

1 **Evidence and characteristics of human-to-human transmission of SARS-CoV-2**

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

9

10 Guangdong Provincial Center for Disease Control and Prevention (M.K., J.W., J.H., B.L., L.L.,
11 L.Z., C.K., H.Z., Y.Z., X.C., Q.Z., J.Y., Y.L., Y.S., X.Z., L.L., P.L., H.D., T.S.)
12 Guangdong Provincial Institute of Public Health, Guangdong Provincial Center for Disease
13 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.)
14 Shenzhen Center for Disease Control and Prevention (S.M., X.T.,)
15 Zhuhai Center for Disease Control and Prevention (F.R., S.X.)
16 Shenzhen Nanshan Center for Disease Control and Prevention (Y.W.)
17 Drs. M. Kang, J. Wu, W. Ma., J. He, J. Lu, T. Liu and B. Li contributed equally to this article.
18 Address reprint requests to Dr. Song at the Guangdong Provincial Center for Disease Control and
19 Prevention, Guangzhou 511430, China, or at tsong@cdcp.org.cn

20 **Abstract**

21 **Background**

22 On December 31, 2019, an outbreak of COVID-19 in humans was reported in Wuhan, and then
23 spread fast to other provinces, China. We analyzed data from field investigations and genetic
24 sequencing to describe the evidence and characteristics of human-to-human transmission in
25 Guangdong Province.

26 **Methods**

27 A confirmed COVID-19 case was defined if a suspected case was verified with positive of
28 SARS-CoV-2 in throat swabs, nasal swabs, bronchoalveolar lavage fluid (BALF), or endotracheal
29 aspirates by real-time reverse transcriptase polymerase chain reaction assay (RT-PCR) or genetic
30 sequencing. Field investigations were conducted for each confirmed case. Clinical and
31 demographic data of confirmed cases were collected from medical records. Exposure and travel
32 history were obtained by interview.

33 **Results**

34 A total of 1,151 confirmed cases were identified as of February 10, 2020 in Guangdong Province,
35 China. Of them, 697 (60.1%) cases were from 234 cluster infections. Two hundred and fourteen
36 (18.6%) were secondary cases, in which 144 cases were from family cluster infections. With the
37 epidemic continuing, although familial cluster events were dominated, community cluster events
38 increased with a nosocomial event. The whole genomes within the same family cluster infections
39 were identical, and presented a few unique single nucleotide variants (SNVs) compared with
40 SARS-CoV-2 identified on December 2019 in Wuhan.

41 **Conclusions**

42 We observed evident human-to-human transmissions of SARS-CoV-2 in Guangdong, China.
43 Although most of them were from family cluster infections, community and nosocomial infections
44 were increasing. Our findings indicate that human-to-human transmission risks are transferring
45 from family to community in Guangdong Province.

46 **Introduction**

47 On December 31, 2019, Wuhan Municipal Health Commission reported 27 cases of pneumonia
48 infection with unknown etiology in Wuhan, Hubei Province, China^{1,2}. A novel coronavirus
49 (SARS-CoV-2) was identified as the causative virus by Chinese authorities on January 7, 2020³.
50 The common symptoms of the cases infected with SARS-CoV-2 include fatigue, fever, cough,
51 shortness of breath and breathing difficulties. Severe infection can cause pneumonia, severe acute
52 respiratory syndrome, kidney failure and even death⁴. As of February 11, 2020, 44,653 confirmed
53 cases with 8,204 critical cases and 1,114 deaths were reported in 34 provinces, regions and
54 municipal cities across China, with 33,366 in Hubei Province (19,558 cases in Wuhan), 1,219 in
55 Guangdong Province, 1,135 in Henan Province, and 1,131 in Zhejiang Province. Additionally, 437
56 cases were reported in other 24 countries. Almost all provinces and regions in China have initiated
57 the highest level of public health emergency response.

58 As an emerging infectious disease, the characteristics of COVID-19 such as natural host and
59 reservoir, incubation period, serial interval, and infectivity of human-to-human transmission
60 remain unknown. We described evidence and characteristics of human-to-human transmission of
61 SARS-CoV-2 in Guangdong Province, which will be valuable to decision making on prevention
62 and control strategies for community spreading.

63 **Methods**

64 **Case definition and identification**

65 The suspected and confirmed human infections with SARS-CoV-2 were defined based on the
66 Diagnosis and Treatment Scheme of SARS-CoV-2 released by the National Health Commission
67 of China (Supplementary appendix)⁵.

68 After the outbreak of COVID-19 in Wuhan announced on December 31, 2019, enhanced
69 surveillance was initiated in Guangdong Province to early recognize suspected infections,
70 especially among those who had history of travelling to Wuhan within 14 days. Once a suspected
71 case was diagnosed, the Center for Diseases Control and Prevention (CDC) at prefecture level
72 would conduct an initial field investigation, and collect respiratory specimens. The respiratory
73 specimens were shipped to Guangdong Provincial CDC (GDCDC) for SARS-CoV-2 test by
74 RT-PCR or genetic sequencing.

