medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

A Territory-wide study of COVID-19 cases and clusters with unknown source in Hong Kong community: A clinical, epidemiological and phylogenomic investigation

Kenneth Siu-Sing LEUNG, Timothy Ting-Leung NG, Alan Ka-Lun WU, Miranda Chong-Yee YAU, Hiu-Yin LAO, Ming-Pan CHOI, Kingsley King-Gee TAM, Lam-Kwong LEE, Barry Kin-Chung WONG, Alex Yat-Man HO, Kam-Tong YIP, Kwok-Cheung LUNG, Raymond Wai-To LIU, Eugene Yuk-Keung TSO, Wai-Shing LEUNG, Man-Chun CHAN, Yuk-Yung NG, Kit-Man SIN, Kitty Sau-Chun FUNG, Sandy Ka-Yee CHAU, Wing-Kin TO, Tak-Lun QUE, David Ho-Keung SHUM, Shea Ping YIP, Wing Cheong YAM, Gilman Kit-Hang SIU*

Department of Health Technology and Informatics, Faculty of Health and Social Science, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region, China (T T-L NG MSc, H-Y LAO BSc, M-P CHOI MSc, L-K LEE BSc, S P YIP PhD, G K-H SIU PhD)

Department of Microbiology, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong Special Administrative Region, China (K S-S LEUNG PhD, K K-G TAM PhD, W-C YAM PhD)

Department of Clinical Pathology, Pamela Youde Nethersole Eastern Hospital, Hong Kong Special Administrative Region, China (A K-L WU MD, M C-Y YAU DHSc)

Department of Medicine, Pamela Youde Nethersole Eastern Hospital, Hong Kong Special Administrative Region, China (K-C LUNG MD)

Department of Medicine & Geriatrics, Ruttonjee Hospital, Hong Kong Special Administrative Region, China (R W-T LIU MD)

Department of Pathology, United Christian Hospital, Hong Kong Special Administrative Region, China (B K-C WONG PhD, K S-C FUNG MD, S K-Y CHAU MD)

Department of Medicine & Geriatrics, United Christian Hospital, Hong Kong Special Administrative Region, China (E Y-K TSO MD)

Department of Pathology, Princess Margaret Hospital, Hong Kong Special Administrative Region, China (A Y-M HO PhD, W-K TO MD)

Department of Medicine, Princess Margaret Hospital, Hong Kong Special Administrative Region, China (W-S LEUNG MD, M-C CHAN MD)

Department of Clinical Pathology, Tuen Mun Hospital, Hong Kong Special Administrative Region, China (K-T YIP MPhil, T-L QUE MD)

Department of Medicine and Geriatric, Tuen Mun Hospital, Hong Kong Special Administrative Region, China (K-M SIN MD, Y-Y NG MD)

Faculty of Health and Social Science, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region, China (D H-K SHUM PhD)

*** Correspondence to: Dr. Gilman KH SIU, Department of Health Technology and Informatics, Hong Kong Polytechnic University, Hong Kong (gilman.siu@polyu.edu.hk)**

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Summary

Background: Initial cases of coronavirus disease 2019 (COVID-19) reported in Hong Kong were mostly imported cases from Mainland China. However, most cases reported in February 2020 were local infections with unknown source, indicating local community transmissions. This study aimed to report the clinical, epidemiological and phylogenomic characteristics of the local cases of COVID-19 in our community.

Methods: We extracted the demographic, clinical and epidemiological data from 50 COVID-19 patients, who accounted for 53·8% of the cases in Hong Kong by the end of February 2020. We used both Nanopore and Illumina platforms to perform whole-genome sequencing (WGS) of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from these patients. Phylogenetic relatedness among these local cases and their placement in the global phylogeny were examined. The evolutionary rate and divergence time of transmission were also determined.

Findings: Of these 50 patients, only three (6·0%) had visited Wuhan while 43 (86·0%) did not have recent travel records. The average interval from symptom onset to hospital admission was 8·5 days. The most common signs and symptoms on admission were cough (74·0%) and fever (58·0%). Radiographic abnormality was found in 46 (92·0%) patients. Three (6·0%) patients required ICU admission. Phylogenetic analysis concurred with epidemiological investigation that 42 (84·0%) cases could be grouped into six transmission clusters. Forty-four (88·0%) cases harboured a common mutation *Orf3a* G251V. Global phylogeny of SARS-CoV-2 revealed that most (88·0%) cases in Hong Kong were clustered in two subclades with the strains from other countries. The estimated time to the most recent common ancestor (tMRCA) of COVID-2019 outbreak in Hong Kong was December 24,

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

2019 with an evolutionary rate of 3.04×10^{-3} substitutions per site per year. The reproduction number value was 1·84 as of February 28, 2020 in Hong Kong.

Interpretation: We provided a territory-wide overview of COVID-19 in Hong Kong, which has borders connecting to Mainland China. Transmission in closed settings especially during family and religious gatherings is a hallmark of the recently reported cases. The reproduction number value indicated an ongoing outbreak in the community. Rapid isolation of suspected or confirmed cases and their close contacts coupled with robust laboratory surveillance and social distancing is crucial to the containment of COVID-19 transmission.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Introduction

"Coronavirus disease 2019 (COVID-19)" refers to the cluster of viral pneumonia cases initially from Wuhan, Hubei Province, China since December 2019. The aetiology was unknown during the early stage of the outbreak until Chinese scientists isolated a novel coronavirus, SARS-CoV-2, from a patient on January 7, 2020 and performed genome sequencing.[1]

Fever was the main symptom of COVID-19 with about one-third of the patients presenting with acute respiratory distress syndrome (ARDS). About 16% of the patients were considered to be in severe condition on admission, and the estimated mortality rate was 1·4%.[2] Sustained human-to-human transmission was confirmed upon identification of cases clustering among families as well as transmission from patients to healthcare workers,[3,4] which triggered China's urgent public health actions and international concern.

