

---

1           **Genome-wide variations of SARS-CoV-2 infer evolution**  
2                           **relationship and transmission route**

3           Lehai Zhang<sup>1,2#</sup>, Shifu Wang<sup>1#</sup>, Qian Ren<sup>1</sup>, Junjie Yang<sup>3</sup>, Yanqin Lu<sup>4</sup>, Lei Zhang<sup>5,6,\*</sup>, Zhongtao Gai<sup>2,\*</sup>

4

5           1. Department of Children's Medical Laboratory Diagnosis Center, Qilu Children's Hospital of  
6           Shandong University, Jinan 250022, China

7           2. Jinan Institute of Pediatrics & Shandong Children's Microbiome Center, Qilu Children's  
8           Hospital of Shandong University, Jinan, Shandong, 250022, China

9           3. College of Life Science, Qilu Normal University, Jinan, Shandong, 250200, China

10          4. Key Laboratory for Biotech-Drugs of National Health Commission, Key Laboratory for Rare  
11          & Uncommon Diseases of Shandong Province, Shandong Medicinal Biotechnology Centre,  
12          Shandong First Medical University & Shandong Academy of Medical Sciences, Jinan  
13          250062, China

14          5. Microbiome-X, Beijing Advanced Innovation Center for Big Data-Based Precision Medicine,  
15          School of Medicine and Engineering, Beihang University, & Key Laboratory of Big  
16          Data-Based Precision Medicine (Beihang University), the Ministry of Industry and  
17          Information Technology of the People's Republic of China, Beijing, 100191, China

18          6. Shandong Institute of Industrial Technology for Health Sciences and Precision Medicine,  
19          Jinan, Shandong, 250101, China

20

21          # These authors contributed equally to this work.

22          \* Zhongtao Gai and Lei Zhang are Co-corresponding authors. Correspondence should be addressed to  
23          Zhongtao Gai, Jinan Institute of Pediatrics, Qilu Children's Hospital of Shandong University, Jinan,  
24          Shandong, 250022, E-mail: [gaizhongtao@sina.com](mailto:gaizhongtao@sina.com).

---

25 **Abstract**

26 In the epidemic evolution of severe acute respiratory syndrome coronavirus 2  
27 (SARS-CoV-2), the issues of mutation, origin, typing and the effect of mutation on  
28 molecular detection remain to be unrevealed. In order to identify the evolutionary  
29 relationship of SARS-CoV-2 and evaluate the detection efficiency of primers that are  
30 currently used in different countries, we retrieved genomic sequences of 373  
31 SARS-CoV-2 strains from multiple databases and performed genome-wide variation  
32 analysis. According to the nucleotide C28144T variation, the SARS-CoV-2 can be  
33 divided into group A (117 strains) and group B (256 strains). The spike protein gene  
34 (S gene) coding region 1841 (total 23403) A1841G, formed a B1 subgroup (40 strains)  
35 in group B, of which 30 strains were from European and American countries in March  
36 (especially Washington, USA). These mutations are likely to be influenced by the  
37 environment or the immunization selection pressure of different populations.  
38 Although the mutation is not in the receptor binding region (RBD) and alkaline  
39 cleavage region, it may also affect the ability of transmission and pathogenicity;  
40 however, the significance is not yet clear. As the ratio of A / B strains in the epidemic  
41 months showed an increasing trend (0.35: 1 in January, 0.62: 1 in February and 0.76:  
42 1 in March), it seems that the transmissibility of group A strains becomes stronger  
43 with time. Based on the variation of 11 nucleotide sites during the epidemic process, it  
44 is speculated that the Washington strain is more like an ancestor type, and the Wuhan  
45 strain is the offspring of the group A virus strain. By comparing the detection  
46 capabilities of primers in different countries, the SARS-CoV-2 nucleotide variation  
47 may only affect molecular detection of very few strains. The differences in the  
48 transmissibility, pathogenicity and clinical manifestations of different types of strains  
49 require further investigations.