75 **Epidemiological data Collection**

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

80 **Metagenomic Sequencing**

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

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

90 **Statistical and Sequences analyses**

91 We used descriptive statistics to summarize the epidemiological characteristics of confirmed cases.
92 The sequencing data were quality controlled (QC) using fastp⁶ and then mapped to the human
93 genome (hg38) using Bowtie2⁷ to remove human related RNA reads. De novo assembling was
94 performed using Spades⁸ and Megahit⁹ for each dataset. Assembled contigs were annotated using
95 blastn¹⁰ using the complete nucleotide database. Twelve genome sequences generated in this study
96 have been submitted to GISAID under the accession numbers EPI403932–EPI403937 and
97 EPI406533–EPI406538. Maximum-likelihood (ML) trees of viral genome were estimated via
98 IQ-Tree¹¹ by integrating closely related sequences from SARs and SARs-like viruses from public
99 database. Two SARS-CoV-2 genomes from introduction cases in Thailand were also included and
100 data analysis was approved by the corresponding submitter. SNV (single nucleotide variant) sites
101 were found through SNV-sites¹². Phylogenetic trees were annotated and visualized with ggtree¹³.

102 **Ethics Approval**

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

106

107 **Results**

108 **Description of the outbreak**

109 As of February 10, 2020, a total of 1,151 cases were identified in 20 of 21 cities in Guangdong
110 Province, with the majority of cases in Shenzhen (n=368) and Guangzhou(n=313). Most cases
111 (81.4%, 937/1151) were imported from other Provinces, especially from Hubei Province (70.1%,
112 841/1,151). The peak of total cases was between January 23 and January 28, and so did for cluster
113 cases and secondary cases (Figure 1). The average age of all cases was 44.5 years, and the
114 percentage of males was 49.1% (565/1,151). Of total cases, 579 (50.3%) and 275 (23.9%) cases
115 had history of travelling to Wuhan, and other cities in Hubei Provinces, respectively (Table 1). The
116 average duration from onset of symptoms to diagnosis was 5.6 ± 3.2 days.

117 Of total cases, 697 cases (60.1%) were from 234 cluster infection events with 537 cases in 195
118 family cluster events and 160 cases in community clusters. Figure 2 shows that most cluster
119 infection events are familial clusters during the epidemic, and community cluster events increased
120 since the end of January. More importantly, on February 4, 1 nosocomial cluster event (3 cases)
121 were reported in a hospital. Two hundred and fourteen (18.6%) secondary cases were identified,
122 where 144 cases were from family cluster infections, 56 from other cluster infections, and 14 from
123 community infections.

124 Here we extracted 13 family cluster infections as representatives of all family cluster cases to
125 describe their characteristics, in which the epidemiological information of three family clusters
126 were described in detail below. We also described two special events, one for community infection
127 and another for nosocomial transmission.

128 **Characteristics of 13 family cluster infections**

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

136 **Family cluster A**

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

151 collected and tested by CDC on January 15, and identified positive of SARS-CoV-2.

152 **Family cluster B**

153 Three cases were identified in family cluster B in Zhuhai, including a father (Case 1), a mother

154 (Case 2) and their daughter (Case 3). Case 1 and Case 2 resided in Wuhan, and visited on Case 3

155 in Zhuhai on January 11. On their trip to Zhuhai, Case 1 began to have symptom of fatigue. On

156 January 12, Case 1 developed dry cough, fatigue, headache, eye pain and muscular soreness, and

157 Case 2 had symptoms of dry cough, fatigue, headache. On January 15, Case 1 and Case 2 visited a

158 hospital. Their throat swabs were collected and tested by CDC on January 17, and laboratory

159 diagnosis showed positive of SARS-CoV-2. Case 3 had symptoms of fever, occasional itching and

160 dry cough on January 17 without history of visiting Wuhan within 14 days prior to the onset of

161 symptoms. She visited a hospital on January 18, and SARS-CoV-2 was identified in throat swab

162 samples by CDC.

163 **Family cluster D**

164 Family cluster D in Zhuhai had three cases. Case 1 stayed at Wuhan from January 8 to 17, and

165 returned Zhuhai on January 17. On the trip, he had symptom of muscular soreness. Case 2 (wife of

166 Case 1) lived with Case 1 in Zhuhai and had cough on January 20 without history of visiting

167 Wuhan within 14 days before the onset. She visited a hospital with Case 1 on January 20, and

168 were transferred to another hospital on January 21 for treatment and isolation. Their nasal and

169 throat swabs were tested and identified with positive of SARS-CoV-2 on January 22. Case3

170 (mother of Case 1) lived with Case1 and Case2 without history of visiting Wuhan in the past 14

171 days. Her nasal and throat swabs were also identified with positive of SARS-CoV-2 on January 22,

172 when she had low fever and fatigue, and was isolated on January 23.

