medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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.

Evidence and characteristics of human-to-human transmission of 2019-nCoV

Min Kang, Ph.D., Jie Wu, Ph.D., Wenjun Ma, Ph.D., Jianfeng He, M.D., Jing Lu, Ph.D., Tao Liu, Ph.D., Baisheng Li, Ph.D., Shujiang Mei, M.D., Feng Ruan, M.D., Lifeng Lin, Ph.D., Lirong Zou, M.D., Changwen Ke, M.D., Haojie Zhong, M.D., Yingtao Zhang, M.D., Xuguang Chen, M.D., Zhe Liu, Ph.D., Qi Zhu, M.D., Jianpeng Xiao, Ph.D., Jianxiang Yu, M.D., Jianxiong Hu, M.D., Weilin Zeng, M.D., Xing Li, M.D., Yuhuang Liao, M.D., Xiujuan Tang, M.D., Songjian Xiao, M.D., Ying Wang, M.D., Yingchao Song, M.D., Xue Zhuang, M.D., Lijun Liang, M.D., Siqing Zeng M.D., Guanhao He, Ph.D., Peng Lin, M.D., Tie Song, M.P.H.

Guangdong Provincial Center for Disease Control and Prevention (M.K., J.W., J.H., B.L., L.L., L.Z., C.K., H.Z., Y.Z., X.C., Q.Z., J.Y., Y.L., Y.S., X.Z., L.L., P.L., T.S.)

Guangdong Provincial Institute of Public Health, Guangdong Provincial Center for Disease Control and Prevention (W.M., J.L., T.L., Z.L., J.X., J.H., W.Z., X.L., G.H., S.Z., W.M.)

Shenzhen Center for Disease Control and Prevention (S.M., X.T.,)

Zhuhai Center for Disease Control and Prevention (F.R., S.X.)

Shenzhen Nanshan Center for Disease Control and Prevention (Y.W.)

Drs. M. Kang, J. Wu, W. Ma., J. He, J. Lu, T. Liu and B. Li contributed equally to this article.

Address reprint requests to Dr. Song at the Guangdong Provincial Center for Disease Control and Prevention, Guangzhou 511430, China, or at tsong@cdcp.org.cn

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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/10.1101/2020.02.03.20019141; this version posted February 5, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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.

Abstract

Background

On December 31, 2019, an outbreak of 2019-nCoV in humans was reported in Wuhan, China. We analyzed data from field investigations and genetic sequencing to provide evidence and characteristics of human-to-human transmission.

Methods

A confirmed case of 2019-nCoV was defined if a suspected case was verified with positive of 2019-nCoV in throat swabs, nasal swabs, bronchoalveolar lavage fluid (BALF), or endotracheal aspirates by real-time reverse transcriptase polymerase chain reaction assay (RT-PCR) or genetic sequencing. Field investigations were conducted for each confirmed case. Clinical and demographic data of the confirmed cases were collected from their medical records. Exposure and travel history were obtained by interviewing confirmed cases.

Results

A total of 188 confirmed cases were identified from January 1 to 27, 2020 in Guangdong Province, China. Of them, 84 (44.6%) cases were from 31 cluster infections. Thirty cases (16.0%) were identified as secondary cases, in which 25 and 9 cases were identified in cluster infections and family cluster infections, respectively. 2019-nCoV were detected in three cases with mild respiratory symptoms, and in two asymptomatic cases. The whole viral genomes within the same family cluster infections were exactly the same, and presented a few unique single nucleotide variants (SNVs) compared with 2019-nCoVs identified in Wuhan on December 2019.

Conclusions

We observed increasing human-to-human transmissions of 2019-nCoV in Guangdong, China, and

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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.

most of them were identified in cluster infections. Our findings indicate that prevention strategies of containing the person-to-person transmission of 2019-nCoV in households, hospitals and communities are urgently needed.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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

On December 31, 2019, Wuhan Municipal Health Commission reported 27 cases of pneumonia infection with unknown aetiology in Wuhan, Hubei Province, China[1,2]. A novel coronavirus (2019-nCoV) was identified as the causative virus by Chinese authorities on January 7, 2020[3]. The common symptoms of the cases infected with 2019-nCoV include fatigue, fever, cough, shortness of breath and breathing difficulties. Severe infection can cause pneumonia, severe acute respiratory syndrome, kidney failure and even death[4]. As of January 28, 2020, 5974 confirmed cases with 1239 critical cases and 132 deaths were reported in 31 provinces, regions and municipal cities across China, with 3554 in Hubei Province, 296 in Zhejiang Province, and 241 in Guangdong Province. Additionally, 47 cases were reported in other countries including United States, Thailand, South Korea, etc. Almost all provinces and regions in China have initiated the highest level of public health emergency response.

