Original Article

Identification of a novel coronavirus causing severe pneumonia in human: a descriptive study

Li-Li Ren[1], Ye-Ming Wang[2,3], Zhi-Qiang Wu[4], Zi-Chun Xiang[1], Li Guo[1], Teng Xu[5], Yong-Zhong Jiang[6], Yan Xiong[7], Yong-Jun Li[5], Hui Li[8], Guo-Hui Fan[9], Xiao-Ying Gu[9], Yan Xiao[1], Hong Gao[10], Jiu-Yang Xu[11], Fan Yang[4], Xin-Ming Wang[1], Chao Wu[1], Lan Chen[1], Yi-Wei Liu[1], Bo Liu[4], Jian Yang[4], Jie Dong[4], Li Li[4], Chao-Lin Huang[12], Jian-Ping Zhao[13], Yi Hu[14], Zhen-Shun Cheng[15], Lin-Lin Liu[5], Zhao-Hui Qian[4], Chuan Qin[10], Qi Jin[4], Bin Cao[2,3], Jian-Wei Wang[1]

[1]NHC Key Laboratory of Systems Biology of Pathogens and Christophe Mérieux Laboratory, IPB, CAMS-Fondation Mérieux, Institute of Pathogen Biology (IPB), Chinese Academy of Medical Sciences (CAMS) & Peking Union Medical College, Beijing 100730, China;

[2]Department of Pulmonary and Critical Care Medicine, Center of Respiratory Medicine, National Clinical Research Center for Respiratory Diseases, China-Japan Friendship Hospital, Beijing 100029, China;

[3]Institute of Respiratory Medicine, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100029, China;

[4]NHC Key Laboratory of Systems Biology of Pathogens, Institute of Pathogen Biology (IPB), Chinese Academy of Medical Sciences (CAMS) & Peking Union Medical College, Beijing 100730, China;

[5] Vision Medicals Co., Ltd, Guangzhou, Guangdong 510000, China;

[6]Hubei Provincial Center for Disease Control and Prevention, Wuhan, Hubei 430079, China;

[7]Wuhan Center for Disease Control and Prevention, Wuhan, Hubei 430022, China;

[8]Clinical and Research Center of Infectious Diseases, Beijing Ditan Hospital, Capital Medical University, Beijing 100015, China;

[9]Institute of Clinical Medical Sciences, China-Japan Friendship Hospital, Beijing 100029, China;

[10]Institute of Laboratory Animal Science, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100021, China;
[11]Department of Basic Medical Sciences, Tsinghua University School of Medicine, Beijing, 100084, China;
[12]Jin Yin-tan Hospital, Wuhan, Hubei 430022, China
[13]Department of Pulmonary and Critical Care Medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Sciences and Technology, Wuhan, Hubei 430030, China;
[14]Department of Pulmonary and Critical Care Medicine, The Central Hospital of Wuhan, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430014, China;
[15]Department of Respiratory medicine, Zhongnan Hospital of Wuhan University, Wuhan, Hubei 430071, China.

Li-Li Ren, Ye-Ming Wang, Zhi-Qiang Wu, Zi-Chun Xiang, Li Guo, Yong-Zhong Jiang, Yan Xiong, and Teng Xu contributed equally to this work.

Correspondence to: Dr. Jian-Wei Wang, No.9 Dong Dan San Tiao, Dongcheng District, Beijing 100730, China
E-Mail: jianwei.wang@ipbcams.ac.cn;
Dr. Bin Cao, No.6 Rong Jing Dong Jie, Daxing District, Beijing 100176, China
E-Mail: caobin_ben@163.com;
Dr. Qi Jin, No.6 Rong Jing Dong Jie, Daxing District, Beijing 100176, China
E-Mail: zdsys@vip.sina.com;
Dr. Chuan Qin, No.5 Pan Jia Yuan Nan Li, Chaoyang District, Beijing 100021, China
E-Mail: qinchuan@pumc.edu.cn

Edited by: Xiu-Yuan Hao and Pei-Fang Wei

Abstract

Background: Human infections with zoonotic coronaviruses (CoVs), including severe acute respiratory syndrome (SARS)-CoV and Middle East respiratory syndrome (MERS)-CoV, have raised great public health concern globally. Here, we

report a novel bat-origin CoV causing severe and fatal pneumonia in humans.

Methods: We collected clinical data and bronchoalveolar lavage (BAL) specimens from five patients with severe pneumonia from Jin Yin-tan Hospital, Wuhan, Hubei province, China. Nucleic acids of the BAL were extracted and subjected to next-generation sequencing. Virus isolation was carried out, and maximum-likelihood phylogenetic trees were constructed.

Results: Five patients hospitalized from December 18 to December 29, 2019 presented with fever, cough, and dyspnea accompanied by complications of acute respiratory distress syndrome. Chest radiography revealed diffuse opacities and consolidation. One of these patients died. Sequence results revealed the presence of a previously unknown β-CoV strain in all five patients, with 99.8–99.9% nucleotide identities among the isolates. These isolates showed 79.0% nucleotide identity with the sequence of SARS-CoV (GenBank NC_004718) and 51.8% identity with the sequence of MERS-CoV (GenBank NC_019843). The virus is phylogenetically closest to a bat SARS-like CoV (SL-ZC45, GenBank MG772933) with 87.6–87.7% nucleotide identity, but is in a separate clade. Moreover, these viruses have a single intact open reading frame gene 8, as a further indicator of bat-origin CoVs. However, the amino acid sequence of the tentative receptor-binding domain resembles that of SARS-CoV, indicating that these viruses might use the same receptor.