173 **Community infection event**

174 Five cases were included in a community transmission event. Case 1 resided in Henan Province,
175 China. On January 11, he and his wife (Case 2) came to Guangzhou to visit on their daughter
176 (Case 3) by train which passed through Wuhan. Case 3 resided on the tenth floor of an apartment
177 in Guangzhou. Case 1 and Case 3 had symptoms of fever on January 19, but they did not visit a
178 hospital. Case 1 had symptoms of fever and dry cough on January 23. Case 1 and Case 2 visited a
179 hospital on January 24 with Case 3, and their throat swabs were tested and identified with positive
180 of SARS-CoV-2. They were isolated in a hospital for treatment. On January 16, Case 1 used the
181 elevator in their apartment. Immediately after he got off the elevator, Case 4, who resided on the
182 eleventh floor, entered the elevator, and touched herself mouth after she pressed the elevator
183 button; she did not wear a mask. Case 4 did not meet Case 1, Case 2 and Case 3 before her onset
184 of illness. On January 21, Case 4 had symptoms of fever, and visited a community hospital. On
185 January 27, Case 4 and her husband (Case 5) visited another hospital, and their throat swabs were
186 tested and identified with positive of SARS-CoV-2. Case 5 had no symptoms before he was
187 diagnosed as COVID-19. Both Case 4 and Case 5 had no exposure to wild animals within 14 days
188 prior to the onset of illness.

189 **Nosocomial infection event**

190 Three COVID-19 cases were included in a hospital-associated transmission event. Case 1, who
191 resided in Guangzhou, and visited her relatives in Hubei Province from January 12 to January 22.
192 She had symptoms of fatigue and leg pain on January 27, and visited a hospital on January 28. She
193 had fever when she was hospitalized. On February 2, she had chest radiologic changes, and her
194 throat swabs were tested and identified with positive of SARS-CoV-2. Case 2 was admitted to the

195 same room in the same hospital with Case 1 on December 31, 2019 because of backache. His
196 white blood counts decreased on February 4, and he had severer cough on February 5 when he
197 was identified with positive of SARS-CoV-2. Case 3 was a nurse who took care of Case 1 for
198 several times, had symptoms of dry cough on February 4, and expectorated on February 5 when
199 she was identified with positive of SARS-CoV-2. She had no exposure to wild animals within 14
200 days prior to the onset of illness, but did not strictly protect herself by using protective products
201 when she nursed Case 1.

202 **Phylogenetic analyses**

203 The phylogeny based on full-length of SARS-CoV-2 genomes (12 generated in this study) showed
204 SARS-CoV-2 could be classified into SARS-like virus with the closest related strains as
205 MG772934 and MG772933 which were collected from Rhinolophus sinicus in China on 2015 and
206 2017, respectively. SARS viruses shared nearly 97.5% nucleotide similarity with closely related
207 SARS-like viruses from bats. By contrast, SARS-CoV-2 had lower nucleotide similarity with
208 related SARS-like viruses (87.4%-87.5%), and a long internal branch was observed between
209 SARS-CoV-2 and related SARS-like viruses (Figure 4A).

210 To characterize the COVID-19 cases from family clusters and investigate potential genetic
211 adaptation, we compared the twelve SARS-CoV-2 genome sequences generated between 14
212 January 2020 and 23 January 2020 in this study and all published SARS-CoV-2^{14,15}. Sequences
213 from the three family cluster infections fell into three corresponding genetic clusters and could be
214 separated from the sequences of sporadic infections (Figure 4B). The viral genome sequences
215 from cases within a family cluster were exactly the same. To find whether there were any genetic
216 changes following the illness progress, we sequenced SARS-CoV-2 viruses from case 1 of cluster

217 B on 17 January and 22 January, which was about 1 week and 2 weeks after the onset of illness,
218 respectively. These two sequences were exactly the same suggesting the genetic stability of
219 SARS-CoV-2 during the infection. Notably, the viral genome from family cluster A share three
220 unique SNV sites which could not be found from other COVID-19 cases in public database.
221 Among the three SNVs, one was predicted to cause amino acid changes (nonsynonymous
222 variations) in ORF8 protein coding sequence led the amino acid change from Leu to Ser. One
223 unique SNV site was identified in family cluster B and the change from C to T at position 21442
224 within the spike protein coding sequence leading to amino acid change from His to Thr. Only one
225 synonymous variation was identified in viral genome sequences from family cluster C. Due to the
226 increasing epidemic activity, we also investigated whether there were any parallel mutations,
227 which was regard as an indicator of potential genetic adaptation, could be identified in recently
228 identified sporadic infection cases. Three SARS-CoV-2 genetic sequences from sporadic cases on
229 22 January and 23 January showed there were no common SNV sites that could be identified in
230 these cases and the SF201_20/01/23 SARS-CoV-2 collected on 23 January had exactly the same
231 genetic sequence with viral sequences collected at the beginning of the epidemic.

232

233 **Discussion**

234 Eighteen years after the first emergence of SARS-CoV in Guangdong at the end of 2002¹⁶, a great
235 challenge presents by the emergence of a new SARS-like coronavirus (SARS-CoV-2). The early
236 epidemiological data suggested that most cases were caused by spill-over infections from
237 nonhuman sources which is still unknown¹⁷. In our study, we described the characteristics of
238 human-to-human transmissions of SARS-CoV-2 using epidemiological and genetic evidence.