The World Health Organization (WHO) declared COVID-19 a global health emergency in late January 2020. As of February 28, 2020, 78 824 patients have been diagnosed to have COVID-19 in Mainland China with 2 788 deaths. The disease also spread to 50 other countries across the world.[5] The first case in Hong Kong was diagnosed on January 23, 2020. The patient was a resident of Mainland China who travelled to Hong Kong from Wuhan via Shenzhen by High-Speed Rail, and was thus considered an imported case. 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, there were 93 COVID-19 cases in Hong Kong with at least 70 (75·3%) cases being local cases and their close contacts.[6,7] Secondary and tertiary transmissions were observed in some clusters of cases. As most index cases of these clusters have unknown sources, a hidden transmission chain was believed to have established in the community.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Here, we report the demographic, clinical and epidemiological data of 50 hospitalised patients, who accounted for 53·8% of COVID-19 cases in Hong Kong at the data cut-off point, including three imported cases and six transmission clusters of local infections contracted through household contacts, family gatherings, and religious assemblies. With the combined use of Nanopore sequencing and Illumina sequencing, the viral genomes of all these cases were characterised. Phylogenetic and phylodynamic analyses were performed to determine the transmission linkage and the evolutionary rate of COVID-19 cases in our community.

Methods

Cases

For this retrospective, multi-centre study, we intentionally enrolled the positive cases of COVID-19 that were laboratory-confirmed at four 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, from January 26 to February 28, 2020.

Sputum specimens and throat swabs pooled with nasopharyngeal aspirates were collected from patients who fulfilled the reporting criteria or enhanced surveillance criteria on admission.[8] A patient was considered to have laboratory-confirmed infection when SARS-CoV-2 was detected by real-time reverse transcription polymerase chain reactions (RT-PCRs), which amplified the envelope (*E*) gene and RNA-dependent RNA polymerase (*RdRp*) gene.[9]

The demographic, clinical, and microbiological data were obtained from patients' medical records. The epidemiological information was retrieved from the Centre for Health Protection (CHP) of the Department of Health[6] and the website wars.vote4.hk - Coronavirus in HK.[7]

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

The definitions of clinical symptoms and complications are based on the WHO guidance.[10] We adopted the case numbering system of CHP, which was based on the date of case confirmation.

This study was approved by the Institutional Review Boards of The Hong Kong Polytechnic University (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*g* for 2 minutes to deplete host cells. Supernatant was used for total nucleic acid extraction using either MagNA Pure 96 System (Roche Diagnostics, Germany) or NucliSENS® easyMAG® (bioMérieux, The Netherlands) according to manufacturers' instructions. DNase treatment was done by TURBO DNA-free Kit (ThermoFisher Scientific, USA) to remove the residual host DNA.

Reverse Transcription and viral genome amplification using multiplex PCR

DNase-treated RNA was reverse-transcribed to single-strand cDNA using random hexamers and SuperScript IV reverse transcriptase (Invitrogen, USA) as previously described.[11] The viral cDNA was then amplified using two PCRs containing tiled, multiplexed primers (table S1) described in the ARTIC network.[12] Details of the multiplex PCR are provided in the Supplementary Appendix.

Nanopore MinION Sequencing

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Ligation-based 1D sequencing (SQK-LSK109, ONT, UK) was carried out according to the manufacturer's instructions. Briefly, multiplex PCR amplicons of each sample were normalised to 1 ng/µL prior to end-repairing and native barcode ligation (EXP-NBD104/114, ONT, UK). Barcoded samples were pooled and ligated to AMII sequencing adaptor. Sequencing was performed with Nanopore MinION device using R9.4.1 flow cell for 48 hours.

Illumina MiSeq sequencing

Multiplex PCR amplicons were subjected to library preparation and dual-indexing using KAPA HyperPrep Kit and Unique Dual-Indexed Adapter Kit (Roche Applied Science, US) according to the manufacturer's instructions. Ligated libraries were then enriched by 6-cycle PCR amplification followed by purification and size selection using AMPure XP beads (Beckman Coulter, USA). The pooled library was sequenced with MiSeq Reagent Kit V2 Nano on Illumina MiSeq System.

Bioinformatic analysis

Nanopore sequencing data were analysed using a modified Artic Network nCoV-2019 novel coronavirus bioinformatics protocol (Supplementary Appendix).[13]

Illumina sequencing reads were mapped with reference to the respective draft genome of each sample constructed from the Nanopore data. Variants were called using freebayes (v1.0.0) with haploid decoding and minimum base quality set at Q30. Consensus genomes were constructed by GATK (v4.1.4.1) based on the VCF file[14]. SPAdes genome assembler (v3.14.0) and minimap2 (v2.17) were used to combine Nanopore and Illumina sequencing

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

results for *de novo* assembly and identify the sequence of the unmapped gap regions. The sequences have been submitted to the GenBank with the accession no. MT232662-MT232711.

Genomic and Phylogenetic analysis

To identify the amino acid change caused by each single-nucleotide polymorphism (SNP), the consensus genome of each specimen was BLASTed against the reference NC_045512.2 using BlastX. The non-synonymous mutations were then identified using custom Python script.

Phylogenetic tree was constructed based on the consensus genomes from the specimens with PhyML (v3.0) using maximum likelihood algorithm. Best-fitting substitution model for phylogenetic analysis was performed by Akaike information criteria, in which the general time reversible model with fixed proportion of invariable sites (+I) was selected.[16] Bootstrap replicates was set at 1000×, and maximum-likelihood phylogenetic tree was rooted on the earliest published genome of SARS-CoV-2 (accession no.: NC_045512.2). Transmission cluster was defined by clear epidemiological and onset-time relationship. Meanwhile, an additional 273 SARS-CoV-2 genomes were downloaded from GISAID Severe acute respiratory syndrome coronavirus 2 data hub.[15] Together with the genomes from this study, phylogenetic tree was constructed using fast likelihood-based aLRT SH-like method and rooted on SARS-CoV-2 genome NC_045512.2.

Estimation of evolutionary rate and divergence time of transmission

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

To reconstruct the evolutionary model of COVID-19 cases in Hong Kong, Bayesian interference through Markov Chain Monte Carlo (MCMC) framework was implemented in BEAST (v2.6.2).[17] Bayesian phylodynamic analysis was performed using both strict clock and relaxed clock model with coalescent exponential growth tree priors. MCMC chains were run for 1×10^9 generations and sampled every 500 steps. Bayesian output, including MCMC convergence, growth rate, tMRCA, and effective sample sizes, was analysed after the results were visualised by Tracer (v1.7.1).[18] All parameters had an effective sample size of >200, which indicated sufficient sampling.