Conclusion: A novel bat-borne CoV was identified that is associated with severe and fatal respiratory disease in humans.

Keywords: Bat-origin; Coronavirus; Zoonotic transmission; Pneumonia; Etiology; Next-generation sequencing

Introduction

Coronaviruses (CoVs) are enveloped viruses with a single positive-stranded RNA genome (~26–32 kb in length). They belong to the subfamily *Orthocoronavirinae* under the family *Coronaviridae*, and are classified into four genera: *Alphacoronaviruses* (α), *Betacoronaviruses* (β), *Gammacoronaviruses* (γ), and *Deltacoronaviruses* (δ).[1,2] The viral genome normally encodes four structural proteins, spike (S), envelope (E), membrane (M), and nucleocapsid (N), as well as several non-structural proteins and multiple unique accessory proteins.[1,2]

CoVs infect humans and a variety of avian and mammalian species worldwide. There are six CoVs known to infect humans, including two α-CoVs (229E and NL63) and four β-CoVs (OC43, HKU1, severe acute respiratory syndrome [SARS]-CoV, and Middle East respiratory syndrome [MERS]-CoV).[1-4] All human CoVs are zoonotic as a distinguishing characteristic.[5] In particular, bats are regarded as a key reservoir of CoVs, and many human CoVs are believed to have originated from bats.[5,6] Since the beginning of this century, two zoonotic CoVs, SARS-CoV and MERS-CoV, have been identified to cause severe human diseases.[3,4,7] The outbreak of SARS-CoV in 2003 was responsible for 8,096 cases and 794 deaths worldwide.[8] Since its discovery in Middle Eastern countries in 2012, MERS-CoV has infected 2,260 people with a current case fatality ratio of 35.5%.[9,10] These outbreaks have raised public health concerns of the potential for the emergence of another novel zoonotic CoV.

Here, we report a previously unknown bat-origin CoV causing severe and fatal pneumonia in five patients from Wuhan, China. Sequence results revealed that this virus, harboring a single open reading frame gene 8 (ORF8), is phylogenetically closest to bat SARS-like CoV, but is in a separate lineage. Furthermore, the amino acid sequence of the tentative receptor-binding domain (RBD) of this new CoV resembles that of SARS-CoV, indicating that they might use the same receptor. These findings highlight the urgent need for regular surveillance of the interspecies transmission of bat-origin CoV to human populations.

Methods

Ethical approval

This study was conducted in accordance with the *Declaration of Helsinki* and was approved by the National Health Commission of China and Ethics Commission of the Jin Yin-tan Hospital (No. KY-2020-01.01). The requirement for written informed

consent was waived given the context of emerging infectious diseases.

Clinical specimen and data collection

Bronchoalveolar lavage fluid (BAL) samples were collected from five patients hospitalized with pneumonia in Jin Yin-tan Hospital, Wuhan, Hubei province, China from December 18 to 29, 2019. Information was gathered, including clinical data, demographic characteristics, underlying medical conditions, clinical signs and symptoms, chest radiographic findings, clinical laboratory testing results, traveling history, recent animal exposure, and outcomes. The data collected for the cases were deemed by the National Health Commission of the People's Republic of China as the contents of a public health outbreak investigation.

Genome sequencing

Nucleic acids were extracted from 200 µl BAL of each sample with the Direct-zol RNA Miniprep kit (Zymo Research, Irvine, CA, USA) and Trizol LS (Thermo Fisher Scientific, Carlsbad, CA, USA) according to the manufacturer instructions in a biosafety III laboratory. A 50-µl elution was obtained from each sample. The DNA/RNA concentrations were measured by a Qubit Fluorometer (Thermo Fisher Scientific, Carlsbad, CA, USA). The sequencing library was constructed by a transposase-based methodology and sequenced on an Illumina Hiseq sequencing platform (Illumina, San Diego, CA, USA). At least 25 million single-end 76-bp reads were generated for each sample on the Illumina NextSeq platform. Quality control processes included removal of low-complexity reads by bbduk (entropy = 0.7, entropy window = 50, entropy k = 5; version: January 25, 2018),[11] adapter trimming, low-quality reads removal, short reads removal by Trimmomatic (adapter: TruSeq3-SE.fa:2:30:6, LEADING: 3, TRAILING: 3, SLIDING WINDOW: 4:10, MINLEN: 70, version: 0.36),[12] host removal by bmtagger (using human genome GRCh38 and yh-specific sequences as reference),[13] and ribosomal reads removal by SortMeRNA (version: 2.1b).[14] Taxonomic assignment of the clean reads was performed with Kraken 2 against the reference databases, including archaea, bacteria, fungi, human, plasmid, protozoa, univec, and virus sequences (software 2.0.7-beta, database version: August 2, 2019).[15] A negative control sample was processed and sequenced in parallel for each sequencing run as a contamination control. The data were classified by simultaneous alignment to the microbial genome databases

comprising viruses, bacteria, fungi, and parasites after filtering of the adapters and human-origin reads. The sequences were confirmed by Sanger sequencing with specific primers and one-step real-time polymerase chain reaction (RT-PCR) Kit (Invitrogen, Carlsbad, CA, USA).