As an emerging infectious disease, the characteristics of 2019-nCoV such as natural host and reservoir, incubation period, serial interval, and the infectivity of human-to-human transmission remain unknown. We described evidence and characteristics of human-to-human transmission of 2019-nCoV in Guangdong Province, which will be valuable to decision making on prevention and control strategies for community spreading.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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.

Methods

Case definition and identification

The suspected and confirmed human infections with 2019-nCoV were defined based on the Diagnosis and Treatment Scheme of 2019-nCoV released by the National Health Commission of China (Supplementary appendix).[5]

After the outbreak of 2019-nCoV in Wuhan announced on December 31, 2019, enhanced surveillance was initiated in Guangdong Province to early recognize suspected infections, especially among those who had history of travelling to Wuhan within 14 days. Once a suspected case was diagnosed, the Center for Diseases Control and Prevention (CDC) at prefecture level would conduct an initial field investigation, and collect respiratory specimens. The respiratory specimens were shipped to Guangdong Provincial CDC (GDCDC) for 2019-nCoV test by RT-PCR or genetic sequencing.

Epidemiological data Collection

Demographic information, date of illness onset, and clinical outcomes of confirmed cases were collected from medical records. Through interviewing with confirmed cases or their relatives, we obtained exposure histories for each case during 14 days before the onset of illness including date, frequency, and patterns of exposures to wild animals.

Metagenomic Sequencing

Clinical samples including bronchoalveolar lavage fluid (BALF), endotracheal aspirates, throat swabs, and nasal swabs were collected for suspected cases. Total RNAs were extracted and RT-PCR was firstly performed according to the standard method provided by the China CDC. Next, the positive samples with cycle threshold value lower than 35 were used for metagenomic

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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.

sequencing to obtain the viral genomic sequences. Sample RNA was treated with TURBO DNase (Invitrogen, Cat. No. AM2239) and then purified by using Agencourt RNAClean XP beads (Beckman Cat. No. A63987). Libraries were prepared using the SMARTer Stranded Total RNA-Seq Kit v2 (Clontech, Cat. No. 634412) according to the manufacturer's protocol. Pair-end sequencing was performed on Miseq (Illumina, USA).

Statistical and Sequences analyses

We used descriptive statistics to summarize the epidemiological characteristics of confirmed cases. The sequencing data were quality controlled (QC) using fastp[6] and then mapped to the human genome (hg38) using Bowtie2[7] to remove human related RNA reads. De novo assembling was performed using Spades[8] and Megahit[9] for each dataset. Assembled contigs were annotated using blastn[10] using the complete nucleotide database. Twelve genome sequences generated in this study have been submitted to GISAID under the accession numbers EPI403932–EPI403937 and EPI406533–EPI406538. Maximum-likelihood (ML) trees of viral genome were estimated via IQ-Tree[11] by integrating closely related sequences from SARs and SARs-like viruses from public database. Two nCoV genomes from introduction cases in Thailand were also included and data analysis was approved by the corresponding submitter. SNV (single nucleotide variant) sites were found through SNV-sites[12]. Phylogenetic trees were annotated and visualized with ggtree[13].

Ethics Approval

Data collection and analysis of cases and close contacts were determined by the National Health Commission of the People's Republic of China to be part of a continuing public health outbreak investigation and were thus considered exempt from institutional review board approval.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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

Description of the outbreak

Between January 1 and 27, 2020, a total of 188 confirmed cases were identified in 16 of 21 cities in Guangdong Province, with most cases in Guangzhou (n=51) (Figure 1). The average age of all cases was 48.8 years, and the percentage of males was 49.5% (93/188). Of total cases, 141 (75.0%) and 17 (9.0%) cases had history of travelling to Wuhan, and other cities in Hubei Provinces, respectively. Thirty (16.0%) cases were identified as secondary cases with 25 cases from cluster infections. The average duration from onset of symptoms to diagnosis was 5.4 days (Table 1).

Of total cases, 84 cases (44.6%) were identified in 31 cluster infections who, compared with non-cluster infections, had lower proportions of travelling to Wuhan (63.1% *vs.* 84.6%) or other cities (7.1% *vs.* 10.6%) in Hubei Province within 14 days prior to the onset of illness, and more of whom were seriously (17.9% *vs.* 7.7%) or critically ill (6.0% *vs.* 1.0%). Of the cluster infections, 37 (44.0%) cases were identified in 13 family cluster infections (Table 1).