Results

Fifty patients with COVID-19 were included in this study with 54·0% being female and with mean age of 55·2 years (table 1). Three cases were considered as imported cases as the patients had stayed in Wuhan before travelling to Hong Kong in mid-January. Four had travelling record to other regions where COVID-19 active community transmission was not confirmed by that time. These cases were considered as possibly local infection. The other 43 were defined as local infection with no recent travel history. Eighteen (36·0%) had chronic illnesses, of which cardiovascular and cerebrovascular diseases were the most common conditions (table 1). In total, 74·0% of the patients presented with cough. Only 58·0% of the patients on admission presented with fever, but fever gradually developed in 64·0% patients during hospitalisation. Other symptoms were less common, such as muscle ache (25·0%), sore throat (24·0%), shortness of breath (24·0%), and diarrhoea (14·3%) (table 2). Two (4·0%) patients were asymptomatic throughout the study period. According to radiological examination, 27 (54·0%) showed bilateral pneumonia, 11 (22·0%) unilateral pneumonia, and

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

17 (34·7%) multiple mottling and ground-glass opacity. None of the patients was co-infected with other respiratory viruses or fungi. Two patients were culture-positive for *Klebsiella aerogenes* and *Ralstonia pickettii* in their sputum specimens. Both presented with ARDS and acute respiratory injury accompanied with septic shock or acute renal injury, and required ICU admission.

Of the 50 cases, 42 (84·0%) could be clustered based on their epidemiological linkages (figure 1). Six transmission clusters (Clusters 1–6) were identified. Cluster 1 involved a family of four members. The father, who travelled to Guangdong Province, China in late January 2020, was believed to infect his wife and subsequently their daughter and son-in-law in a family gathering. Clusters 2 and 3 were family clusters of local infection with unknown source. Both clusters involved three household members who had no recent travel history. Cluster 4 was a super-spreading event (SSE) associated with a party where a family of 19 members had barbecue and hotpot in late January. Ten of them had developed symptoms two days after the party. A colleague of one infected member, who was not present at the party, was also subsequently diagnosed to have COVID-19. Cluster 5 initiated from a resident of a public estate, who was diagnosed on January 30. Eleven days later, three members of a household, who lived in the same building as, but 10-storey below from, the index case, were also confirmed to be infected. Two of the household members attended a family gathering of 29 people at a Chinese seafood restaurant during their incubation period. Three of them were diagnosed consecutively around two weeks after the gathering. Additionally, a Filipino domestic helper of one infected member, who was absent in the family gathering, was tested positive for SARS-CoV-2. The earliest reported case of Cluster 6 was a woman who visited a Buddhist worship hall during the Chinese New Year holidays. In addition, eight individuals who had visited the same Buddhist worship hall during this period were later diagnosed. At

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

the data cut-off point for this study (February 28, 2020), at least four other household members who had never been to the worship hall were also tested positive for SARS-CoV-2.

The consensus genomes of all 50 cases were first constructed based on Nanopore sequencing and further refined by Illumina sequencing. On average, 62 387 and 18 747 reads were obtained per genome sequencing at a depth of 550× and 132× for Nanopore and Illumina platforms respectively. The size of the consensus genomes was ~29.9 kbp with a GC content of 38%. The genomes were highly conserved in comparison with the first COVID19 genome with an average sequence identity of 99·98% (range: 99·94%–100·0%) and 0–11 SNPs identified for each strain. A total of 64 nonsynonymous substitutions were identified from these 50 genomes (table 3). Of them, G251V in *Orf3a* is the most frequent amino acid substitution with 44/50 (88·0%) of SARS-CoV-2 strains harbouring this mutation, followed by H3233Y (30/50; 60·0%) in *Orf1ab* and L8V (27/50; 54·0%) in the *S* gene.

The genomic-wide SNPs were used to contextualise the phylogenetic placement of Hong Kong strains in the global phylogeny of SARS-CoV-2. In comparison with SARS-CoV-2 strains isolated from other regions, Hong Kong strains showed limited genomic variability, and tended to aggregate in two clusters (figure S1). In the first cluster, a total of 13 Hong Kong strains formed a subclade (yellow rectangular box) with the strains from France (n=5), Singapore (n=4), Australia (n=2), South Korea (n=2), Sweden (n=1) and USA (n=1). The second cluster composed of 31 Hong Kong strains, which formed another distinct subclade (pink rectangular box) with the strains from Japan (n=3), China (n=2), England (n=2), Australia (n=2) and New Zealand (n=1).

Regarding the phylogenetic relationship among the Hong Kong cases, the branching and the clustering of samples were consistent with the actual epidemiological linkage (figure 2). Four

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

distinctive nodes were identified. The first node belonged to two imported cases and the cases in Cluster 1 of which the genomes had very close genetic distance (0–2 SNPs) with the reference genome. The second node represented cases in Cluster 5, of which four patients lived in different apartments of the same building in a public housing estate, and were believed to have transmitted the virus to other members in a subsequent family gathering. The third node was associated with the cases in Cluster 4 believed to originate from a family hotpot gathering. Notably, the genomes of Cluster 4 were highly similar to those in Cluster 3 and two singleton cases (i.e. Case 23 and Case 43), which shared the same missense mutations at L8V in the *S* gene, H3233Y in *Orf1ab* and G251V in *Orf3a*. Finally, the fourth node belonged to the SSE occurring in Buddhist worship hall (Cluster 6), in which *Orf1ab* G295V and *Orf3a* G251V were identified in all the strains in the cluster.

According to Bayesian time-scaled phylodynamic analysis, the estimated tMRCA of COVID-2019 outbreak in Hong Kong was postulated to be December 24, 2019 (95% Bayesian credible interval [BCI]: December 11, 2019 to January 5, 2020) with an evolutionary rate 3.04×10^{-3} substitutions per site per year (95% BCI: $2{\cdot}04\times10^{-3}$ to $4{\cdot}09\times10^{-3}$ substitutions per site per year). Based on the demographic data, the time interval from symptom onset to hospital admission was ~8·5 days. The estimated reproduction number was 1·84 (95% BCI: 1·37 to 2·35).