239 As of February 10, increasing secondary cases were identified in Guangdong Province. In all 214
240 secondary cases, none of them had history of residence or traveling to Wuhan within 14 days prior
241 to onset of the illness. Our finding confirms the human-to-human transmission of SARS-CoV-2
242 reported in a familial cluster infection in a previous study¹⁸, and further suggests that it has strong
243 transmissibility between humans. We further observed several cases with mild symptoms or no
244 symptoms, which indicates that current monitoring measures such as fever testing may not work
245 effectively to identify these asymptomatic cases. With the epidemic ongoing, we found two cases
246 infected by sharing elevator with an index case in a building, and increasing cases in gathering
247 activities or communities. These findings suggest that the human-to-human transmission risks are
248 transferring from familial cluster infections to community outbreak. More importantly, we
249 observed a nosocomial transmission event in Guangdong Province, which was also reported in a
250 hospital in Wuhan, where 57 patients including 40 health-care workers were infected¹⁹. “Super
251 spreading events” may occur in some specific circumstances such as hospitals if the epidemic was
252 not control on time. As the extended Chinese Spring Festival holiday ends, millions of people will
253 travel back to Guangdong Province from regions across China, which will cause more cases to be

254 imported to the Province. This will further increase the risk of community transmission. Therefore,

255 we should persist in rigorous measures of cutting human-to-human transmission taken by the

256 governments in communities especially in high risk places such as hospitals, factories and schools.

257 The evidence of strong human-to-human transmission raise much concern on 1) whether there are

258 any genetic mutations associated with these cluster infections, and 2) are there any potential viral

259 adaptations leading to the increasing epidemic activity? These questions, which are fundamental

260 to the outbreak control, have not been adequately answered yet. In our study, the viruses

261 sequenced from two secondary cases are the same with the viruses from corresponding index

262 cases in their families, and these secondary cases had no history of traveling to Wuhan and no

263 exposure to wild animals, which indicated that these viruses likely represent strains which can

264 transmit from human to human. We also searched any genetic clues that might be associated with

265 viral adaptation to a human transmission by comparing the viruses from family cluster infections

266 with SARS-CoV-2 from the beginning of epidemic. Only three and one unique SNV sites are

267 identified from viruses sequenced in family infection cluster A and B, respectively. These SNV

268 sites are predicted to result in two synonymous change from Leu to Ser in Orf 8 and from His to

269 Thr in spike protein, respectively. It is still unknown whether these amino acid changes provide or

270 increase the virus capability of human transmission. However, the previous study from SARS

271 suggests Orf8 is either noncoding or coding for a functionally unimportant putative protein²⁰. As a

272 result, few nonsynonymous changes detected between COVID-19 cases from secondary cases in

273 Guangdong and early COVID-19 cases in Wuhan indicates SARS-CoV-2 could already transmit

274 among humans at the beginning of the epidemic when it was still regarded as spill-over infections.

275 In this circumstance, more human to human transmissions should occur at the end of December
276 2019 than we previously expected. Another reason that we believe SARS-CoV-2 may be already
277 well established in humans since its first identification is that the SARS-CoV-2 we identified on
278 January, 23, 2020 were exactly the same with viruses identified in Wuhan on 30 December 2019.
279 Together with the SARS-CoV-2 genome sequences generated on 15 and 22 January, we did not
280 find any clues of adaptations within the genome of SARS-CoV-2 accompany with the increasing
281 epidemic. These data suggest the genetic stability of SARS-CoV-2 during the epidemic, which is a
282 contrast to our previous findings on SARS viruses for which the strong positive selection on spike
283 protein coding sequence and deletions in ORF8 gene are observed at the early phase of the
284 outbreak²⁰.

285 Our findings indicate that with the epidemic ongoing, prevention measures of containing
286 person-to-person transmission in communities are urgently enhanced. Persistent surveillance for
287 COVID-19 in communities, early detection and isolation of contacts of confirmed cases, and
288 virologic analyses to assess genetic changes that might suggest increased transmissibility among
289 humans are all critical to inform prevention and control efforts and assessing the global pandemic
290 potential of COVID-19.

291

292 **Acknowledgements**

293 We thank all the medical and nursing staff who assisted in the care of patients; the members from
294 health department and CDC in Guangdong Province for their contribution in data collection,
295 COVID-19 control and prevention; Dr. Pilailuk Okada and his colleagues for sharing their
296 sequence data. This work was supported by National Key Research and Development Program of
297 China (2018YFA0606200, 2018YFA0606202), the Science and Technology Program of
298 Guangdong Province (2018B020207006, 2019B020208005, 2019B111103001), Guangzhou
299 Science and technology Plan Project (201804010383).