Phylogenetic analysis

Multiple sequence alignment was performed with the ClustalW program using MEGA software (version 7.0.14). Phylogenetic trees were constructed by means of the maximum-likelihood method with MEGA software (version 7.0.14). The full-genome viral sequences were deposited in the dataset of Global Initiative on Sharing All Influenza Data (GISAISD, No. EPI_ISL_402123, EPI_ISL_403928-31) and the Genome Warehouse in National Genomics Data Center, Beijing Institute of Genomics (BIG), Chinese Academy of Sciences, under Project ID PRJCA002165 that is publicly accessible at https://bigd.big.ac.cn/gwh as of January 2020.

Virus isolation

The BAL specimens were inoculated onto Vero cells (ATCC, CCL-81). All cultures were observed daily for a cytopathic effect (CPE). Maintenance medium containing tosyl-phenylalanine chloromethyl-ketone (TPCK) enzyme at a final concentration of 1 μg/mL was replenished at day 4, and cultures were terminated 7 days after inoculation. The viral particles were negative stained with 1% solution of phosphotungstic acid (pH 7.0) and the morphology was characterized by using 120 kV TECNAI (Thermo Fisher Scientific, Hillsboro, OR, USA) electron microscopy and camera of Gatan832 (Gatan, Pleasanton, CA, United States). The culture supernatants of cells demonstrating CPE were mixed with paraformaldehyde, dried onto formvar/carbon-coated grids, and stained. Viral nucleic acids were confirmed by RT-PCR with specific primers (Supplementary table 1).

Immunofluorescence assay

Spot slides were prepared by applying 20 μl of the virus-infected or non-infected cell suspension onto 12-well Teflon-coated slides. The cells were fixed with 4% paraformaldehyde in 1× phosphate-buffered saline (PBS) for 30 min, washed three times with PBS, blocked, and stained with serum from a convalescent patient or serum from a healthy person for 30 min at 37°C at a dilution of 1:200. Goat

anti-human immunoglobulin G conjugated with fluorescein isothiocyanate was used as the secondary antibody (Jackson Immuno Research Laboratories, Inc., West Grove, PA, USA). Nuclei and the cytoplasm were counterstained with 4′,6′-diamidino-2-phenylindole (DAPI) and Evans blue (Sigma-Aldrich, St. Louis, MO, USA). Fluorescent images were obtained and analyzed using laser-scanning confocal microscopy (Airyscan LSM880, Zeiss, Berlin, Germany).

Results

Genereal information of patients

Patient 1 was a 65-year-old man who reported a high fever and cough, with little sputum production, at the onset of illness. He had a continuous fever and developed severe shortness of breath 16 days later. He was a vendor at the Huanan seafood market, Wuhan, Hubei Province, China. Patient 2, a 49-year-old woman, presented with high fever and dry cough. Five days later, she developed dyspnea and was admitted to the hospital. She was also a worker in the Huanan seafood market. Patient 3 was a 52-year-old woman who did not report any market exposure. She was admitted to hospital because of fever, cough, and ground-glass opacity in the chest computed tomography (CT) scan. Patient 4 was a 41-year-old man who also presented with high fever and dry cough at the onset of the illness. He developed acute respiratory distress syndrome 7 days later. This patient had no known history of exposure to the Huanan seafood market. Patient 5, a 61-year-old man, was admitted to a local hospital with a 7-day history of fever, cough, and dyspnea. He also worked in the market.

With regards to medical history, Patient 4 had hypertension, Patient 5 had chronic liver disease and myxoma of abdominal, whereas none of the other patients had a record of underlying diseases. The demographic and clinical characteristics of the five patients are summarized in Table 1.

Novel CoV identification by next-generation sequencing

The resultant clean reads accounted for 12.0–92.0% of the raw reads. Most of the reads could be successfully assigned. Notably, 80.3% of the reads mapped to the viral genome for sample from Patient 5 with the highest proportion of viral reads

among the five samples. Nearly all of the viral reads (97%) were classified as *Coronaviridae*. Similarly, in the other four patients, most of the viral reads were assigned to β-CoVs [Table 1]. Based on *de novo* assembly and careful curation, a consensus sequence of this CoV was obtained.

A substantial proportion of all sequencing reads mapped to the newly reported CoV genome (BWA mem, version: 0.7.12),[16] ranging from 71,883 (0.27% among all reads) in Patient 4 to 37,247,818 (85%) in Patient 5 [Figure 1]. In addition, very few reads mapped to known bacterial pathogens, including *Streptococcus*, *Acinetobacter baumannii*, and *Pseudomonas* [Figure 1].