50

51 **Key words:** SARS-CoV-2; Molecular typing; Source of infection; Mutation  
52 transmission route

53

54

---

55 **Introduction:**

56 The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) started to  
57 emerge in Wuhan in December 2019. As of April 20, 2020, more than 2.4  
58 million cases have been reported in more than 211 countries and regions.  
59 While the outbreak and pandemic of SARS-CoV-2 poses a great threat to public  
60 health, in-depth studies have been conducted on key issues such as  
61 transmission ability, epidemiology, clinical characteristics, pathogenic  
62 mechanism, clinical treatment, and laboratory testing. Regarding the origin of  
63 SARS-CoV-2, bat coronavirus RaTG13 was found to be highly similar to  
64 SARS-CoV-2, so it was speculated that bats might be the original host of the  
65 virus <sup>[1]</sup>. It was further found that the virus isolated from pangolin was similar  
66 to SARS-CoV-2, suggesting that pangolin might be an intermediate host <sup>[2, 3]</sup>.  
67 By comparing the genomes of existing coronavirus such as SARS-CoV-2,  
68 MERS-CoV and SARS-CoV, it was found that RNA editing in the human body  
69 is related to certain mutations obtained by virus evolution <sup>[4]</sup>. In order to  
70 reveal more clues of the origin of SARS-CoV-2, which is still an urgent problem to  
71 be solved, we retrieved the publicly available virus sequences from global  
72 database sources such as GISAID and NCBI to analyze the emerged mutation,  
73 evolution and typing of SARS-CoV-2, and to evaluate the influence of its  
74 mutations on molecular detection in current clinical settings.

75

76 **Materials and methods**

77

78 By February 25, 2020, our first retrieval downloaded a total of 188  
79 complete genomes from GISAID, NCBI, the Global Coronavirus Data Sharing  
80 and Analysis System of the National Microbiology Data Center of China  
81 (<http://nmdc.cn/coronavirus>) and the 2019 Novel Coronavirus Resource  
82 (2019nCoV) of the China National Center for Bioinformation  
83 ([https://bigd.big.ac.cn/ncov/release\\_genome?lang=en](https://bigd.big.ac.cn/ncov/release_genome?lang=en)). BioEdit was used to  
84 compare the genome sequence and remove those with ambiguous loci. After  
85 continued retrieval, 373 full genomes were finally obtained by March 31, 2020,  
86 and used to construct the phylogenetic tree by MEGA-X and FigTreev1.4.3.  
87 The variation percentage and pairwise distance were calculated using  
88 MEGA-X32. BioEdit was used to calculate the number of mutation sites, and

---

89 DNAsp was used to calculate the nonsynonymous mutation rate and  
90 synonymous mutation rate.

91

## 92 **Results**

93

### 94 **1. Mutation and evolution of SARS-CoV-2**

95

96 The total number of nucleotides was 29,838bp after alignment of 373 full  
97 genome sequences and the deleted sequences carried no mutations.  
98 Representative strains, such as USA/WA1-F6, Chongqing/YC01, Fujian/8,  
99 MN908947.3, MN938384.1, WA-UW82, were selected from each cluster of  
100 the phylogenetic tree to analyze the mutation rate of nucleic acids, the  
101 nonsynonymous mutation rate and synonymous mutation rate, the amino acids  
102 variations of nonsynonymous mutations, and their evolutionary relationship  
103 (Table 1, 2 and 3).

104 Based on the position and nucleotide variation of the phylogenetic tree,  
105 we compared the evolutionary variation between SARS-CoV-2 representative  
106 strains. WA1-F6 (sample collected on January 25, 2020) strain has one  
107 nucleotide variation compared with the closed bat strain (bat/Yunnan/RaTG13).  
108 Compared with WA1-F6 strain, we find that Chongqing/YC01, Fujian/8 and  
109 HKU-SZ-002a strain (in the same evolutionary branch with WA1-F6 strain)  
110 showed 2, 1 and 2 nucleotide variation respectively; While Wuhan epidemic  
111 strain Wuhan-Hu-1 (sample collected in November 2019), which is far from  
112 the outer group, has 3 nucleotide variation; and WA-UW82, which is far from  
113 the outer group, has 8 nucleotide variation (sample collected in March 2020).  
114 Interestingly, the WA-UW82 strain appeared one-and-a-half month after the  
115 WA1-F6 strain became popular in the Washington area of the United States.  
116 The WA-UW82 strain mutated and evolved from WA1-F6 in the 5'UTR,  
117 ORF1ab, S, ORF3a and ORF8 coding regions, whereas no positive selection  
118 was found (Table 3).