300

301 References

- 302 1. Wuhan Municipal Health Commission, Hubei Province, China. Wuhan Municipal Health
303 Commission briefing on the pneumonia epidemic situation, 31 December 2019. 2019.
304 (<http://wjw.wuhan.gov.cn/front/web/showDetail/2019123108989.>)
- 305 2. The 2019-nCoV Outbreak Joint Field Epidemiology Investigation Team, Li Q. An Outbreak
306 of NCIP (2019-nCoV) Infection in China — Wuhan, Hubei Province. China CDC Weekly
307 2020;2:79-80.
- 308 3. World Health Organization. WHO Statement Regarding Cluster of Pneumonia Cases in
309 Wuhan, China. 2020. (<https://www.who.int/china/>.)
- 310 4. World Health Organization. Coronavirus. 2020.
311 (<https://www.who.int/health-topics/coronavirus.>)
- 312 5. National Health Commission of the People's Republic of China. Prevention and Control
313 Scheme of Novel Coronavirus. 2020.
314 (<http://www.nhc.gov.cn/xcs/zchengcwi/202001/808bbf75e5ce415aa19f74c78ddc653f.shtml.>)
- 315 6. Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor.
316 Bioinformatics 2018;34:i884-i90.
- 317 7. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods
318 2012;9:357-9.
- 319 8. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its
320 applications to single-cell sequencing. J Comput Biol 2012;19:455-77.
- 321 9. Li D, Liu CM, Luo R, Sadakane K, Lam TW. MEGAHIT: an ultra-fast single-node solution
322 for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics

- 323 2015;31:1674-6.
- 324 10. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. BMC
325 Bioinformatics 2009;10:421.
- 326 11. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: A Fast and Effective
327 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. Mol Biol Evol
328 2014;32:268-74.
- 329 12. Page AJ, Taylor B, Delaney AJ, et al. SNP-sites: rapid efficient extraction of SNPs from
330 multi-FASTA alignments. Microb Genom 2016;2:e000056.
- 331 13. Yu G, Smith DK, Zhu H, Guan Y, Lam T-T, et al. GGTREE: an r package for visualization
332 and annotation of phylogenetic trees with their covariates and other associated data. Methods
333 Ecol Evol 2017;8:28–36.
- 334 14. Zhou P, Yang X-L, Wang X-G, et al. A pneumonia outbreak associated with a new
335 coronavirus of probable bat origin. Nature 2020 (in press). Doi: 10.1038/s41586-020-2012-7 .
- 336 15. Zhu N, Zhang D, Wang W, et al. A Novel Coronavirus from Patients with Pneumonia in
337 China, 2019. N Engl J Med 2020. Doi: 10.1056/NEJMoa2001017.
- 338 16. Drosten C, Günther S, Preiser W, et al. Identification of a novel coronavirus in patients with
339 severe acute respiratory syndrome. N Engl J Med 2003;348:1967-76.
- 340 17. Wu F, Zhao S, Yu B, et al. A new coronavirus associated with human respiratory disease in
341 China. Nature 2020 (in press). Doi: 10.1038/s41586-020-2008-3
- 342 18. Chan JF-W, Yuan S, Kok K-H, et al. A familial cluster of pneumonia associated with the 2019
343 novel coronavirus indicating person-to-person transmission: a study of a family cluster. The
344 Lancet 2020 (in press). Doi: 10.1016/S0140-6736(20)30154-9
- 345 19. Wang D, Hu B, Hu C, et al. Clinical characteristics of 138 hospitalized patients with 2019
346 novel coronavirus–infected pneumonia in Wuhan, China. JAMA 2020 (in press). Doi:

347 10.1001/jama.2020.1585

348 20. He JF, Peng GW, Min J, et al. Molecular Evolution of the SARS Coronavirus, during the

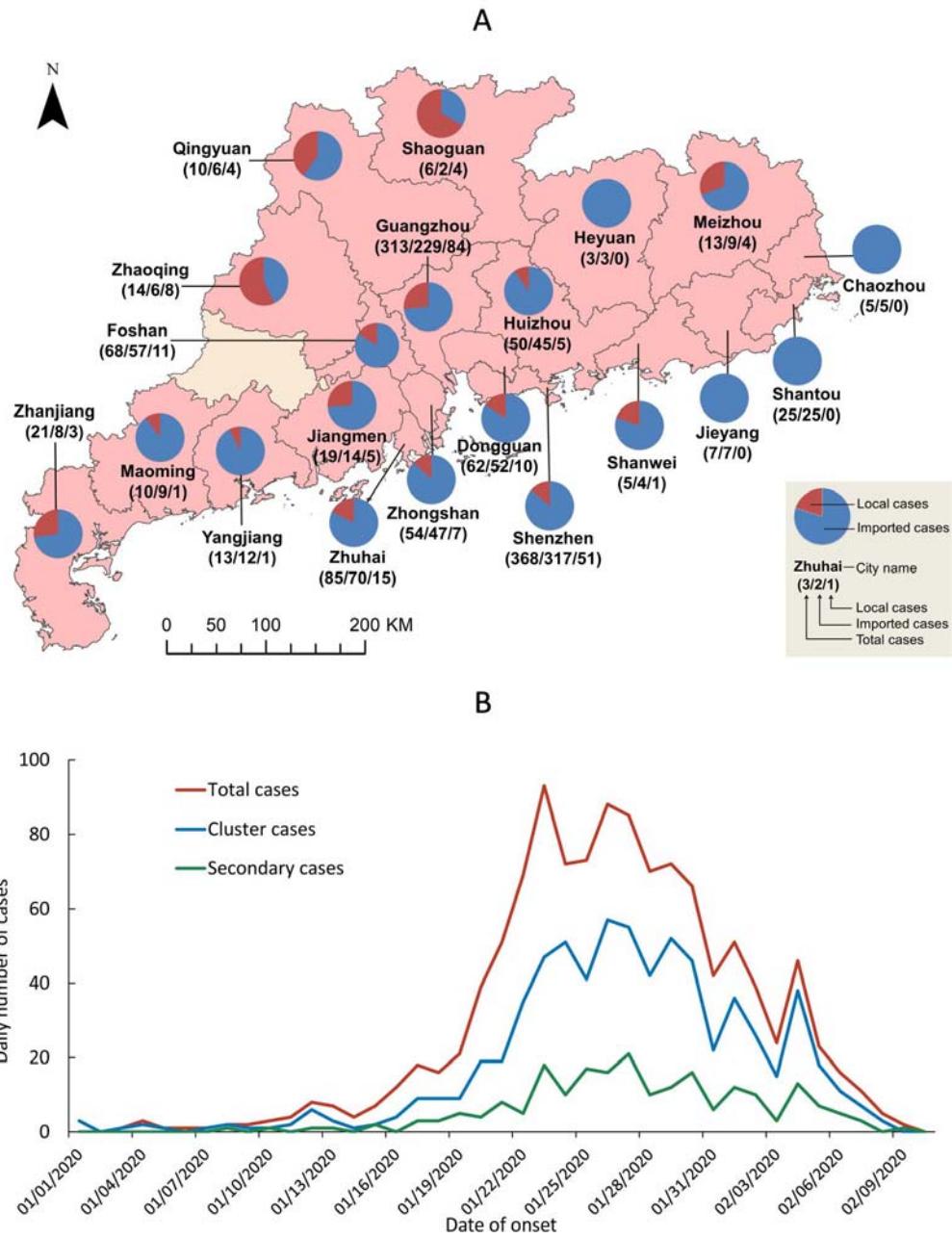
349 Course of the SARS Epidemic in China. Science 2004;303:1666-9.