Characteristics of 13 family cluster infections

The 13 family cluster infections were reported in Shenzhen (n=4), Zhuhai (n=3), Guangzhou (n=2), Foshan (n=2), Shaoguan (n=1) and Yangjiang (n=1) (Figure 2). The most common symptoms in 37 cases included fever, fatigue, and cough (Table S1). Of 37 cases, 3 and 10 cases were critically and severely ill, respectively. Three cases in family cluster E had only mild respiratory symptoms such as mild cough, and 2 cases in family cluster M did not report any symptoms. Thirty-four cases had no exposure to wild animals within 14 days prior to the onset of illness. Nine cases were identified as secondary cases. Here we described the five family cluster infections with 9 secondary cases, and the detailed information of other family cluster infections

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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.

were shown in the appendix (Figure S1 and Table S1).

This is the first family cluster infections identified in Guangdong Province. There were 5 cases in family cluster A. Case 1, Case 2, Case 3 and Case 4 stayed in Wuhan during December 29, 2019 to January 4, 2020. Case 1 had symptoms of cough, fatigue, fever and diarrhea on January 3, 2020 when they were in Wuhan. Case 2 (wife of Case 1) had fatigue and fever on January 4 after they returned Shenzhen. The couple visited two hospitals on January 6 and January 10 when they had fever (>38□) and chest radiologic changes. Their endotracheal aspirates were collected and tested by Center for diseases control and prevention (CDC) on January 13, and were identified positive infections of 2019-nCoV. Case 3 (son-in-low of Case 1) and Case 4 (grandson of Case 1) had fever, stomachache, diarrhea and muscular soreness on January 1 in Wuhan, and visited a hospital on January 11 in Shenzhen. Their blood and throat swab samples were also identified positive with 2019-nCoV. These four cases had no history of exposure to wild animals. Case 5 (mother of Case 3) did not visit Wuhan within 14 days before the onset of symptoms, but lived with Case 3 in an apartment in Shenzhen. On January 8, she had low fever and breathless. On January 14, she visited a hospital and was isolated for treatment. Her blood and throat swab samples were collected and tested by CDC on January 15, and identified positive of 2019-nCoV.

Family cluster B

Three cases were identified in family cluster B in Zhuhai, including a father (Case 1), a mother (Case 2) and their daughter (Case 3). Case 1 and Case 2 resided in Wuhan, and visited Case 3 in Zhuhai on January 11. On their trip to Zhuhai, Case 1 began to have symptom of fatigue. On January 12, Case 1 developed dry cough, fatigue, headache, eye pain and muscular soreness, and Case 2 had symptoms of dry cough, fatigue, headache. On January 15, Case 1 and Case 2 visited a

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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.

hospital. Their throat swabs were collected and tested by CDC on January 17, and laboratory diagnosis showed positive of 2019-nCoV. Case 3 had symptoms of fever, occasional itching and dry cough on January 17 without history of visiting Wuhan within 14 days prior to the onset of symptoms. She visited a hospital on January 18, and 2019-nCoV was identified in throat swab samples by CDC.

Family cluster C

There were three cases in family cluster C in Shaoguan. Case 1 resided in Ezhou, Hubei Province, and took a high-speed train at Wuhan station to visit her parents, brothers and sister (Case2) in Shaoguan on January 15. She had fever, chills and itchy throat on January 16, and visited a hospital with Case 2 on January 19, when Case 2 also developed fever, cough and sore throat. They visited a hospital again and were transferred to another hospital on January 20. Their throat swabs were collected and tested by CDC, and laboratory diagnosis showed positive of 2019-nCoV. Case 3 (husband of Case 2) had symptoms of dizziness and fever, and identified with positive of 2019-nCoV on January 22. Case2 and Case3 had no history of visiting to Wuhan within 14 days before the onset of the symptoms.

Family cluster D

Family cluster D in Zhuhai had three cases. Case 1 stayed at Wuhan from January 8 to 17, and returned Zhuhai on January 17. On the trip, he had symptom of muscular soreness. Case 2 (wife of Case 1) lived with Case 1 in Zhuhai and had cough on January 20 without history of visiting Wuhan within 14 days before the onset. She visited a hospital with Case 1 on January 20, and were transferred to another hospital on January 21 for treatment and isolation. Their nasal and throat swabs were tested and identified with positive of 2019-nCoV on January 22. Case3 (mother

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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.

of Case 1) lived with Case1 and Case2 without history of visiting Wuhan in the past 14 days. Her nasal and throat swabs were also identified with positive of 2019-nCoV on January 22, when she had low fever and fatigue, and was isolated on January 23.

Family cluster E

There were 4 cases in family cluster E in Guangzhou. Case 1 resided in Wuhan and visited her daughter in Guangzhou on January 19. Case 1 had fever, dry cough and slight headache on January 20 and visited a hospital with Case 2 (daughter of Case 1) and Case 3 (husband of Case 2) on January 21. Throat swabs of Case 1 were collected and tested by CDC on January 22, and laboratory diagnosis showed positive of 2019-nCoV. Case 1 had no history of wildlife exposure. CDC also collected biological samples from Case2, Case3, and Case4 (daughter of Case 2 and Case 3) on January 23, and laboratory diagnosis showed all of them were positive of 2019-nCoV. Case 2 and Case 3 had only mild symptoms of occasional cough, and Case 4 had occasional cough and low fever. Case2, Case 3 and Case 4 had no history of visiting in Wuhan and exposure to wild animals in past 14 days.