The reads mapping to CoVs were assembled, and their genome sequences were confirmed by Sanger sequencing. The nucleotide (nt) similarity among the obtained five whole-genome sequences was 99.8–99.9%. The full length of the obtained genome was 29,870 bp with a GC content of 37.99–38.02%. The genome organization, 5′-ORF1ab–S–E–M–N–3′, was similar to that of the most well-known bat SARS-like (SL)-CoV. In addition, unique accessory open reading frames (UA-ORFs) were identified that are characterized in the subgenus *Sarbecovirus*, encoding putative ORF3, ORF6, ORF7, and ORF8 proteins reading from the 5′-terminus to the 3′-terminus between the structural proteins [Figure 2A].

Phylogenetic analysis

Homology assessment showed that full-length viral genome sequences have 79.0% nt identity with that of SARS-CoV Tor2 (GenBank NC_004718), 51.8% with that of MERS-CoV (GenBank NC_019843), and 87.6–87.7% with those of bat SL-CoV ZC45 and ZXC21 (GenBank MG772933, MG772934), isolated from Chinese horseshoe bats (*Rhinolophus sinicus*) [Table 2], indicating that the novel CoVs are most similar to bat SL-CoVs.

Compared with bat SL-CoV ZC45, the novel CoVs showed 75.9%, 98.6%, 93.2–93.4%, and 91.1% nt identities in the S, E, M, and N genes, respectively. Overall, ORF1ab showed 89.0% nt identity between the novel CoVs and bat SL-CoV ZC45. Surprisingly, RNA-dependent RNA polymerase (RdRp), which is the most highly conserved sequence among different CoVs,[1, 4] only showed 86.3–86.5% nt identities with bat SL-CoV ZC45. According to the International Committee on Taxonomy of Viruses criteria, a new CoV species could be defined if the nt identity is less than 90%

for the conserved RdRP sequence.[4] Thus, we considered that the novel CoVs should be classified as a new species under the subgenus *Sarbecovirus* of the genus *Betacoronavirus.*

The phylogenetic trees constructed with the sequences of the RdRp, S, and N genes, and the whole genome using a maximum-likelihood model showed that all five novel CoVs were closely related to bat SL-CoVs ZXC21 and ZC45, but in a separate evolutionary lineage under the subgenus *Sarbecovirus* [Figure 2B–E], which is consistent with the homology assessment results.

ORF3 and intact ORF8 gene regions were present in the novel CoVs, which are the characteristic features of bat-origin CoVs.[17,18] ORF3 of the novel CoVs showed 87.8% nt and 90.9% amino acid (aa) identities with bat SL-CoV ZC45, but less than 76.8% nt and 76.0% aa identities with the other members in the subgenus *Sarbecovirus*. In addition, ORF8 of the novel CoVs showed 88.5% and 94.2% nt and aa identity with bat SL-CoV ZC45, respectively, and less than 67.8% and 58.6% nt and aa identity, respectively, with other members of *Sarbecovirus*. These findings further indicated that the novel CoVs are of bat origin.

The RBD in the CoV S protein determines the host range.[19] The RBD aa sequences of the novel CoV showed several distinct features, including higher aa identities with those of SARS-CoV (73.8–74.8%) and human angiotensin-converting enzyme 2 (hACE2)-using SL-CoVs (76.4–76.9%) than those of SL-CoVs incapable of using hACE2 (61.5–64.1%). The novel CoV does not possess the deletions commonly found in the RBD of SL-CoVs incapable of using hACE2 as a receptor [Figure 2F].[20] In addition, the five critical aa residues interacting with hACE2 in SARS-CoV RBD (Y442, L472, N479, D480, T487) differ from the corresponding residues in the novel CoVs (L, F, Q, S, N), although these residues possess similar polarity.[20,21] These results suggested that the novel CoVs might still use hACE2 as the receptor.

Viral culture

CPE were observed in 30% of Vero cells inoculated with the new CoV after two passages [Figure 3A]. The cells showed a round, refractive, and syncytium appearance. The Vero cells with CPE were further examined using negative-staining electron microscopy, demonstrating characteristic CoV particles with surface projections [Figure 3B]. Immunofluorescent assays of the culture of Vero cells showing CPE with the convalescent serum from patients showed green signals in the

cytoplasm, with no signals detected in wells containing control serum, indicating the presence of viral particles in the cells [Figure 3C].

Clinical features and outcomes of the patients

The clinical features and laboratory test results of the five patients are summarized in Table 2. Fever, cough, and dyspnea were the most common symptoms. The white blood cell counts varied among these patients, but the lymphocyte counts were generally low. The alanine aminotransferase and serum creatine levels were normal or only slightly increased. Bilateral ground-glass opacities and consolidation were observed on chest radiography from two representative patients, Patient 2 based on aortic arch scan [Figure 4A] and pulmonary vein scan [Figure 4B] on day 10 after symptoms onset and Patient 5 taken on day 12 [Figure 4C] and 13 [Figure 4D] after symptoms onset.

Several complications were observed in these patients. Four of the five patients (except for Patient 3) developed acute respiratory distress syndrome requiring oxygen therapy, and two patients were given extracorporeal membrane oxygenation. Two patients (Patients 1 and 5) experienced secondary infections, and Patient 5 later developed septic shock as well as acute kidney injury, and ultimately died of multi-organ failure. Patient 3 was discharged on January 8, 2020 (day 17 after symptoms onset). The other three patients were still hospitalized at the time of manuscript preparation.