Discussion

This study provides a territory-wide overview of COVID-19 cases and clusters with unknown source in Hong Kong, an international city with borders connecting to Mainland China, by integrating the demographic, clinical, epidemiological, phylogenomic, and phylodynamic data.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Fever was reported to be the most common symptom as 81·8%–98·0% of the patients had fever on admission during the initial phase of the outbreak in China.[19-22] However, a large-scale study by Guan *et al* found that fever was presented in only 43·8% of patients on admission[2]. In the present study, fever was identified in only 58·0% of the patients on admission, but developed in 64.0% after hospitalisation. Afebrile patients might be missed if the surveillance case definition relied solely on fever. Therefore, laboratory surveillance should be extended to in-patients and out-patients with respiratory symptoms. Admission to ICU was relatively uncommon (6·0%) in our cohort when compared with previous studies.[19-21] Two of the three patients presented with severe complications including ARDS, acute renal damage, and septic shock. Of note, both these patients were found to have bacterial co-infection, which was absent in other cases. In addition to the virulence factors of the pathogens, the host's immune status, old age, and presence of chronic medical illness might be associated with enhanced severity of the disease. Prompt administration of antibiotics to prevent infection and strengthening of immune supportive treatment might reduce complications and mortality.

In Hong Kong, initial cases recorded in January 2020 were mostly imported cases. Since February 1, the majority were local cases and their close contacts, indicating local community transmissions. Transmission in closed settings especially during family and religious gatherings is a hallmark of the recent cases recorded in Hong Kong. Among six clusters identified based on epidemiological linkage, three (Clusters 4–6; figure 1) were considered SSEs because they involved more individuals (n=8–13). To investigate the phylogenetic relatedness and the transmission linkage among these clusters and other singleton cases, WGS was performed on all 50 cases.

The SARS-CoV-2 strains isolated in Hong Kong have 99·98% identity to the reference genome (NC_045512.2), indicating that the virus has not undergone major genome

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

modification since the initial outbreak of COVID-19 in Wuhan. Forty-four (88·0%) of our strains shared a common mutation *Orf3a* G251V. Except one imported case, all belonged to local cases and their close contacts without travel records (figure 1). Unlike other regions, such as China and Europe, where the SARS-CoV-2 genomes scattered across different branches in global phylogeny, the majority (88·0%) of Hong Kong strains were clustered together in two subclades (figure S1). This suggested that the COVID-19 outbreak in Hong Kong community was mostly arisen from two ancestors.

Regarding the local phylogenetic analysis, although the bootstrap value is only marginally supportive because of the highly conserved genomic sequences, the clustering of samples was highly concordant to the epidemiological linkage. Cluster 1 demonstrated the closest genetic distance to the reference genome amongst all cases reported in Hong Kong (figures 1 and 2). The index case of Cluster 1 (Case 66) was previously defined as possibly local infection as the patient travelled to Guangdong province, which was not considered to have active community transmission by that time. However, our sequencing result showed that the genome of Case 66 was 100% identical to the Wuhan reference genome, and all cases in Cluster 1 did not harbour *Orf3a* G251V, which was recognized as a hallmark of the local cases with unknown source in our community. Therefore, instead of possibly local infections, Cluster 1 was more likely imported from Mainland China via the index case 66.

Cluster 5 originated from a public estate, in which a family of three members (Cases 42, 48, and 49) was speculated to get infected from a confirmed case (Case 12) who lived in the same building, but 10-storey above, through a faulty sewage pipe setup. Based on the phylogenetic analysis, the viral genomes of these cases and other patients in Cluster 5 shared a similar genetic distance from the reference genome and were assigned to the same branch of the tree. This supports a potential transmission linkage among these cases.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Cluster 4 was an SSE associated a family gathering during the Chinese New Year. In concert with epidemiological information, all 11 cases from Cluster 4 shared three common missense mutation, namely L8V in the *S* gene, H3233Y in *Orf1ab* and G251V in *Orf3a*. Seven of them even had identical genomes. Considering the fast-evolving property of RNA viruses, the identical genetic sequences among the strains implied that the transmission occurred in a short period or even in a single event. Meanwhile, two singleton cases (Case 23 and Case 43) as well as cases from another local cluster (Case 38, Case 39, and Case 40 of Cluster 3) shared highly similar genomes to those of patients in Cluster 4 (figures 1 and 2). While there was no apparent epidemiological linkage between these clusters, the high degree of genome similarity suggests that these cases might be originated from a single source. The speculation was further supported by the geographical distribution of residences of these patients that they lived in close proximity to each other, and might share overlapping living circles (figure S2).

Cluster 6 was an SSE occurring in a Buddhist worship hall. Two missense mutations G295V and L3606F in *Orf1ab* were unique to this cluster. Epidemiological investigation identified a monk (Case 102; pink rectangular box of Cluster 6, figure 1), who was the abbot of the worship hall and had travelled to Mainland China in early January. SARS-CoV-2 was isolated from his respiratory specimen when he was sent to a quarantine centre in late February after being found associated with a series of confirmed cases connected with the worship hall. He was completely asymptomatic throughout the study period. Phylogenetic analysis showed that this case was closest to the root of the cluster (top portion with strains shown in red, figure 2), suggesting that Case 102 could be the index patient of the cluster. By the time of data cut-off, the cluster has already involved 13 patients and the spread was still ongoing. Here we demonstrated the possibility that “hidden spreader” could be a source of community outbreak of COVID-19. It also highlights the importance of rapid quarantine of

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

the close contacts of confirmed cases regardless of the presence of signs and symptoms in order to halt the spreading of COVID-19 in the community.