Phylogenetic analyses

The phylogeny based on full-length of 2019-nCoV genomes (12 generated in this study) showed 2019-nCoV could be classified into SARS-like virus with the closest related strains as MG772934 and MG772933 which were collected from Rhinolophus sinicus in China on 2015 and 2017, respectively. SARS viruses shared nearly 97.5% nucleotide similarity with closely related SARS-like viruses from bats. By contrast, 2019-nCoV had lower nucleotide similarity with related SARS-like viruses (87.4%-87.5%), and the long internal branch was observed between 2019-nCoV and related SARS-like viruses (Figure 4A).

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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 characterize the 2019-nCoVs from family clusters and investigate potential genetic adaptation, we compared the twelve 2019-nCoV genome sequences generated between 14 January 2020 and 23 January 2020 in this study and all published 2019-nCoV [14,15]. Sequences from the three family cluster infections fell into three corresponding genetic clusters and could be separated from the sequences of sporadic infections (Figure 4B). The viral genome sequences from cases within a family cluster were exactly the same. To find whether there were any genetic changes following the illness progress, we sequenced 2019-nCoV viruses from case 1 of cluster B on 17 January and 22 January which was about 1 week and 2 weeks after the onset of illness, respectively. These two sequences were exactly the same suggesting the genetic stability of 2019-nCoV during the infection. Notably, the viral genome from family cluster A share three unique SNV sites which could not be found in other sequenced 2019-nCoVs in public database. Among the three SNVs, one was predicted to cause amino acid changes (nonsynonymous variations) in ORF8 protein coding sequence led the amino acid change from Leu to Ser. One unique SNV site was identified in family cluster B and the change from C to T at position 21442 within the spike protein coding sequence leading to amino acid change from His to Thr. Only one synonymous variation was identified in viral genome sequences from family cluster C. Due to the increasing epidemic activity, we also investigated whether there were any parallel mutations could be identified in recently identified sporadic infection cases which regarded as an indicator of potential genetic adaptation. Three 2019-nCoVs genetic sequences from sporadic cases on 22 January and 23 January showed there were no common SNV sites could be identified in these cases and the SF201_20/01/23 2019-nCoV collected on 23 January had exactly the same genetic sequence with viral sequences collected at the beginning of the epidemic.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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/10.1101/2020.02.03.20019141; this version posted February 5, 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.

Discussion

Eighteen years after the first emergence of SARS coronavirus in Guangdong at the end of 2002[16], a great challenge presents by the emergence of a new SARS-like coronavirus. The early epidemiological data suggested most cases were caused by spill-over infections from nonhuman sources which is still unknown. In this study, we provided the epidemiological and genetic evidence of human-to-human transmissions of 2019-nCoV.

As of January 27, increasing secondary cases were identified in Guangdong Province. In all 30 secondary cases treated in hospital, none of them had history of residence or traveling to Wuhan within 14 days prior to onset of the illness. Moreover, 14 medical professionals were also reported to be infected by a case in Hubei province. This suggests that 2019-nCoV transmission pattern has been transferring from exposure to nonhuman resources to person-to-person. We observed several confirmed cases with mild symptoms or no symptoms, which indicates that current monitoring measures such as fever testing may not work effectively to identify these asymptomatic cases, and recent clinical evidence suggested that asymptotic cases could also transmit 2019-nCoV to other humans. These findings suggest that the human-to-human transmission of 2019-nCoV may extend from families to communities. “Super spreading events” may occur in some specific circumstances such as hotels and hospitals if asymptotic cases were not detected earlier. Therefore, much more rigorous measures of cutting person-to-person transmission are urgent. Health education must be made to improve the perception and awareness in the general public.