350

351 Table 1. General characteristics of confirmed COVID-19 cases in Guangdong Province, China

| | Total (n, %) | Cluster cases (n, %) | Non-cluster cases (n, %) | χ^2 | P |
|---|--------------|-------------------------|--------------------------------|----------|--------|
| Sex | | | | | |
| Male | 565 (49.1) | 334 (47.9) | 231 (50.9) | 0.850 | 0.357 |
| Female | 586 (50.9) | 363 (52.1) | 223 (49.1) | | |
| Age (years) | | | | | |
| <30 | 225 (19.5) | 141 (20.2) | 84 (18.5) | | |
| 30~ | 258 (22.4) | 151 (21.7) | 107 (23.6) | | |
| 40~ | 176 (15.3) | 99 (14.2) | 77 (17) | 9.295 | 0.098 |
| 50~ | 216 (18.8) | 121 (17.4) | 95 (20.9) | | |
| 60~ | 209 (18.2) | 138 (19.8) | 71 (15.6) | | |
| ≥70 | 67 (5.8) | 47 (6.7) | 20 (4.4) | | |
| History of travelling to Wuhan and other cities in Hubei | | | | | |
| Wuhan | 579 (50.3) | 320 (45.9) | 259 (57) | 33.66 | <0.001 |
| Other cities in Hubei | 275 (23.9) | 160 (23.0) | 115 (25.3) | | |
| No | 294 (25.5) | 217 (31.1) | 77 (17.0) | | |
| Unclear | 3 (0.3) | 0 (0.0) | 3 (0.7) | | |
| Duration from onset to diagnosis (days) | | | | | |
| <2 | 101 (8.8) | 81 (11.6) | 20 (4.4) | | |
| 2~ | 231 (20.1) | 160 (23) | 71 (15.6) | 43.26 | <0.001 |
| 4~ | 268 (23.3) | 157 (22.5) | 111 (24.4) | | |
| 6~ | 212 (18.4) | 132 (18.9) | 80 (17.6) | | |
| ≥8 | 290 (25.2) | 140 (20.1) | 150 (33) | | |
| Unclear | 49 (4.3) | 27 (3.9) | 22 (4.8) | | |
| Severity of illness | | | | | |
| Mild | 993 (86.3) | 604 (86.7) | 392 (86.3) | 0.46 | 0.792 |
| Serious | 113 (9.8) | 66 (9.5) | 47 (10.4) | | |
| Critical | 42 (3.6) | 27 (3.9) | 15 (3.3) | | |