In the evolutionary clock study, the death rate δ (which refers to the duration for the case to become non-contagious) was determined as the lag time between symptom onset date and the hospital admission date (i.e. 8·5 days, which is equivalent to 366/8.5=45·18/year). While δ was normally calculated based on recovery date,[23] we used the admission date for calculation because the transmission link in Hong Kong was practically stopped once the patient was hospitalised. Assuming δ=45·18/year in our calculation, the reproduction number within Hong Kong up to February 28, 2020 was estimated to be 1·84 (95% BCI: 1·37 to 2·35). The value strongly indicated that the outbreak in Hong Kong is still ongoing, but was smaller than the estimated reproduction number of 2·6 in Wuhan as of February 4, 2020.[24] The smaller value is a combined outcome of reduced growth rate and increased δ. The former one is attributed to very strong public health awareness among Hong Kong people who had greatly reduced their social activities and always put on surgical masks during this period,[25,26] whereas the latter is the result of robust laboratory surveillance and fast quarantine time in response to COVID-19 outbreak. In addition, tMRCA for the cases in Hong Kong was determined to be December 24, 2019, which is ~25 days since the first symptom onset case (Case 2–the case with the earliest symptom onset, January 18, 2020; figure 1) of our study cohort.

There are several limitations in our study. First, despite we have included 53·8% of the cases reported in Hong Kong up to February 28, another 43 cases including two fatal cases were not analysed in this study. Second, as most cases in this study were of unknown sources, the incubation period of these cases might vary widely. Studies have shown that the incubation period can vary from 4·5 to 15·8 days,[24] and can be even longer for patients presenting with mild symptoms. However, as the patient might already be infectious during the incubation

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

period, the resultant reproductive number in this study could still be underestimated. Moreover, the current calculation was solely based on phylodynamic analysis and the value could be different from those based on epidemiological models. Finally, gap regions were observed in some consensus genomes. This is mainly because WGS was performed on respiratory specimens instead of viral culture. The variable viral load in specimens could affect the yield of sequencing libraries. Nevertheless, the uncovered area only accounted for approximately 1% of the entire viral genome while the remaining mapped regions had an average coverage of >100X, which should provide sufficient and accurate information for our subsequent phylogenetic and phylodynamic analyses.

In conclusion, phylogenetic and phylodynamic data were consistent with the epidemiological finding that an active community outbreak is occurring in Hong Kong although the clinical presentations of the confirmed cases were generally mild. Vigilant infection control measures, such as rapid isolation of suspected or confirmed cases and their close contacts, and social distancing should be maintained to contain further transmission of COVID-19 in the community.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

References

1. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N Engl J Med* 2020; **382**: 727–33.
2. Guan WJ, Ni ZY, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* 2020; published online Feb 28, 2020. DOI: 10.1056/NEJMoa2002032.
3. Chan JF, Yuan S, Kok KH, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. *Lancet* 2020; **395**: 514–23.
4. Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus-Infected Pneumonia. *N Engl J Med* 2020; published online Jan 29, 2020. DOI: 10.1056/NEJMoa2001316.
5. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect Dis* 2020; published online Feb 19, 2020. DOI: 10.1016/S1473-3099(20)30120-1.
6. Centre_for_Health_Protection. Latest local situation of Severe Respiratory Disease associated with a Novel Infectious Agent. 2020. https://www.chp.gov.hk/files/pdf/enhanced_sur_pneumonia_wuhan_eng.pdf (accessed Feb 24, 2020).
7. Leung YF, Wong HB, Shing Y, Wong HW, Wong WK. wars.vote4.hk - Coronavirus in HK. 2020. https://wars.vote4.hk/en/ (accessed Feb 25, 2020).
8. Centre_for_Health_Protection. Severe respiratory disease associated with a novel infectious agent—letters to doctors. 2020. https://www.chinadaily.com.cn/a/202001/25/WS5e2bb1b6a31012821727333a.html (accessed Feb 6, 2020).
9. Corman VM, Landt O, Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Euro Surveill* 2020; **25**. DOI: 10.2807/1560-7917.ES.2020.25.3.2000045..
10. WHO. Clinical management of severe acute respiratory infection when Novel coronavirus (nCoV) infection is suspected: interim guidance. 2020. https://www.who.int/internal-publications-detail/clinical-management-of-severe-acute-respiratory-infection-when-novel-coronavirus-(ncov)-infection-is-suspected (accessed Jan 11, 2020).
11. Peiris JS, Lai ST, Poon LL, et al. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* 2003; **361**: 1319–25.
12. Quick J. Artic Network-nCoV 2019 sequencing protocol. 2020. https://artic.network/ncov-2019 (accessed Feb 24, 2020).
13. Loman N, Rambaut A. nCoV-2019 novel coronavirus bioinformatics protocol. 2020. https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html (accessed Mar 10, 2020).
14. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; **43**(5): 491-8.
15. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* 2017; **1**: 33–46.
16. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* 2010; **59**: 307–21.
17. Bouckaert R, Heled J, Kuhnert D, et al. BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* 2014; **10**: e1003537.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

18. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol* 2018; **67**: 901–4.
19. Chen N, Zhou M, Dong X, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020; **395**: 507–13.
20. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* 2020; **395**: 497–506.
21. Wang D, Hu B, Hu C, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus-Infected Pneumonia in Wuhan, China. *JAMA* 2020; published online Feb 7, 2020. DOI: 10.1001/jama.2020.1585.
22. Wu J, Liu J, Zhao X, et al. Clinical Characteristics of Imported Cases of COVID-19 in Jiangsu Province: A Multicenter Descriptive Study. *Clin Infect Dis* 2020; published online Feb 29, 2020. DOI: 10.1093/cid/ciaa199.
23. Boskova V, Stadler T, Magnus C. The influence of phylodynamic model specifications on parameter estimates of the Zika virus epidemic. *Virus Evol* 2018; **4**: vex044.
24. Lai A, Bergna A, Acciarri C, Galli M, Zehender G. Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. *J Med Virol* 2020; published online Feb 25, 2020. DOI: 10.1002/jmv.25723.
25. Sung AD, Sung JAM, Thomas S, et al. Universal Mask Usage for Reduction of Respiratory Viral Infections After Stem Cell Transplant: A Prospective Trial. *Clin Infect Dis* 2016; **63**: 999–1006.
26. Seto WH, Tsang D, Yung RW, et al. Effectiveness of precautions against droplets and contact in prevention of nosocomial transmission of severe acute respiratory syndrome (SARS). *Lancet* 2003; **361**: 1519–20.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