The sharp rise in number of confirmed cases and the emergence of more and more cluster infections raise much concern on 1) whether there are any genetic mutations associated with these

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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 infections, and 2) whether there are any potential viral adaptations leading the increasing epidemic activity. All these questions are fundamental to decision making of disease control strategy, which have not been fully answered yet. In this study, the viruses sequenced from two secondary cases are exactly the same with the viruses from corresponding index cases in the family, and these secondary cases had no history of travelling to Wuhan and no exposure to wild animals, which indicated that these viruses likely represent strains which are able to transmit from human to human. We also searched any genetic clues that might be associated with viral adaptation to human by comparing the viruses from family cluster infection with currently published 2019-nCoV sequences. Only three and one unique SNV sites are identified from viruses sequenced in family infection cluster A and B, respectively. These SNV sites are predicted to result in two synonymous change from Leu to Ser in Orf 8 and from His to Thr in spike protein, respectively. It is still unknown whether these amino acid changes provide or increase the virus capability of human transmission. However, the previous study from SARS suggests Orf8 is either noncoding or coding for a functionally unimportant putative protein[17]. As a result, few SNVs detected between 2019-nCoVs from secondary cases in Guangdong and 2019-nCoVs reported in Wuhan indicates 2019-nCoV could already transmit among humans at the beginning of the epidemic when it was still regarded as spill-over infections.So, more asymptomatic transmission may occur at the end of December 2019 than we previously expected. Another reason that we believe 2019-nCoVs may be already well established in humans since its first identification is that the 2019-nCoVs we identified onJanuary, 23, 2020 were exactly the same with viruses identified in Wuhan on 30 December 2019. Together with the 2019-nCoV genome sequences generated on 15 and 22 January, we did not find any parallel mutations within the genome of 2019-nCoV

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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.

accompany with the increasing epidemic. These data suggest the genetic stability of 2019-nCoV during the epidemic, which is contrast to SARS viruses for which the deletions and common SNVs are observed in the early phase of the outbreak [17].

Our findings indicate that with the progress of the outbreak, prevention strategies of containing person-to-person transmission of 2019-nCoV in communities are urgently needed in the near future. Enhanced surveillance for 2019-nCoV in communities, early detection and isolation of contacts of confirmed cases, and virologic analyses to assess genetic changes that might suggest increased transmissibility among humans are all critical to informing prevention and control efforts and assessing the global pandemic potential of 2019-nCoV.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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.

Acknowledgements

We thank all the medical and nursing staff who assisted in the care of patients; the members from health department and CDC in Guangdong Province for their contribution in data collection, 2019-nCoV control and prevention; Dr. Pilailuk Okada and his colleagues for sharing their sequence data. This work was supported by National Key Research and Development Program of China (2018YFA0606200, 2018YFA0606202), the Science and Technology Program of Guangdong Province (2018B020207006, 2019B020208005, 2019B111103001), Guangzhou Science and technology Plan Project (201804010383).

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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/10.1101/2020.02.03.20019141; this version posted February 5, 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. Wuhan Municipal Health Commission, Hubei Province, China. Wuhan Municipal Health Commission briefing on the pneumonia epidemic situation, 31 December 2019. 2019. (http://wjw.wuhan.gov.cn/front/web/showDetail/2019123108989.)

2. The 2019-nCoV Outbreak Joint Field Epidemiology Investigation Team, Li Q. An Outbreak of NCIP (2019-nCoV) Infection in China — Wuhan, Hubei Province. China CDC Weekly 2020;2:79-80.

3. World Health Organization. WHO Statement Regarding Cluster of Pneumonia Cases in Wuhan, China. 2020. (https://www.who.int/china/.)

4. World Health Organization. Coronavirus. 2020. (https://www.who.int/health-topics/coronavirus.)

5. National Health Commission of the People's Republic of China. Prevention and Control Scheme of Novel Coronavirus. 2020. (http://www.nhc.gov.cn/xcs/zhengcwj/202001/808bbf75e5ce415aa19f74c78ddc653f.shtml.)

6. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 2018;34:i884-i90.

7. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods 2012;9:357-9.

8. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 2012;19:455-77.

9. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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.

2015;31:1674-6.

10. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. BMC Bioinformatics 2009;10:421.

11. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol 2014;32:268-74.

12. Page AJ, Taylor B, Delaney AJ, et al. SNP-sites: rapid efficient extraction of SNPs from multi-FASTA alignments. Microb Genom 2016;2:e000056.

13. Yu G, Smith DK, Zhu H, Guan Y, Lam T-T, et al. GGTREE: an r package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol Evol 2017;8:28–36.

14. Zhou P, Yang X-L, Wang X-G, et al. Discovery of a novel coronavirus associated with the recent pneumonia outbreak in humans and its potential bat origin. bioRxiv 2020:2020.01.22.914952.

15. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med 2020. Doi: 10.1056/NEJMoa2001017.

16. Drosten C, Günther S, Preiser W, et al. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med 2003;348:1967-76.

17. He JF, Peng GW, Min J, et al. Molecular Evolution of the SARS Coronavirus, during the Course of the SARS Epidemic in China. Science 2004;303:1666-9.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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.