119

### 120 **2. Non-synonymous mutation and phylogenetic clustering**

---

121 Compared with bat/Yunnan/RaTG13, Wuhan-Hu-1 has a  
122 non-synonymous mutation (S84L) at position 251 (28,144 in total) in the  
123 coding region of ORF8, which is speculated to be closely related to its  
124 evolutionary position. The nucleotides and amino acids in the ORF8 region of  
125 the sequence were analyzed, and the positions of different mutated strains in  
126 the evolutionary tree were compared in Figure 1. It was found that nucleotides  
127 T251C (28144 in total) or 197 strains of amino acid L84S formed two clusters  
128 in the phylogenetic tree. To test this assumption, all 373 SARS-CoV-2 genome  
129 sequences were integrated into the phylogenetic tree in batches, and the results  
130 remained consistent. According to the nucleotide C28144T variation,  
131 SARS-CoV-2 can be clustered into two large groups. The group A virus has a C  
132 at position 28,144, whereas group B virus has a T. Among the 39 strains of the  
133 group B virus, the A1841G non-synonymous mutation (D614G) occurred in  
134 the 1841 position (23403 positions) of the S gene coding region during the  
135 SARS-CoV-2 pandemic, forming a B1 subgroup (Supplementary Figure 1 and  
136 Table 1).

137 Compared with the other 5 SARS-CoV-2 representative strains, the  
138 non-synonymous mutation of A1841G (D614G) occurred in the S gene coding  
139 region 1841 (total 23403), but no mutation occurred in the receptor binding  
140 region (RBD, S protein amino acid 451-509) and alkaline cleavage region  
141 (667-694) in WA-UW82. Further analysis of the S-coding regions of 373  
142 strains of SARS-CoV-2 genome collected from December 2019 to March 31,  
143 2020 revealed that 39 strains had mutations of D614G, and 334 strains had no  
144 mutations. Of the 39 strains, only BetaCoV/Germany/BavPat1/2020 was  
145 mutated at the end of January, 10 strains were mutated in February, and the  
146 remaining 28 strains were mutated in March. Except for the  
147 Wuhan/HBCDC-HB-06/2020 strain from Asia, the vast majority of mutations  
148 were from European and American countries (Figure1).

149

### 150 3. The influence of the variation to the nucleic acid detection

151 In order to evaluate the impact of nucleotide variation on clinical test, we  
152 compared the 373 SARS-CoV-2 genome sequences with currently used  
153 primers that were reported by the Chinese Center for Disease Control and  
154 Prevention (CDC)<sup>[5]</sup>, Europe and Hong Kong<sup>[6]</sup> (Table 4).

---

155 The primer sequences reported by China CDC and Europe did not find  
156 nucleotide variation in the ORF1ab and RdRp amplification regions,  
157 respectively, while several nucleotides in the N amplification region mutated,  
158 including 10 strains from Mexico, Netherlands, Sweden, Finland, Portugal and  
159 Peru. The nucleotides corresponding to the 5' initiation segment of the forward  
160 primer of N region have all been mutated from GGG to AAC (Table 4), but  
161 this variation may not lead to false-negative results in PCR detection. Whereas  
162 the nucleotide mutations have occurred in the N or E amplification region near  
163 the 3' end of the European primers, that may affect the amplification results.  
164 Although European primers use degenerate primers at the RdRp amplification  
165 region, fortunately there is no mutation observed yet. The difference between  
166 German and European primers is that the former amplifies RdRP, E, and N  
167 region, and the latter amplifies only RdRP and E region. The RdRP reverse  
168 primer "S" is designed according to the base C of SARS-CoV-2 and the base G  
169 of the bat SARS-related coronavirus, while no mutation of base T has been  
170 found at this location in SARS-CoV-2. Therefore, the use of Chinese CDC and  
171 European primer sequence in PCR detection will not be affected by this  
172 mutation, which can ensure the accuracy and quality of clinical PCR screening  
173 of COVID - 19.