Discussion

In this study, we identified a previously unknown CoV from patients suffering from severe pneumonia. The whole-genome sequences of the viruses were obtained by a next-generation sequencing approach from all five patients, demonstrating overwhelmingly dominant viral reads in the BAL samples. Among the five novel CoV genome sequences, the nt identities reached up to 99.8–99.9%. The viruses successfully isolated from the patients could also be effectively recognized by serum from a convalescent patient. These findings primarily indicate that the novel CoV is associated with the pneumonia that developed in these patients. However, it remains to be determined whether this novel CoV is capable of causing similar diseases in experimental animals.

Sequence homology analysis of the viral genome showed that the CoV identified in this study is distinct from any of the known human CoVs, including SARS-CoV and MERS-CoV. The most closely related known viruses are two bat SL-CoVs (GenBank accession nos. MG772933, 772934) identified in 2005 in Zhoushan, Zhejiang, China, which is geographically distant from Wuhan;[22] however, the nt identities among the viruses are only 85.7–86.8%. Phylogenetic analysis showed that this virus forms a single clade. Collectively, these data indicate that this CoV should be considered a new species. The outbreak of SARS in 2003 largely boosted awareness of threats caused by emerging CoVs. Consequently, great efforts have been made to monitor novel emerging CoVs and to trace their origins so as to establish a risk assessment and alert system for preventing potential epidemics in the human population. Clarification of the coronavirome in animals, particularly in bats as a key reservoir of a wide range of CoVs, should be a priority for any task force.[23,24]

A few striking features of these novel CoVs indicated that they are of bat-origin. First, the genome sequences of the novel CoVs show high similarity with that of bat SL-CoV ZC45. Second, the phylogenetic analysis indicated that these viruses are evolutionarily close to bat SL-CoVs ZXC21 and ZC45. Third, all of these novel CoVs contain ORF3 and intact ORF8 gene regions, which are characteristic features of bat-origin CoVs. [17,20] Moreover, the aa sequences of the N-terminal domains (NTDs) of the novel CoVs were very similar to those of ZC45 and ZXC21, whereas the RBD of the novel CoV showed higher aa sequence identity to that of SARS-CoV than to those of ZC45 and ZXC21, suggesting that a recombination event might have occurred at the region between the NTD and RBD of the S gene, facilitating the interspecies transmission.

Owing to the lack of epidemic information at present, the transmission modes of the novel CoV remain obscure. It is notable that four of the five patients had a history of recent exposure to a seafood market in Wuhan. However, the origin of infection is unknown at the time of manuscript preparation. It is assumed that the zoonotic CoV jumped to humans through an intermediate host; for example, camel is suspected as the intermediate host of MERS-CoV, whereas the palm civet may have contributed to the interspecies transmission of SARS-CoV to humans.[25,26] Bat CoVs may evolve to adapt to using humans as a host during their circulation in a mammalian host, thereby enabling them to effectively infect humans.[26] However, two of our patients did not have a history of exposure to the seafood market. Therefore, further investigation will

be needed to determine the potential of multiple infection sources responsible for this uncommon outbreak.

One of the most striking and concerning features of this virus is its ability to cause severe respiratory syndrome. The disease progressed rapidly with a major presentation of lower respiratory pathology. Notably, no obvious upper respiratory tract symptoms such as a sore throat and rhinorrhea were present in these patients. Therefore, further exploration is needed on the distribution of the viral receptor in the organs to potentially account for pathogenesis development. In addition, the possibility of unrecognized mild infections or subclinical infections should be clarified, as identification of such infections is critical to control spread of the disease. Development of serological assays would be largely beneficial to detect such types of infection at the population level.

In conclusion, we identified a novel bat-borne CoV associated with a severe and fatal respiratory disease in humans. The emergence of this virus poses a potential threat to public health. Therefore, clarification of the source and transmission mode of these infections is urgently needed to prevent a potential epidemic.

Funding

This study was supported by grants from the Chinese Academy of Medical Sciences (CAMS) Innovation Fund for Medical Sciences (No. 2016-I2M-1-014), the National Major Science & Technology Project for Control and Prevention of Major Infectious Diseases in China (Nos. 2017ZX10103004, 2018ZX10305409, 2017ZX10204401), and the National Natural Science Foundation (No. 81930063) .

Acknowledgments

We would like to thank Dr. Chen Wang, Dr. Zhen-Dong Zhao, Dr. Fei Guo (Chinese Academy of Medical Sciences & Peking Union Medical College) and Dr. Ming-Kun Li (Beijing Institute of Genomics, Chinese Academy of Sciences) for critical reading of the manuscript and helpful discussions. We thank the colleagues who contributed to sample collection and experiments.

Conflicts of interest

None.

References

1. Woo PC, Lau SK, Huang Y, Yuen KY. Coronavirus Diversity, Phylogeny and Interspecies Jumping. Exp Biol Med (Maywood) 2009;234:1117-1127. doi: 10.3181/0903-MR-94.

2. Perlman S, Netland J. Coronaviruses post-SARS: update on replication and pathogenesis. Nat Rev Microbiol 2009;7:439-450. doi: 10.1038/nrmicro2147.