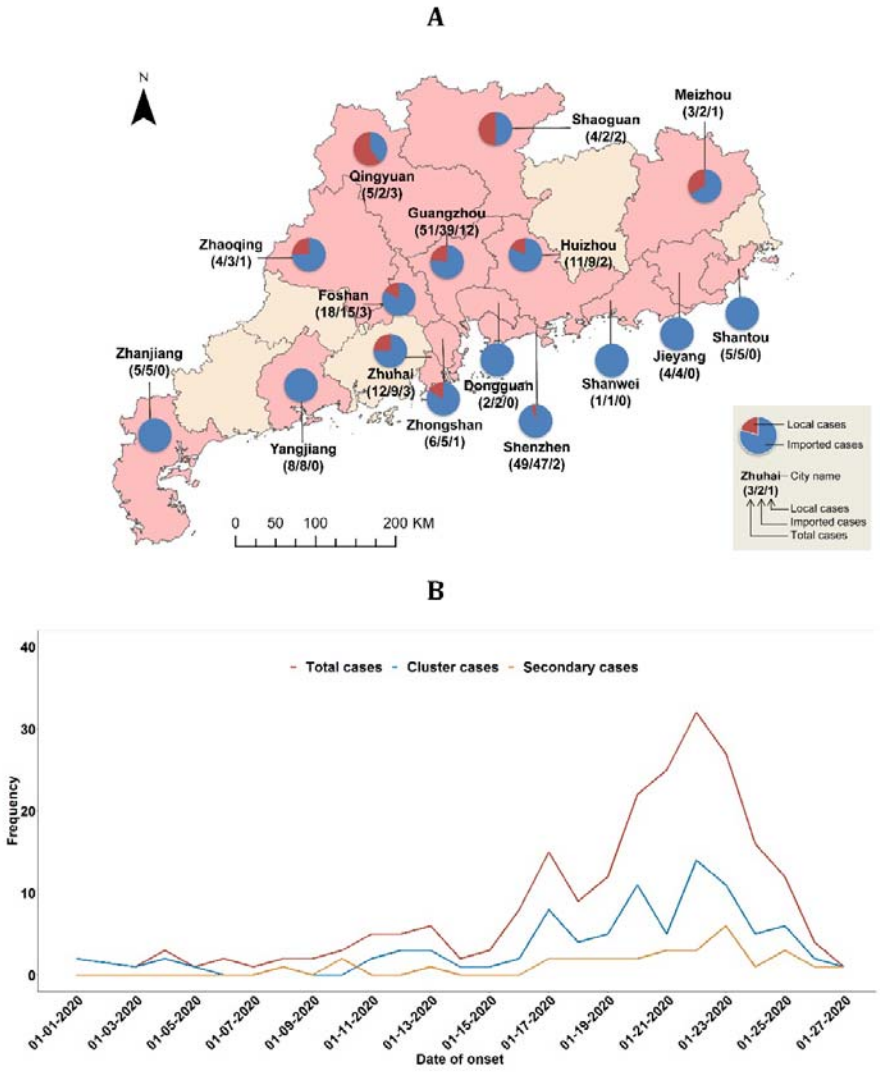

Figure 1. tempo-Spatial distribution of 188 confirmed cases of 2019-nCoV in Guangdong Province, China, January 1-27, 2020

Chart A: The spatial distribution of all confirmed cases.

Chart B: The temporal distribution three types of cases.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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.

Table 1. General characteristics of confirmed 2019-nCoV cases in Guangdong Province, China

	Total (n, %)	Cluster cases (n, %)	Non-cluster cases (n, %)	χ^2	P
Sex					
Male	93 (49.5)	41 (48.8)	52 (50.0)	0.002	0.988
Female	95 (50.5)	43 (51.2)	52 (50.0)		
Age (years)					
<30	34 (18.1)	12 (14.3)	22 (21.2)		
30~	30 (16.0)	14 (16.7)	16 (15.4)		
40~	21 (11.2)	6 (7.1)	15 (14.4)	10.79	0.056
50~	42 (22.3)	16 (19.0)	26 (25.0)		
60~	42 (22.3)	27 (32.1)	15 (14.4)		
≥70	19 (10.1)	9 (10.7)	10 (9.6)		
History of travelling to Wuhan and other cities in Hubei					
Wuhan	141 (75.0)	53 (63.1)	88 (84.6)	21.61	<0.001
Other cities in Hubei	17 (9.0)	6 (7.1)	11 (10.6)		
No	30 (16.0)	25 (29.8)	5 (4.8)		
Duration from onset to diagnosis (days)					
<2	19 (10.1)	13 (15.5)	6 (5.8)		
2~	50 (26.6)	20 (23.8)	30 (28.9)	6.46	0.168
4~	42 (22.3)	16 (19.1)	26 (25.0)		
6~	35 (18.6)	18 (21.4)	17 (16.3)		
≥8	42 (22.3)	17 (20.2)	25 (24.0)		
Severity of illness					
Mild	155 (82.4)	61 (72.6)	94 (90.4)		
Serious	23 (12.2)	15 (17.9)	8 (7.7)	10.82	0.013
Critical	6 (3.2)	5 (6.0)	1 (1.0)		
Recovery	4 (2.1)	3 (3.6)	1 (1.0)		

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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.