174

## 175 **Discussion**

176 Since the outbreak of COVID-19 in Wuhan in December 2019, it has  
177 infected more than 1.7 million people in more than 200 countries and regions  
178 worldwide, resulting in 160,000 deaths (as of 2020-4-20 9:30). The origin of  
179 SARS-CoV-2 has been described by international authorities <sup>[7, 8]</sup>, and four  
180 distinct characteristics of the virus have been found. There have also been  
181 in-depth studies on the evolution of variation <sup>[9, 10]</sup>, which however only  
182 studied the variation and evolution of SARS-CoV-2 strain in the early stage of  
183 the pandemic.

184 In this study, 373 SARS-CoV-2 genome sequences were downloaded from  
185 GISAID, NCBI and the Global Coronavirus Data Sharing and Analysis System  
186 of the National Microbiology Data Center of China from the end of 2019 to  
187 March 31, 2020. More than 188 virus sequences from different epidemic times

188 and countries and regions were selected to construct the phylogenetic tree.  
189 Several distinct clusters were shown in the phylogenetic tree, and the  
190 evolutionary relationships of each cluster were analyzed. It was found that the  
191 strains of bat/Yunnan/RaTG13 clustered at the lower part of the phylogenetic  
192 tree had non-synonymous mutations (T251C, S84L) at 251 (28,144 in total),  
193 and those without mutations were in other clusters. More importantly, there  
194 were such mutations in the further analysis of 373 strains <sup>[9]</sup>. Accordingly,  
195 popular strain can be divided into group A (117 strains) and group B (256  
196 strains) clusters. The A/B ratio showed an interesting trend of increase of  
197 strains (January 0.35:1, February 0.62:1 and March 0.76:1), indicating that the  
198 transmissibility of class A strain got stronger over time. However, the  
199 differences in the transmission ability, pathogenic ability and clinical  
200 performance of the two types of strains have not been reported, and it is  
201 unknown whether there is a difference in the mortality caused by them. It is  
202 worthy of further follow-up studies by molecular epidemiology experts and  
203 clinical experts.

204 Singapore <sup>[11]</sup> reported that the regionally popular strains spanned ORF7b  
205 and ORF8 mutations, including ORF8 transcriptional regulatory sequence  
206 (TRS), which eliminated ORF8 transcription, thereby enhancing the  
207 transcription efficiency of downstream N-terminal genes. The ORF8 region is  
208 considered to be an evolutionary hotspot of SARS related coronaviruses  
209 (SARSr-CoVs) and an important region for the evolution of SARSr-CoVs. The  
210 ORF8 region has been shown to play an important role in the virus replication  
211 efficiency of SARS-CoV, interspecies transmission, and adaptation to human  
212 hosts. It is speculated that it may be the result of the adaptation of  
213 SARS-CoV-2 to the human host, and it has undergone reorganization selection  
214 during the transmission from animal to human <sup>[12]</sup>. In the SARS-CoV-2  
215 pandemic, 39 strains of the same group B (group B1) had A1841G  
216 non-synonymous mutation (D614G) in the 1,841 position (23,403 positions in  
217 total) in the S gene coding region. Of these, 30 strains were isolated from  
218 European and American countries in March (especially Washington, USA),  
219 whereas in comparison, only 1 strain was isolated in January and 10 strains  
220 were isolated in February.

221 At the end of the writing this article, we downloaded 62 new sequences

222 that were deposited in April into the GISAID database. After deleting the  
223 low-quality sequences, 45 were retained for analysis (data not shown).  
224 According to the above typing method (C28144T, A23403G), except for  
225 hCoV-19/England/20144018704 and hCoV-19/ Belgium/ULG-10095/2020  
226 (Group B), the other 43 strains are all divided into B1 subgroup. With the  
227 exception of 3 strains from South Africa and 1 from Georgia, the remaining 39  
228 strains were all from Europe and the United States. This may be mutated by  
229 the selection pressure of the environment or the immunity of different  
230 population. Although the mutation is not in the RBD and alkaline lysis zone, it  
231 may also affect the transmission and pathogenic ability, but its significance is  
232 not yet clear. These results suggest that the virus may have undergone new  
233 mutations under the pressure of environmental selection, forming a new  
234 epidemic strain. The difference in the ability of the mutation to spread,  
235 pathogenicity, clinical manifestations and outcome of the strain needs to be  
236 determined by molecular epidemiologists and clinicians in further follow-up  
237 studies.