352

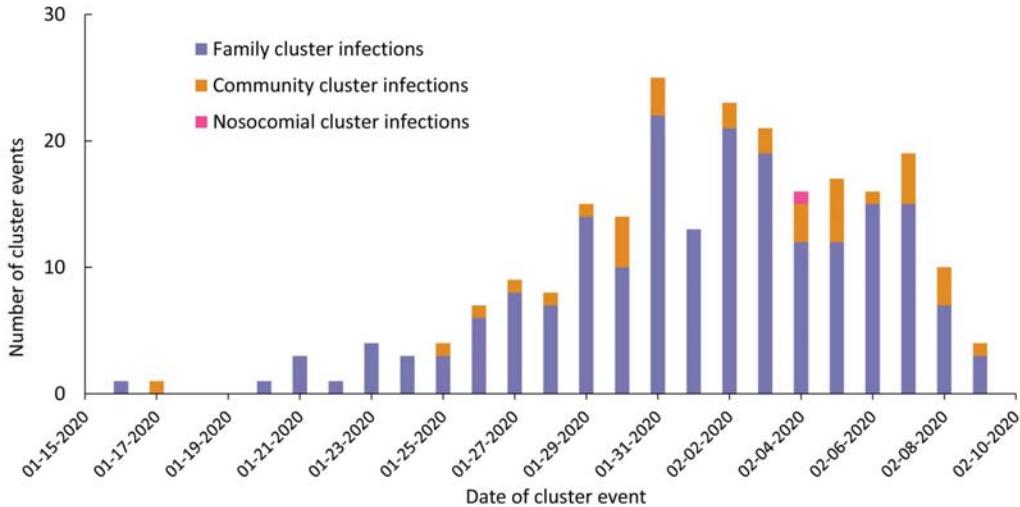


353

354 Figure 1. Spatio-temporal distribution of cases in Guangdong Province, China, by February 10,
355 2020

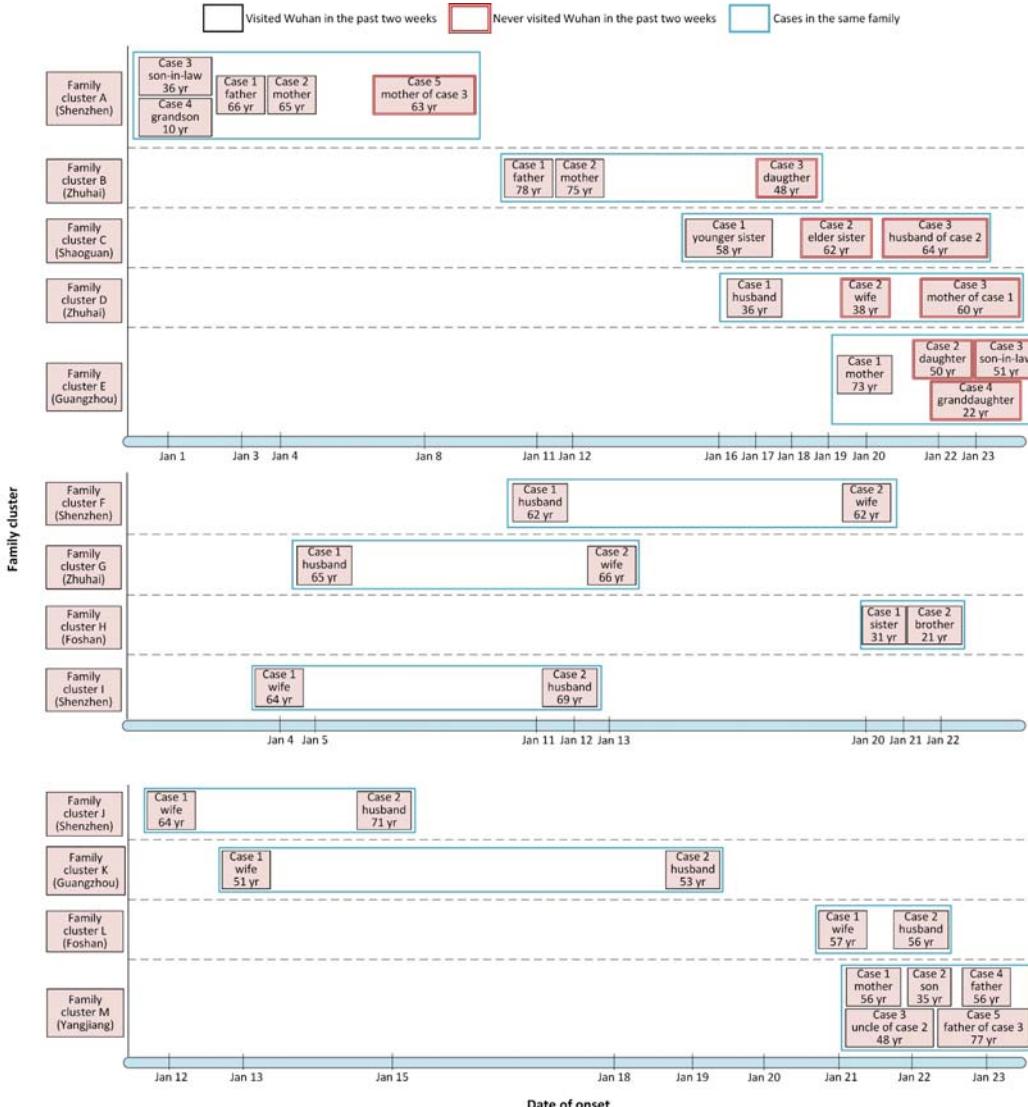
356 **Chart A:** The spatial distribution of all confirmed cases.

357 **Chart B:** The temporal distribution three types of cases, and number of cluster infections.