3. Zhong NS, Zheng BJ, Li YM, Poon LL, Xie ZH, Chan KH, et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. Lancet 2003;362:1353-1358. doi: 10.1016/s0140-6736(03)14630-2.

4. Zaki AM, van BS, Bestebroer TM, Osterhaus AD, Fouchier RA. Isolation of a Novel Coronavirus from a Man with Pneumonia in Saudi Arabia. N Engl J Med 2012;367:1814-1820. doi: 10.1056/NEJMoa1211721.

5. Corman VM, Muth D, Niemeyer D, Drosten C. Hosts and Sources of Endemic Human Coronaviruses. Adv Virus Res 2018;100:163-188. doi: 10.1016/bs.aivir.2018.01.001.

6. Brook CE, Dobson AP. Bats as 'special' reservoirs for emerging zoonotic pathogens. Trends Microbiol 2015;23:172-180. doi: 10.1016/j.tim.2014.12.004.

7. de WE, van DN, Falzarano D, Munster VJ. SARS and MERS: recent insights into emerging coronaviruses. Nat Rev Microbiol 2016;14:523-534. doi:10.1038/nrmicro.2016.81.

8. Summary of probably SARS cases with onset of illness from 1 November 2002 to 31 July 2003. Geneva: World Health Organization. Available from: http://www.who.int/csr/sars/country/table2004_04_21/en/. [Last accessed on 2020 January 20].

9. Kim KH, Tandi TE, Choi JW, Moon JM, Kim MS. Middle East respiratory syndrome coronavirus (MERS-CoV) outbreak in South Korea, 2015: epidemiology, characteristics and public health implications. J Hosp Infect 2017;95:207–213. doi:10.1016/j.jhin.2016.10.008.

10. Middle East respiratory syndrome coronavirus (MERS-CoV) – Saudi Arabia. Geneva: World Health Organization. Available from: http://www.who.int/csr/don/26-april-2016-mers-saudi-arabia/en/. [Last accessed on 2020 January 20].

11. BBDuk Guide. Joint Genome Institute. Available from: https://jgi.doe.gov/data-and-tools/bbtools/bb-tools-user-guide/bbduk-guide/. [Last accessed on 2019 August 1]

12. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114-2120. doi:10.1093/bioinformatics/btu170.

13. Wang J, Wang W, Li R, Li Y, Tian G, Goodman L, et al. The diploid genome sequence of an Asian individual. Nature 2008;456:60-65. doi: 10.1038/nature07484.

14. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 2012; 28: 3211-3217. doi: 10.1093/bioinformatics/bts611.

15. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol 2019;20:257. doi:10.1186/s13059-019-1891-0.

16. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-1760. doi:10.1093/bioinformatics/btp324.

17. Wu Z, Yang L, Ren X, Zhang J, Yang F, Zhang S, et al. ORF8-Related Genetic Evidence for Chinese Horseshoe Bats as the Source of Human Severe Acute Respiratory Syndrome Coronavirus. J Infect Dis 2016;213:579-583. doi: 10.1093/infdis/jiv476.

18. Hu B, Zeng LP, Yang XL, Ge XY, Zhang W, Li B, et al. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. Plos Pathog 2017;13:e1006698. doi: 10.1371/journal.ppat.1006698.

19. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, et al. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature 2013;503:535-538. doi: 10.1038/nature12711.

20. Li F. Receptor recognition and cross-species infections of SARS coronavirus. Antiviral Res 2013;100:246-254. doi:10.1016/j.antiviral.2013.08.014.

21. Hu D, Zhu C, Ai L, He T, Wang Y, Ye F, et al. Genomic characterization and infectivity of a novel SARS-like coronavirus in Chinese bats. Emerg Microbes Infect 2018;7:154. doi: 10.1038/s41426-018-0155-5.

22. Wong ACP, Li X, Lau SKP, Woo PCY. Global Epidemiology of Bat Coronaviruses. Viruses 2019;11:174. doi:10.3390/v11020174.

23. Smith I, Wang LF. Bats and their virome: an important source of emerging

viruses capable of infecting humans. Curr Opin Virol 2013;3:84-91. doi:10.1016/j.coviro.2012.11.006.

24. Drexler JF, Corman VM, Drosten C. Ecology, evolution and classification of bat coronaviruses in the aftermath of SARS. Antiviral Res 2014;101:45-56. doi:10.1016/j.antiviral.2013.10.013.

25. Guan, Y. Isolation and Characterization of Viruses Related to the SARS Coronavirus from Animals in Southern China. Science 2003;302:276-278. doi: 10.1126/science.1087139.

26. Cui J, Li F, Shi ZL. Origin and evolution of pathogenic coronaviruses. Nat Rev Microbiol 2019;17:181-192. doi: 10.1038/s41579-018-0118-9.

Figure legends

Figure 1: Microbial species profiles (proportions) in BAL specimens from Patient 1 (A), Patient 2 (B), Patient 3 (C), Patient 4 (D), and Patient 5 (E) analyzed by deep sequencing.