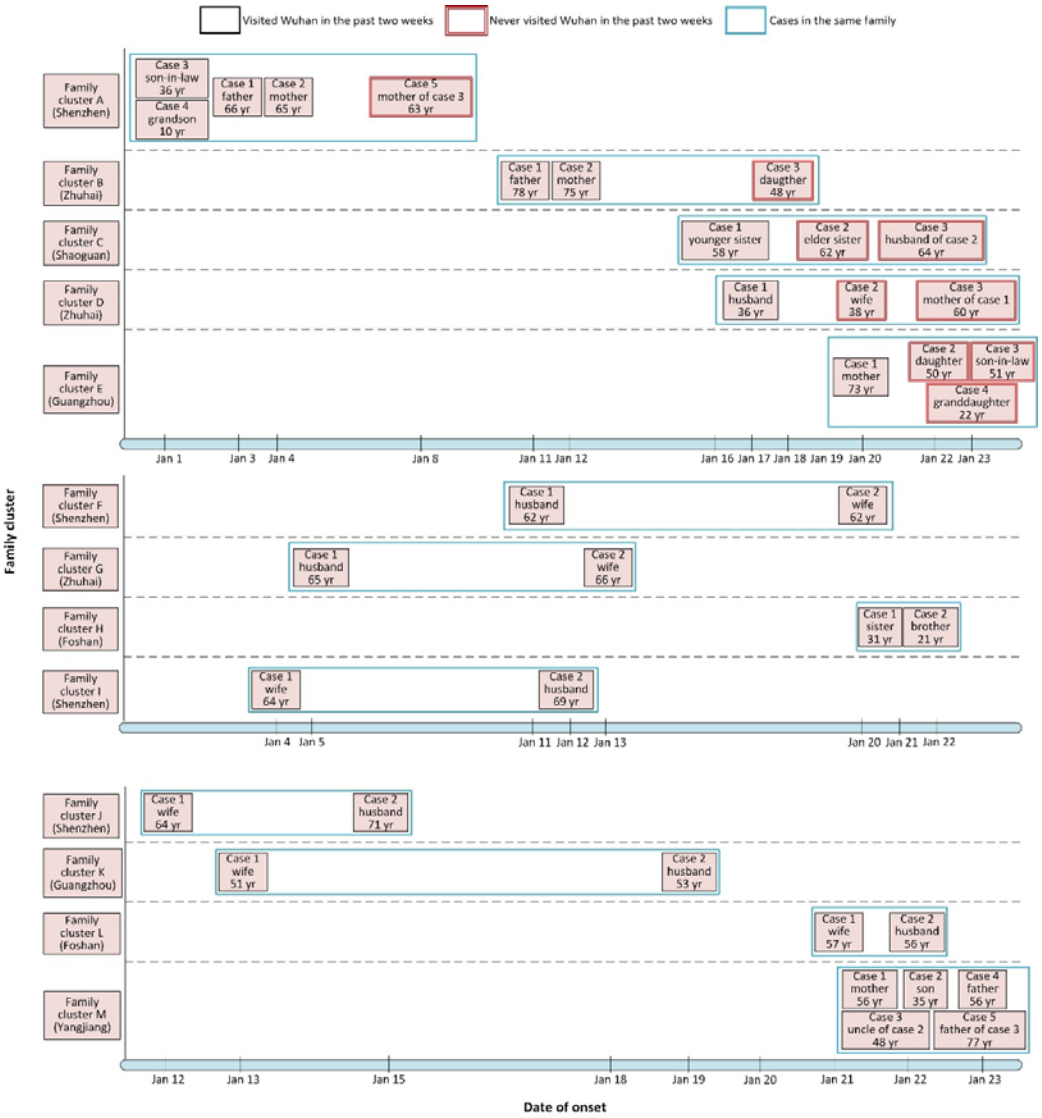

Figure 2. Onset date of 37 confirmed 2019-nCoV cases in 13 family cluster infections in Guangdong Province

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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.