358

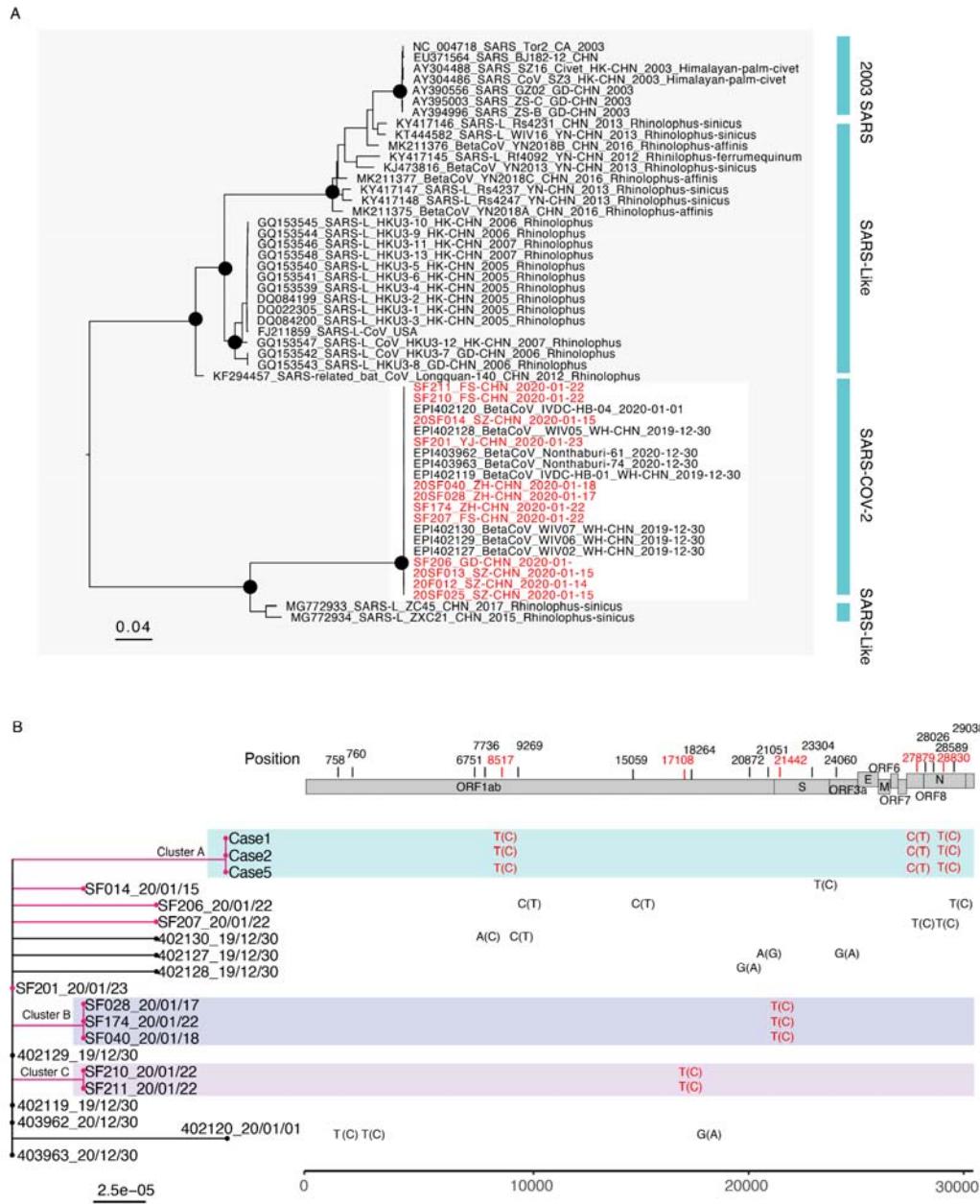
359 Figure 2. Time distribution of cluster infection events in Guangdong Province, by cluster types



360

361 Figure 3. Onset date of 37 confirmed COVID-19 cases in 13 family cluster infections in

362 Guangdong Province



365 Figure 4. Phylogenetic analysis of the complete genome of SARS-CoV-2 identified in Guangdong
366 and closely related viral genome from the public database.
367 **Chart A:** Maximum-likelihood tree was constructed by including 12 new generated SARS-CoV-2
368 sequences in this study, 6 published SARS-CoV-2 sequences and sequences from closely related
369 SARS-like viruses. Two SARS-CoV-2 sequences from Thailand were integrated into the analysis
370 with the permission of the submitter. Black circles indicate bootstrap support >0.9 at the root node.

371 of selected clade. The cluster of SARS-CoV-2 was highlighted with red box. The sequences
372 generated in this study were highlighted in red.

373 **Chart B:** Phylogenetic relationship among SARS-CoV-2 and SNV (single nucleotide variant)
374 sites were identified by aligning twenty complete genomes of SARS-CoV-2 collected between 30
375 December, 2019 and 23 January, 2020 were found by the multiple genome sequences alignment.
376 To ignore the sequences uncertainty at 5' and 3' terminal, the aligned sequences were started from
377 position 266 to 29856 of SARS-CoV-2 genome (use the EPI402119 as the reference genome). The
378 phylogeny branches of Guangdong SARS-CoV-2 sequences were colored in magenta and different
379 family clusters were marked with different color boxes. Unique SNVs and their corresponding
380 positions in reference viral genome identified in family cluster infections were shown on right of
381 the phylogeny and the relative consensus sequences were shown in brackets. The nonsynonymous
382 SNVs in cluster infections were marked with #.