238 Of the first set of 41 infected patients in Wuhan, 27 were associated with  
239 the South China Seafood Market of Wuhan city, and the infected subjects  
240 spread them to others outside the market, however, this conclusion has been  
241 controversial <sup>[13]</sup>. In this study, the strains that have a close phylogenetic  
242 relationship with the outer group bat/Yunnan/RaTG13 are all group A strains.  
243 There is no obvious difference in the time of initial epidemic between group A  
244 and B strains. The original virus strains belonging to group A, such as  
245 Chongqing (Chongqing/YC01), Fujian (Fujian/8), Shenzhen (HKU-SZ-002a),  
246 and Washington (USA/WA1-F6), etc., are all phylogenetically far away from  
247 the group B strain that were popular in Wuhan. Four family members from  
248 Shenzhen (Wuhan-Hu-1) visited Wuhan from December 29, 2019 to January 4,  
249 2020 <sup>[14]</sup>, and the first confirmed patient in the United States (USA / WA1-F6)  
250 visited Wuhan on January 15, 2020 and returned to Washington State <sup>[15]</sup>, and  
251 the Chongqing (Chongqing/YC01) patients returned to their hometowns after  
252 working in Wuhan. Thus, it is speculated that there are at least two or more  
253 original infection sources, or that two types of strains are spreading  
254 simultaneously in a certain infection source.

255 By analyzing the 11 nucleotide sites that can be mutated in the epidemic

---

256 process, we found that the first patient strain (WA1) in Washington, USA, was  
257 mutated by 1 nucleotide based on the bat strain bat/Yunnan/RaTG13, the  
258 Chongqing strain Chongqing/YC01 and Shenzhen strain HKU-SZ-002a  
259 mutated 2 nucleotides at different positions respectively, while the American  
260 WA-UW82 mutated 9 nucleotides. It can be considered that the Washington  
261 strain of the United States is more like an ancestor type, although its onset  
262 time is later than Wuhan patients. The epidemic strain in Wuhan should be a  
263 variant of group A strains. We thus speculated that the group A strains were  
264 already prevalent before the outbreak in Wuhan at the end of December 2019.

265 Through analysis of 10 strains from Mexico, the Netherlands, Sweden,  
266 Finland, Portugal, and Peru, it was found that the nucleotides corresponding to  
267 the 5' initiation segment of the N region forward primer were all mutated from  
268 GGG to AAC. This mutation may cause false negative results in PCR  
269 amplification detection, leading to the missed diagnosis and the spread of the  
270 virus. However, the primer sequences provided by the Chinese CDC and  
271 Europe did not find nucleotide variations in the amplified regions of ORF1ab  
272 and RdRp. Therefore, the quality of screening COVID -19 by clinical PCR  
273 could be guaranteed by using the detection kits that include Chinese CDC and  
274 European primer.

275 In conclusion, the global spread of SARS-CoV-2 has not been effectively  
276 controlled at present, . Group B should be the mutated strain of group A, and  
277 the group A strain is easier to spread. We should pay close attention to the  
278 variation and significance in the epidemic process, especially the  
279 pathogenicity, transmissibility, clinical process and prognosis of different  
280 groups of strains are not clear. Its origin is still a major question to be  
281 explored.

282

283 **Acknowledgments:** This study was supported by Shandong Province Science  
284 Foundation for Talent Cultivation [ZR2015PH059], the outstanding young  
285 technical backbone training project of Jinan [Shi-fu Wang], the State Major  
286 Infectious Disease Research Program [2012ZX10004206-007 and  
287 2017zx10103004-007]. Lei Zhang is supported by the TaiShan Scholars  
288 Program of Shandong Province (No. tshw20120206) and TaiShan Industrial

289 Experts Program (No. tscy20190612).ZT.G. and L.Z. design the research; LH.Z.  
290 and SF.W. performed research; LH.Z. and J.J.Y. analyzed data; LH.Z. Q.R.  
291 and Y.Q.L.performed statistical analyses; LH.Z., SF.W.and L.Z. wrote the  
292 manuscript.