	Patients (n=50)
Age (yo)	
Mean (SD)	55·2 (19·5)
Range	22–96
≤30	8 (16·0%)
31-40	5 (10·0%)
41-50	6 (12·0%)
51-60	11 (22·0%)
61-70	10 (20·0%)
≥71	10 (20·0%)
Gender	
Female	27 (54·0%)
Male	23 (46·0%)
Travel record within 14 days before symptom onset	7 (14·0%)
Wuhan, Hubei Province, China	3 (6·0%)
Other regions in Mainland China	1 (2·0%)
Regions outside Mainland China	3 (6·0%)
No travel record	43 (86·0%)
Chronic medical illnesses	18 (36·0%)
Cardiovascular and cerebrovascular diseases	14 (28·0%)
Endocrine system diseases	11 (22·0%)
Nervous system diseases	5 (10·0%)
Digestive system diseases	4 (8·0%)
Malignant tumour	1 (2·0%)

Table 1: Demographics, travel record and baseline medical history of 50 patients recruited in this study

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

	Patients (n=50)
Signs and symptoms	
Fever (on admission)	29 (58·0%)
Fever (during hospitalisation)	32 (64·0%)
Cough	37 (74·0%)
Sore throat	12 (24·0%)
Shortness of breath	12 (24·0%)
Muscle ache	12 (25·0%)[e]
Diarrhoea	7 (14·3%)[f]
Rhinorrhoea	4 (8·0%)
Nausea and vomiting	4 (8·2%)[f]
Confusion	1 (2·0%)
More than one sign or symptom	41 (82·0%)
Asymptomatic	2 (4·0%)
Complications[a]	
Acute respiratory distress syndrome	2 (4·0%)
Acute respiratory injury	1 (2·0%)
Acute renal injury	5 (10·0%)
Septic shock	1 (2·0%)
More than complications	2 (4·0%)
No complications	45 (90·0%)
Radiological findings	
Unilateral pneumonia	11 (22·0%)
Bilateral pneumonia	27 (54·0%)
Multiple mottling and ground-glass opacity	17 (34·7%)[f]
No abnormality	4 (8·0%)
Co-infection	
Other viruses	0
Bacteria	2 (4·0%)[g]
Fungi	0
Clinical outcome[b]	
In serious condition (Admission to ICU)	3 (6·0%)
Hospitalized; in stable condition	27 (54·0%)
Discharged	20 (40·0%)
Interval from symptom onset to hospital admission (Days)[c]	
Average (SD)	8·5 (3·9)
Range	1–26
Length of hospital stay (Days)[d]	
Average (SD)	17·7 (7·7)
Range	8–35

[a] The definitions of the complications are provided in Supplementary materials.

[b] The data cut-off of the clinical outcome of the patients is Feb 28, 2020.

[c] The data from symptomatic patients were excluded.

[d] This is calculated based on the 18 patients who had been discharged as of Feb 28, 2020.

[e] Data were missing for two patients.

[f] Data were missing for one patient

[g] Moderate growth of *Klebsiella aerogenes* and *Ralstonia pickettii* were obtained from sputum specimens collected from case no. 38 and 70 respectively.

Table 2: Clinical characteristics and outcomes of 50 patients recruited in this study

Amino acid substitutions[a]	All cases (n=50)	Cases and Transmission Clusters of COVID-19 in the Hong Kong Community						Local Singletons[b] (n=5)	Imported Cases[c] (n=3)
		Cluster 1 (n=4)	Cluster 2 (n=3)	Cluster 3 (n=3)	Cluster 4 (n=11)	Cluster 5 (n=8)	Cluster 6 (n=13)		
Orf1ab									
G295V	13 (26·0%)	-	-	-	-	-	13 (100·0%)	-	-
P380L	1 (2·0%)	-	-	-	-	1 (12·5%)	-	-	-
E444K	1 (2·0%)	-	-	-	-	-	-	1 (20·0%)	-
A599D	1 (2·0%)	-	-	-	-	1 (12·5%)	-	-	-
V682A	1 (2·0%)	-	1 (33·3%)	-	-	-	-	-	-
A702T	1 (2·0%)	-	-	-	-	1 (12·5%)	-	-	-
Q1001E	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
D1323A	1 (2·0%)	-	-	-	-	1 (12·5%)	-	-	-
V1399E	1 (2·0%)	-	-	-	-	1 (12·5%)	-	-	
T1471I	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
V1483I	1 (2·0%)	-	-	-	-	-	-	1 (20·0%)	-
T1864M	1 (2·0%)	-	-	-	-	1 (12·5%)	-	-	-
Y1868N	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
A1983V	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
F2215S	1 (2·0%)	-	-	-	-	1 (12·5%)	-	-	-
W2232C	1 (2·0%)	-	-	-	-	1 (12·5%)	-	-	-
H2799Q	2 (4·0%)	-	-	-	-	1 (12·5%)	1 (7·7%)	-	-
D2833G	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
R2875G	2 (4·0%)	-	-	-	-	-	2 (15·4%)	-	-
S2972P	1 (2·0%)	-	-	-	-	1 (12·5%)	-	-	-
L3116F	1 (2·0%)	-	-	-	-	-	1 (7.7%)	-	-
M3131V	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
H3233Y	30 (60·0%)	-	-	3 (100·0%)	11 (100·0%)	-	12 (92·3%)	4 (80·0%)	-
I3257L	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
Y3500F	1 (2·0%)	-	1 (33·3%)	-	-	-	-	-	-
F3604L	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
L3606F	10 (20·0%)	-	-	-	-	-	10 (76·9%)	-	-