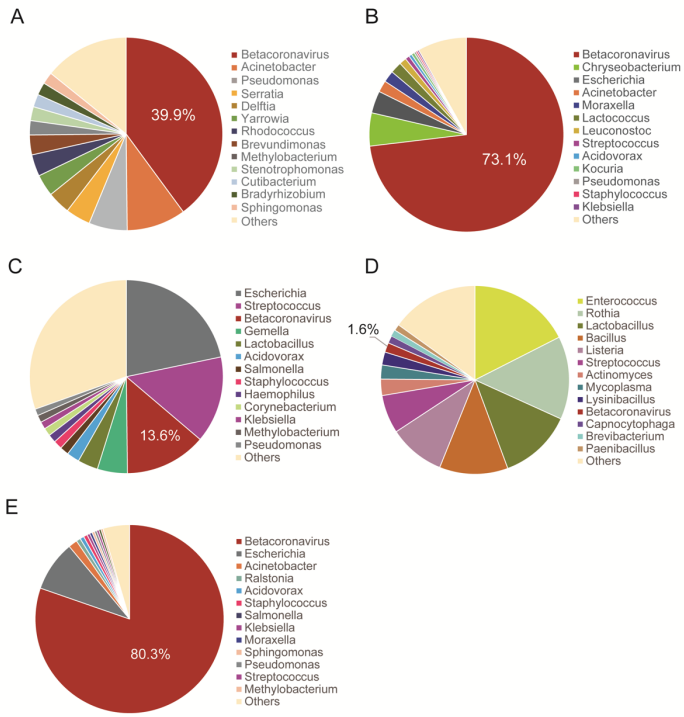

Figure 2: Characteristics of viral genes. The schematic diagram of the novel coronavirus (CoV)'s genome (A). Phylogenic analysis of viral whole genome (B), spike (C), nucleocapsid (D) and RNA-dependent RNA polymerase (E) genes. The novel CoV identified and closely related viruses are in red in the phylogenetic trees. Other outgroup viruses from public database are shown in blue. Evolutionary distances were calculated with the maximum likelihood method. Amino acid sequence of the putative receptor binding domain (RBD) of IPBCAMS-WH-01/2019 compared with strains of severe acute respiratory syndrome coronavirus (SARS-CoV) and bat SARS-like CoV (F).

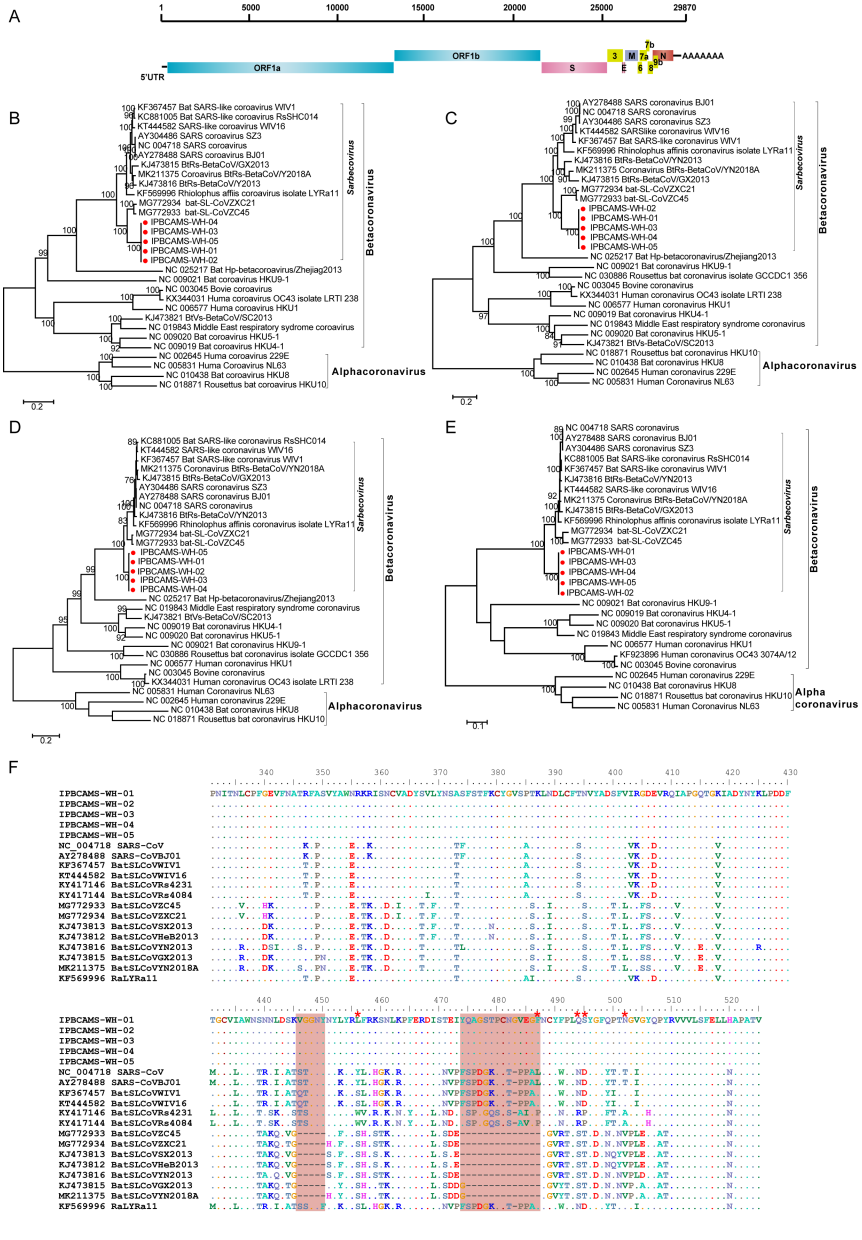

Figure 3: Viral isolation and identification. The untreated control (left) and cytopathic effect (right) on Vero cells visualized with an inverted microscope (A, original