293 ***Declaration of interest statement:*** All authors declare that there is no conflict of  
294 interest.This research plan was approved by the Medical Ethics Committee of the  
295 Qilu Children’s Hospital of Shandong University.

296

## 297 **References :**

- 298 [1]Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat  
299 origin[J/OL].Nature,[2020-03-23].<https://www.nature.com/articles/s41586-020-2012-7>.DOI:10.1038/s41586-020-2012-7.
- 300 [2]Tommy Tsan-Yuk Lam, Marcus Ho-Hin Shum, Hua-Chen Zhudentification , Identification of 2019-nCoV related  
301 coronaviruses in Malayan pangolins in southern China. [J/OL]. BioRxiv, [2020-03-23]. [https://www.biorxiv.org/content/](https://www.biorxiv.org/content/10.1101/2020.02.13.945485v1)  
302 [10.1101/2020.02.13.945485v1](https://www.biorxiv.org/content/10.1101/2020.02.13.945485v1). DOI: 10.1101/2020.02.13.945485.
- 303 [3]Kangpeng Xiao, Junqiong Zhai, Yaoyu Feng, et al. Isolation and Characterization of 2019-nCoV-like Coronavirus from  
304 Malayan Pangolins.[J/OL].BioRxiv, [2020-03-23]. <https://www.biorxiv.org/content/10.1101/2020.02.17.951335v1> .  
305 DOI:10.1101/2020.02.17.951335
- 306 [4]Salvatore Di Giorgio, Filippo Martignano, Maria Gabriella Torcia, et al.Evidence for RNA editing in the transcriptome of  
307 2019 Novel Coronavirus. [J/OL]. BioRxiv, [2020-03-23]. <https://www.biorxiv.org/content/10.1101/2020.03.02.973255v1>. DOI:  
308 10.1101/2020.03.02.973255
- 309 [5] [http://ivdc.chinacdc.cn/kjz/202001/t20200121\\_211337.html](http://ivdc.chinacdc.cn/kjz/202001/t20200121_211337.html)
- 310 [6]Corman V M,Landt O,Kaiser M, et al. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR.  
311 [J/OL].eurosurveillance,[2020-03-26].<https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045>.DOI:  
312 10.2807/1560-7917.ES.2020.25.3.2000045.
- 313 [7]Andersen1 K G, Rambaut A,Lipkin W I, et al. The proximal origin of SARS-CoV-2. [J/OL]. Nature,[2020-04-8].  
314 <https://doi.org/10.1038/s41591-020-0820-9>.DOI: 10.1038/s41591-020-0820-9.
- 315 [8] Boni1 MF, Lemey P, Xiaowei Jiang XW et al.Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for  
316 the COVID-19 pandemic. [J/OL]. bioRxiv,[2020-04-8]. <https://www.biorxiv.org/content/10.1101/2020.03.30.015008v1>. doi:  
317 10.1101/2020.03.30.015008
- 318 [9]Tang XL, Wu CC, Li X. et al.On the origin and continuing evolution of SARS-CoV-2. [J/OL].National Science Review.  
319 [2020-04-8]. <https://doi.org/10.1093/nsr/nwaa036>. doi: 10.1093/nsr/nwaa036
- 320 [10]Yu WB, Tang GD, Zhang L, et al. Decoding the evolution and transmissions of the novel pneumonia  
321 coronavirus(SARS-CoV-2) using whole genomic data.[J/OL].chinaXiv.[2020-04-8]. <http://eprint.las.ac.cn/abs/202002.00033>. doi:  
322 10.12074/202002.00033.
- 323 [11] Su Y C F, Anderson D E, Young B E, et al.,Discovery of a 382-nt deletion during the early evolution of SARS-CoV-2.  
324 [J/OL]. bioRxiv,[2020-04-8].<https://www.biorxiv.org/content/10.1101/2020.03.11.987222v1>. doi: 10.1101/2020.03.11.987222
- 325 [12] Lau, S.K., et al., Severe Acute Respiratory Syndrome (SARS) Coronavirus ORF8 Protein Is Acquired from SARS-Related  
326 Coronavirus from Greater Horseshoe Bats through Recombination. [J/OL]. J Virol, 2015. 89(20): p. 10532-47
- 327 [13]Cohen J. Wuhan seafood market may not be source of novel virus spreading globally. Science 2020:  
328 10.1126/science.abb0611.