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

L3715S	1 (2·0%)	-	-	-	-	-	-	-	1 (33·3%)
W4124C	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
T4164I	1 (2·0%)	-	-	-	1 (9·1%)	-	-	-	-
A4247V	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
I4352L	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
Y4379F	2 (4·0%)	-	1 (33·3%)	-	-	-	1 (7·7%)	-	-
D4695Y	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
D4729V	1 (2·0%)	-	-	-	-	1 (12·5%)	-	-	-
G4895V	1 (2·0%)	-	-	-	-	1 (12·5%)	-	-	-
S5360P	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
V5384G	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
I5445V	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
I5547V	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
K6275N	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
P6318S	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
D7091E	2 (4·0%)	-	-	-	-	-	-	1 (20·0%)	1 (33·3%)
S									
L8V	27 (54·0%)	-	-	3 (100·0%)	11 (100·0%)	-	9 (69·2%)	4 (80·0%)	-
D138Y	3 (6·0%)	-	-	-	3 (27·3%)	-	-	-	-
L249F	1 (2·0%)	1 (25·0%)	-	-	-	-	-	-	-
K310E	1 (2·0%)	-	-	-	-	-	-	1 (20·0%)	-
V367F	1 (2·0%)	-	-	-	-	-	-	-	1 (33·3%)
K458N	1 (2·0%)	1 (25·0%)	-	-	-	-	-	-	-
R509K	1 (2·0%)	-	-	-	-	1 (12·5%)	-	-	-
V510L	1 (2·0%)	-	-	-	-	1 (12·5%)	-	-	-
P715H	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
D820N	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
P1263L	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
Orf3a									
G251V	44 (88·0%)	-	3 (100·0%)	3 (100·0%)	11 (100·0%)	8 (100·0%)	13 (100·0%)	5 (100·0%)	1 (33·3%)
M269T	1 (2·0%)	-	-	-	-	-	-	-	1 (33·3%)
M									

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

D209Y	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
Orf8									
V62L	1 (2·0%)	-	-	-	-	-	-	-	1 (33·3%)
L84S	1 (2·0%)	-	-	-	-	-	-	-	1 (33·3%)
N									
T24N	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-
R226S	1 (2·0%)	-	-	1 (33·3%)	-	-	-	-	-
G316E	1 (2·0%)	-	-	-	-	-	-	1 (20·0%)	-
M322I	1 (2·0%)	-	-	-	-	1 (12·5%)	-	-	-
K363M	1 (2·0%)	-	-	-	-	-	1 (7·7%)	-	-

"-" indicates 0 (0·0%).

[a] The location of amino acid substitutions was based on SARS-CoV-2 reference genome (NC_045512.2).

[b] The singleton cases did not have recent travel history.

[c] The patients had record of travel to Wuhan City.

Table 3: Amino acid substitutions identified in SARS-CoV-2 isolated from the 50 patients in this study

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure Legend

Figure 1. Demographics of patients included in this study

A rectangular box stands for male and eclipse for female. Travel history within the 14-day incubation period prior to symptom onset is highlighted in 1) Cyan for local case without travel history; 2) Pink for travel history to Wuhan within 14-days from symptom onset; 3) Orange for travel history to other regions in Mainland China, and 4) Green for travel history to regions outside Mainland China. The case numbers are those used by the Centre of Health Protection, Department of Health, Hong Kong.[6] Within each cluster, the cases of primary infection (leftmost column) are arranged from top to bottom in the order of the date (dd/m) of symptom onset.

Figure 2. Maximum-likelihood phylogenetic tree of the 50 COVID-19 cases in this study

The tree was constructed using all 50 COVID-19 cases included in this study and rooted on the earliest published genome of SARS-CoV-2 (accession no.: NC_045512.2). Bootstrap value was set at 1000×. Samples were colour-coded by epidemiological link as follows: 1) Magenta represents Cluster 1: Imported cluster; 2) Cyan represents Cluster 2: unknown source; 3) Green represents Cluster 3: local; 4) Orange represents Cluster 4: Hotpot party-related Superspreading event (SSE); 5) Blue represents Cluster 5: Public housing estate-related SSE; and 6) Red represents Cluster 6: Buddha worship hall-related SSE. Case 84 and Case 102 were asymptomatic at the time of sample collection, and were marked with asterisks (*) in the diagram. Each case is indicated with a case number (see figure 1) followed by the date (dd/mm/yyyy) of symptom onset.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Contributors

KSSL and GKHS designed the study, collected, analysed and interpreted the data, conducted literature search, and drafted and critically reviewed the manuscript; and GKHS was also responsible for securing funding for this study. TTLN, HYL, MPC, KKGT, and LKL conducted experiments and analysed the data. AKLW, MCYY, BKCW, AYMH, KTY, KCL, RWTL, EYKT, WSL, MCC, YYN, KMS, KSCF, SKYC, WKT, and TLQ were responsible for collecting and analysing the data, and finalizing the manuscript. DHKS and SPY analysed the data, and wrote and critically reviewed the manuscript. WCY were involved in data analysis, data interpretation and writing of the manuscript. All authors reviewed and approved the final version of the manuscript.