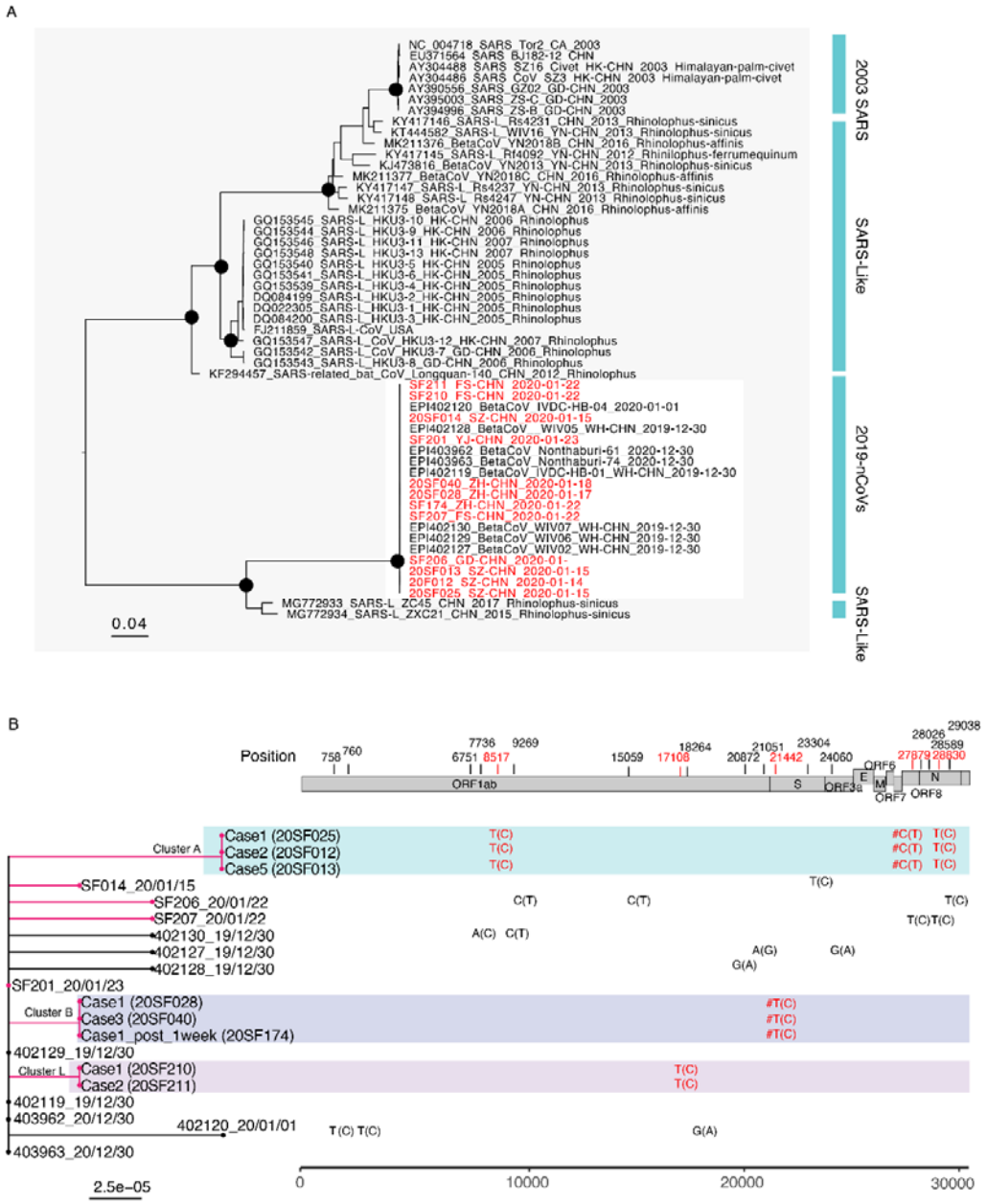

Figure 3. Phylogenetic analysis of complete genome of 2019-nCoVs identified in Guangdong and closely related viral genome from public database

Chart A: Maximum-likelihood tree was constructed by including 12 new generated 2019-nCoV sequences in this study, 6 published 2019-nCoV sequences and sequences from closely related SARS-like viruses. Two 2019-nCoV sequences from Thailand were integrated into the analysis with the permission of the submitter. Black circles indicate bootstrap support >0.9 at the root node of selected clade. The cluster of 2019-nCoVs were highlighted with red box. The sequences generated in this study were highlight with red.

medRxiv preprint doi: https://doi.org/10.1101/2020.02.03.20019141; this version posted February 5, 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.

Chart B: Phylogenetic relationship among 2019-nCoVs and SNV (single nucleotide variant) sites were identified by aligning twenty complete genomes of 2019-nCoVs collected between 30 December, 2019 and 23 January, 2020were found by the multiple genome sequences alignment. To ignore the sequences uncertainty at 5' and 3' terminal, the aligned sequences were started from position 266 to 29856 of 2019-nCoVs genome (use the EPI402119 as the reference genome). The phylogeny branches of Guangdong 2019-nCoVs sequences were highlight red and different family clusters were marked with different color boxes. Unique SNVs and their corresponding positions in reference viral genome identified in family cluster infections were highlighted with red. The nonsynonymous SNVs in cluster infections were marked with #.