- 329 [14] Chan JF-W, Y uan S, Kok K-H, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating  
330 person-to-person transmission: a study of a family cluster. *Lancet* 2020: 10.1016/S0140-6736(20)30154-9.  
331 [15] Holshue ML, DeBolt C, Lindquist S, et al. First Case of 2019 Novel Coronavirus in the United States. *N Engl J Med* 2020:  
332 10.1056/NEJMoa2001191.

333 **Appendices:**

334

335

**Figure legend**

336

337 **Table1.** The number of nucleotide variation in the genome of SARS-CoV-2 represents strain

338 **Table2.** The nucleotide variation sites in the genome of SARS-CoV-2 represent strains

339 **Table3.** The genome evolution analysis of the SARS-CoV-2 represents strain

340 **Table4.** The primers for nucleic acid detection of SARS-CoV-2 in China CDC and Europe

341 **Supplementary Table1.** Group A evolved into group B, subgroup B1, and amino acid changes of  
342 protein S

343

344 **Figure1.** The geographical distribution of mutant strain in the S gene coding region D614G

345 **Supplementary Figure 1.** the comparison and analysis of subgroups of SARS-CoV-2 A/B group

346

347 **Table 1** Number of nucleotide variation in the genome of the representative  
 348 strains of SARS-CoV-2

Strains	WA1-F6	Chongqing/YC01	Fujian/8	MN908947.3	MN938384.1	WA-UW82	bat/Yunnan/RaTG13
WA1-F6	0	2	0	3	2	<u>9</u>	1156
Chongqing/YC01	2	0	2	5	4	<u>9</u>	1158
Fujian/8	0	2	0	3	2	<u>8</u>	1156
MN908947.3	3	5	3	0	3	<u>5</u>	1159
MN938384.1	2	4	2	3	0	<u>6</u>	1156
WA-UW82	<u>9</u>	<u>9</u>	<u>8</u>	<u>5</u>	<u>6</u>	0	1163
bat/Yunnan/RaTG13	1156	1158	1156	1159	1157	1163	0

349 Note: WA1-F6, BetaCoV/USA/WA1-F6/202 ; Chongqing/YC01, BetaCoV/Chongqing/YC01/2020 ; Fujian/8,  
 350 BetaCoV/Fujian/8/2020; MN908947.3, Wuhan-Hu-1; MN938384.1, 2019-nCoV\_HKU-SZ-002a\_2020;  
 351 bat/Yunnan/RaTG13, BetaCoV/bat/Yunnan/RaTG13/2013; WA-UW82, hCoV-19/USA/WA-UW82/2020.  
 352

353

354 **Table 2** the nucleotide variation loci in the genome of the representative  
355 strains of SARS-CoV-2

Strains	5'UTR	ORF1ab					S	ORF3a	ORF8	N		Type
	241	1059	1342	3037	8782	18060	23403	25563	28144	29095	29200	
bat/Yunnan/RaTG13	C	C	C	C	T	T	A	A	C	T	C	A
WA1-F6	C	C	C	C	T	T	A	<b>G</b>	C	C	C	A
Chongqing/YC01	C	C	<b>T</b>	C	T	T	A	G	C	C	<b>T</b>	A
Fujian/8	C	C	<b>C</b>	C	T	T	A	G	C	C	C	A
MN938384.1	C	C	C	C	T	<b>C</b>	A	G	C	<b>T</b>	C	A
MN908947.3	C	C	C	C	<b>C</b>	<b>C</b>	A	G	<b>T</b>	C	C	B
WA-UW82	<b>T</b>	<b>T</b>	C	<b>T</b>	<b>C</b>	<b>C</b>	<b>G</b>	<b>T</b>	<b>T</b>	C	C	B1

356 Note: 5'UTR, 1...265; ORF1ab, 266...21555; S, 21563...25384; ORF3a, 25393...26220; ORF8,

357 27894...28259; N, 28274...29533.