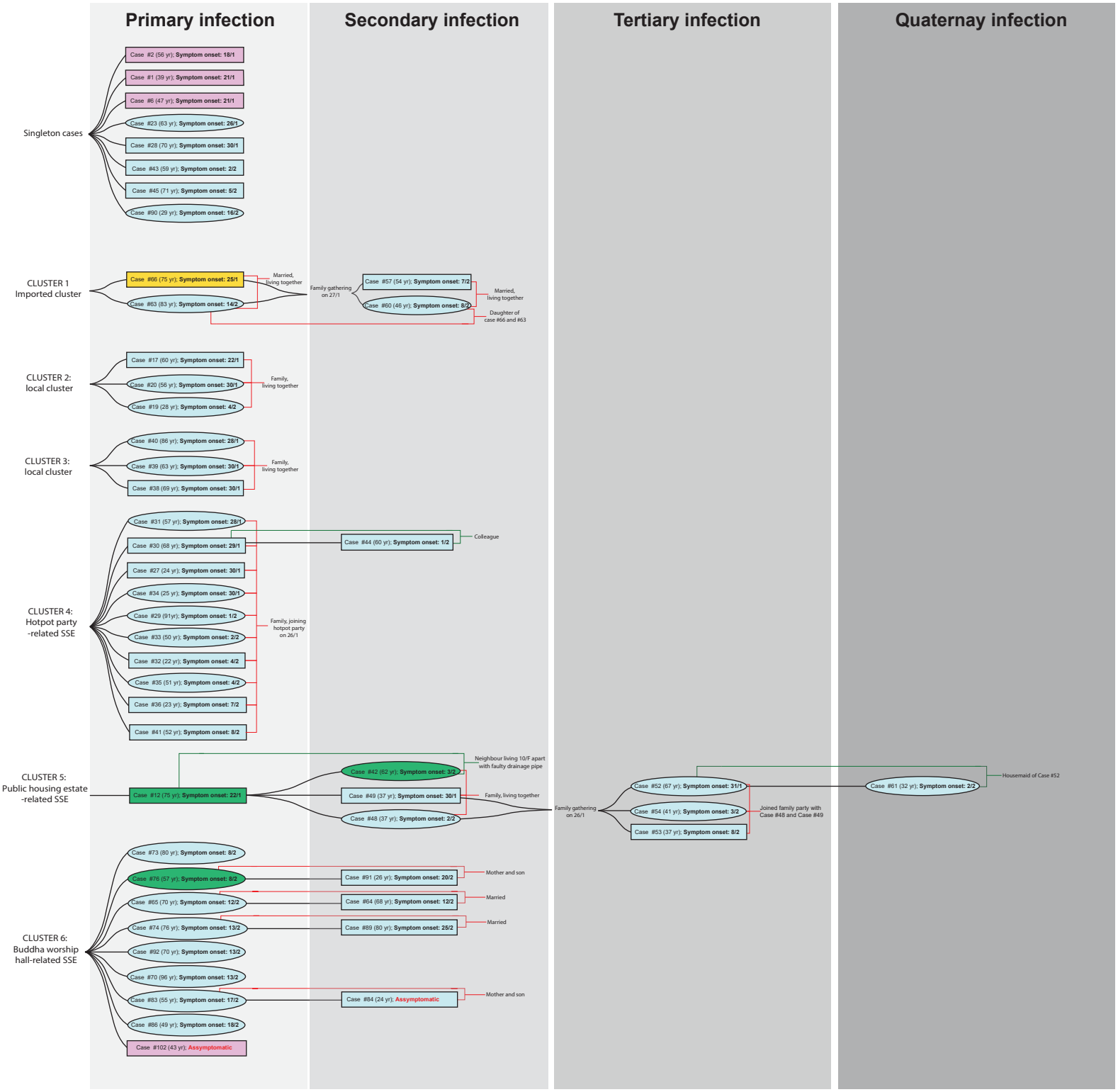
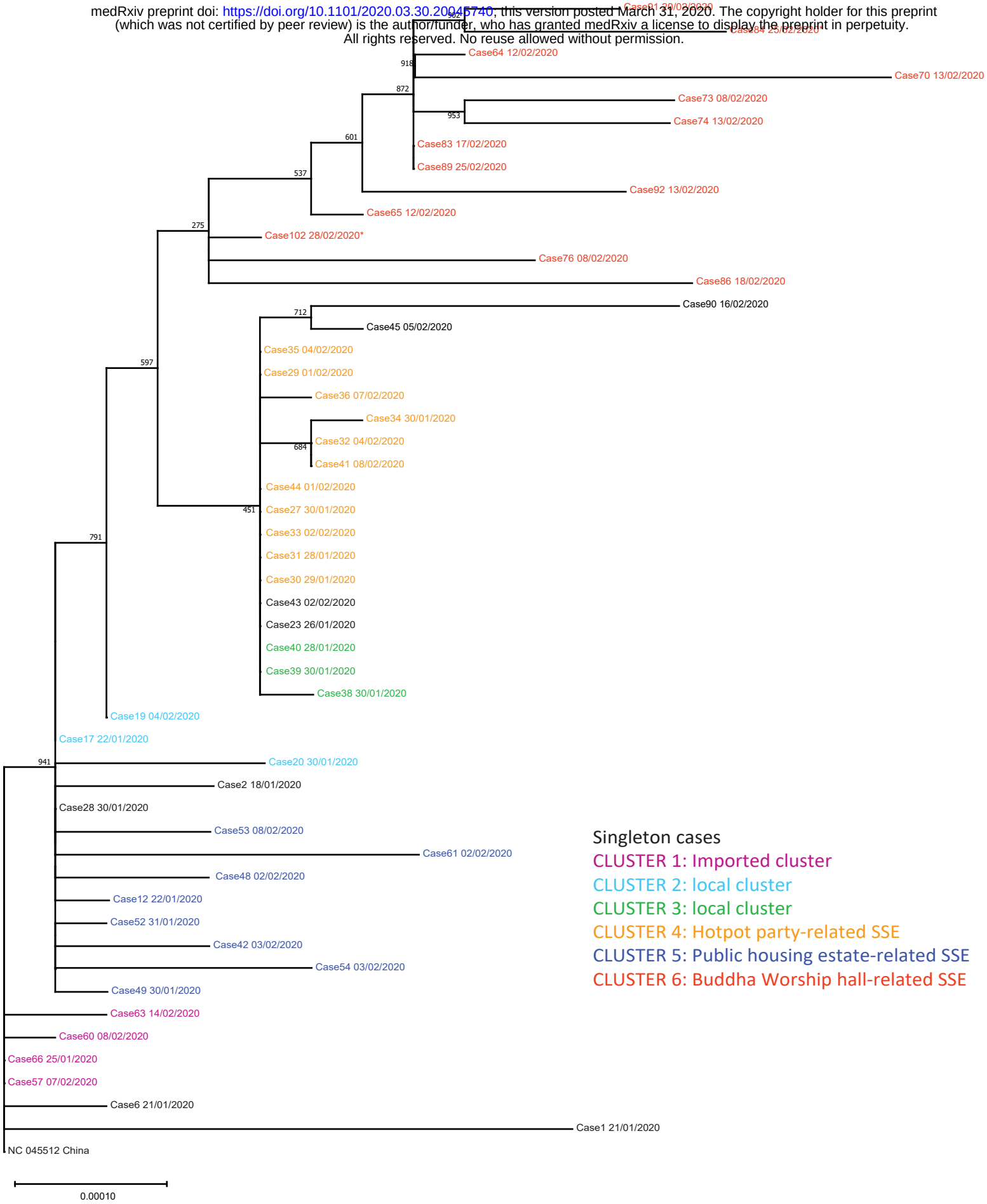
Declaration of interests

We declare no competing interests.

Acknowledgments

This study was supported by Faculty of Health and Social Science and Department of Health Technology and Informatics of The Hong Kong Polytechnic University. 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.

medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740; this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.



medRxiv preprint doi: https://doi.org/10.1101/2020.03.30.20045740, this version posted March 31, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.