscript; and G.K.-H.S. was also responsible for securing funding for this study. T.T.-L.N., H.-Y.L., M.-P.C., K.K.-G.T., and L.-K.L. conducted experiments and analyzed the data. A.K.-L.W., M.C.-Y.Y., B.K.-C.W., A.Y.-M.H., K.-T.Y., K.-C.L., R.W.-T.L., E.Y.-K.T., W.-S.L., M.-C.C., Y.-Y.N., K.-M.S., K.S.-C.F., S.K.-Y.C., W.-K.T., and T.-L.Q. were responsible for collecting and analyzing the data and finalizing the manuscript. D.H.-K.S. and S.P.Y. analyzed the data and wrote and critically reviewed the manuscript. W.C.Y. was involved in data analysis, data interpretation, and writing of the manuscript. All authors reviewed and approved the final version of the manuscript.

About the Author

Dr. Leung obtained his PhD in Microbiology at the University of Hong Kong in 2019 and received bioinformatic and phylogenetic analysis training under HKU Pasteur institute. His research interests are molecular diagnosis and epidemiology of emerging infectious disease such as Mycobacterium tuberculosis and HIV.

References


Address for correspondence: Gilman Siu, Room Y928, 9/F, Block Y, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong; email: gilman.siu@polyu.edu.hk
Supplemental Methods

Case Definitions

Imported cases referred to case-patients who had records of traveling to Wuhan City or Hubei Province, China ≤14 days before symptom onset. Possible local cases referred to case-patients who had records of traveling to other regions of China or other countries where active community transmission was not confirmed at the time of the respective case. Local cases referred to case-patients who did not have a travel history ≤14 days before symptom onset.

Amplification of SARS-CoV-2 Genome using Multiplex PCR

The viral cDNA was amplified by using 2 PCRs containing tiled, multiplexed primers as described in the ARTIC network (1). These primers (n = 196) generated overlapping 400-bp amplicons across the entire genome of COVID-19 (accession no. NC_045512). Each PCR contained a primer pool of 98 primers (Appendix 1 Table 1, https://wwwnc.cdc.gov/EID/article/27/1/20-1543-App1.xlsx). Eventually, the PCR mastermix (50µL) included 2·5µL of cDNA, 5·0µL of 5X Q5 Reaction Buffer, 0·5µL of 10mM dNTP mix, 0·25µL of Q5 Hot Start DNA Polymerase (New England Biolabs, https://www.neb.com/) and 3·6µL of 10µM primer pool 1 or 2, and 13·15µL of nuclease-free water. The mixtures were incubated at 98°C for 30 s, followed by 35 cycles at 98°C for 15 s and 65°C for 5 min. The PCR amplicons were then purified by 1X Agencourt AMPure XP beads (Beckman Coulter, https://www.beckmancoulter.com/).

Bioinformatic analysis of Nanopore Sequencing Data

Nanopore sequencing data were analyzed using a protocol modified from Artic Network nCoV-2019 novel coronavirus bioinformatics protocol (2). In brief, sequencing results obtained from Nanopore sequencing were first based-called using guppy-basecaller (v3.4.5). The
sequencing reads were then demultiplexed by Porechop (v0.2.4) to correct the significant barcoding misassignments and mapped against the SARS-CoV-2 reference genome Wuhan-Hu-1 (accession no. NC_045512) using BWA (v0.7.17) and Samtools (v1.7). The BAM file was used for variant calling by Medaka (v0.11.5) with threshold value set at 1 to ensure haploid decoding. To evaluate the performance of variant calling by medaka, variants were validated again using Nanopolish (v0.11.0) with haploid decoding, and candidate variants from the aligned reads were extracted when the variant frequency was >40%. The consensus genome was assembled based on the VCF from Medaka (v0.11.5) and BAM file. Depth per base required for variant calling was set at >20X with minimum required base Phred score of 10.

**Modified Artic Network nCoV-2019 novel coronavirus bioinformatic protocol**

```
porechop -i [$fastq_file] -o [$output_file] -t [$thread no.]

source activate artic-ncov2019


source ~/.bashrc

source activate medaka #medaka v0.11.5

medaka consensus [$filename.primertrimmed.sorted.bam] [$filename.hdf]


source ~/.bashrc

source activate artic-ncov2019


[$filename.primertrimmed.medaka.vcf] [$filename.primertrimmed.sorted.bam] >[$filename.consensus.fasta]
```
Appendix 2 Figure 1. Maximum-likelihood phylogenetic tree constructed based on 50 severe acute respiratory syndrome coronavirus 2 genomes in Hong Kong, February 2020, and genomes collected from GISAID databank. A total of 478 worldwide genomes as of February 28, 2020 were available from GISAID Global Cases COVID-19 database. All genomes were downloaded and the phylogenetic tree was built by using Maximum Likelihood with bootstrap value set and rooted on the earliest published genome of severe acute respiratory syndrome coronavirus 2 (GenBank accession no. NC_045512.2). Bootstrap value was set at 500$\times$ and nodes with bootstrap value $>50\%$ were shown. Branch lengths were measured in number of substitutions/site. All Hong Kong strains are highlighted in red. Hong Kong strains showed limited genetic variability and tended to aggregate in two lineages, highlighted in green and cyan.

References:

   https://artic.network/ncov-2019

Appendix 2 Figure 2. Bayesian maximum clade creditability time-scaled phylogeny of the early COVID-19 outbreak in Hong Kong. Bayesian time-scaled phylogeny plotted using 50 COVID-19 genomes collected in this study. The analysis was undertaken using GTR+G+I model, coalescent exponential population, strict clock setting, sampling 100,000 trees from 1 billion generations. Uncertainty for the date of each node (95% highest posterior density intervals) is displayed in blue bars. Nodes with posterior higher than 70% were displayed.