358

359

360 **Table 3** Genome evolution analysis of the representative strains of  
361 SARS-CoV-2

Sequence 1	Sequence 2	Ks	Ka
USA/WA1-F6	Chongqing/YC01	0.0003	0.0000
USA/WA1-F6	Fujian/8	0.0000	0.0000
USA/WA1-F6	MN908947.3	0.0003	0.0000
USA/WA1-F6	MN938384.1	0.0001	0.0000
USA/WA1-F6	bat/Yunnan/RaTG13	0.1169	0.0117
USA/WA1-F6	USA/WA-UW82/2020	0.0007	0.0002
Chongqing/YC01	Fujian/8	0.0003	0.0001
Chongqing/YC01	MN908947.3	0.0006	0.0002
Chongqing/YC01	MN938384.1	0.0004	0.0002
Chongqing/YC01	bat/Yunnan/RaTG13	0.1173	0.0177
Chongqing/YC01	USA/WA-UW82/2020	0.0010	0.0002
Fujian/8	MN908947.3	0.0003	0.0000
Fujian/8	MN938384.1	0.0001	0.0001
Fujian/8	bat/Yunnan/RaTG13	0.1169	0.0117
Fujian/8	USA/WA-UW82/2020	0.0007	0.0002
MN908947.3	MN938384.1	0.0004	0.0000
MN908947.3	bat/Yunnan/RaTG13	0.1172	0.0178
MN908947.3	USA/WA-UW82/2020	0.0004	0.0002
MN938384.1	bat/Yunnan/RaTG13	0.1167	0.0178
MN938384.1	USA/WA-UW82/2020	0.0009	0.0002
bat/Yunnan/RaTG13	USA/WA-UW82/2020	0.1174	0.0179

362 Note: Ks, synonymous mutation rate; Ka: non-synonymous mutation rate

363

364

365 **Table 4** Primers for nucleic acid detection of SARS-CoV-2 in China CDC and

366 Europe

	Primers	Amplification area	Mutant strain	Variation sequence
China CDC	ORF1ab-F 5'-CCCTGTGGGTTTTACACTTAA-3'	13342-13460	No variation	No variation
	ORF1ab-R 5'-ACGATTGTGCATCAGCTGA-3'			
	ORF1ab-P5'-CCGTCTGCGGTATGTGGAAAGGTT ATGG-3'			
	N-F 5'- <b>GGG</b> GAACTTCTCCTGCTAGAAT-3'	28881-28979	hCoV-19/Mexico/CDMX/InDRE_01 hCoV-19/Netherlands/Tilburg_1364286 hCoV-19/Netherlands/Rotterdam_1364740 hCoV-19/Switzerland/1000477377 hCoV-19/Switzerland/1000477796 hCoV-19/Finland/FIN03032020C hCoV-19/Portugal/CV62 hCoV-19/Switzerland/1000477806 hCoV-19/Switzerland/1000477797 hCoV-19/Peru/010	<b>AAC</b> GAACTTCTCCTGCTA GAAT
	N-R 5'-CAGACATTTTGTCTCAAGCTG-3'			
	N-P 5'-PTTGCTGCTGCTTGACAGATT-3'			
Europe	RdRp-F 5'-GTGARATGGTCATGTGTGGCGG-3'	15431-15530	No variation	No variation
	RdRp-P2 5'-CAGGTGGAACCTCATCAGGAGATGC-3'			
	RdRp-P1 5'-CCAGGTGGWACRTCATCMGGTGATGC-3'			
	RdRp-R 5'-CARATGTAAASACACTATTAGCATA-3'	26269-26381	BetaCoV/South_Korea/SNU01/2020	ACACTAGCCATCCTTACTG CG <b>A</b> TCG
	E-F 5'-ACAGGTACGTTAATAGTTAATAGCGT-3'			
	E-P 5'-ACACTAGCCATCCTTACTGCG <b>T</b> TTCG-3'			
	E-R 5'-ATATTGCAGCAGTACGCACACA-3'	28706-28833	BetaCoV/Shenzhen/SZTH-004	CACATTGGCA <b>T</b> CCGCAAT C
	N-F 5'-CACATTGGCA <b>CC</b> CGCAATC-3'			
E-P5'-ACTTCTCAAGGAACAACATTGCCA-3'				
E-R 5'-GAGGAACGAGAAGAGGCTTG-3'				

367

368

**Figure 1** the geographical distribution of mutant strain in the S gene coding region D614G