magnification ×20). Viral particles are negative stained with 1% solution of phosphotungstic acid and observed by using electron microscopy analysis (B, scale bar 200 nm). Immunofluorescent assays with convalescent serum and anti-human IgG conjugated FITC show signals (green) in cytoplasma in Vero cells. The Nuclei and cytoplasma were counterstained with 4′,6′-diamidino-2-phenylindole (DAPI, blue) and Evans blue (red), respectively (C, original magnification ×40).

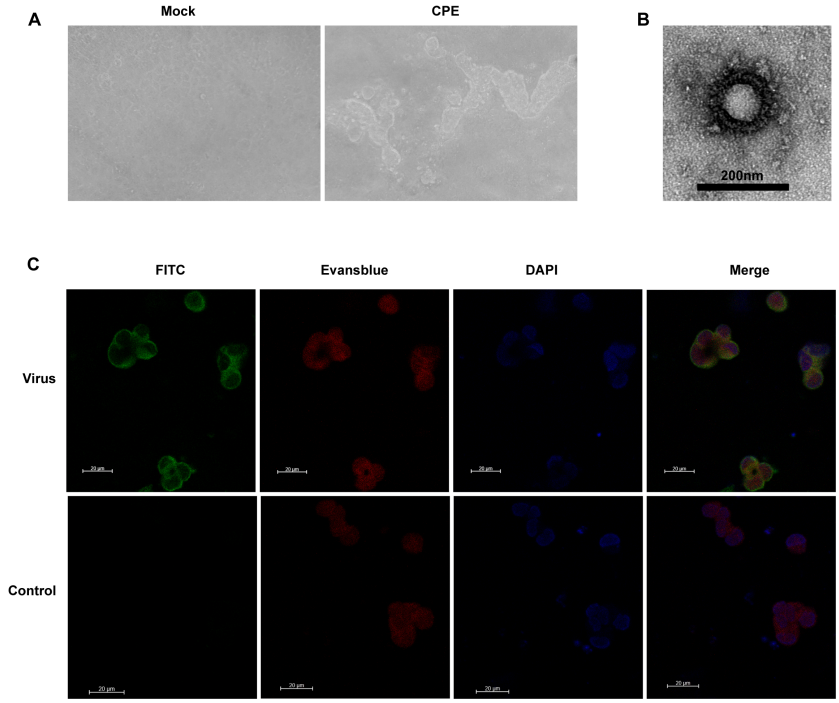

Figure 4: Computed tomographic chest radiographs of Patient 2, obtained on day 10 from illness onset at aortic arch (A) and pulmonary vein (B) scan demonstrating bilateral ground-glass opacity and consolidation, and Patient 5 on day 12 (C) and 13 (D) after illness onset demonstrating white lungs.

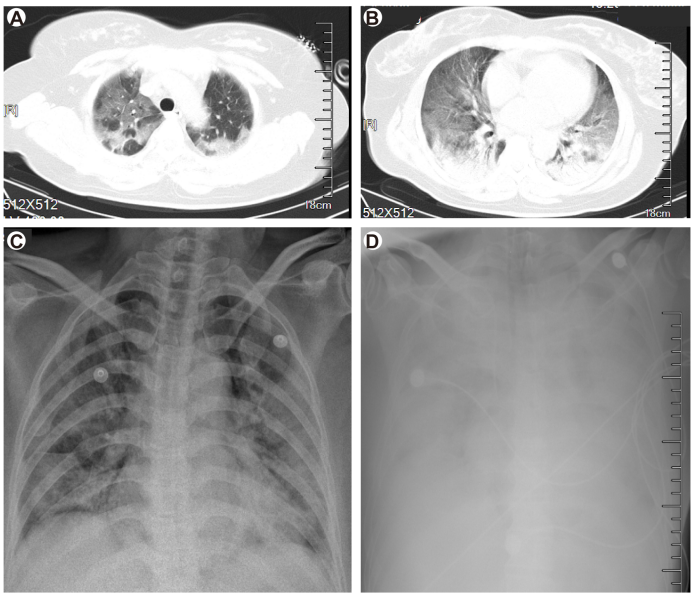
A
B
C
D
512X512
18cm

Table 1: Demographic, epidemiologic, clinical presentations, and treatment of the patients.

Characteristics	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
Age, year	65	49	52	41	61
Gender	Male	Female	Female	Male	Male
Huanan Seafood Market exposure	Yes	Yes	No	No	Yes
Smoking	No	No	No	No	No
Underlying conditions	No	No	No	Hypertension	Chronic liver disease, myxoma of abdominal cavity
Date of illness onset	Dec 15, 2019	Dec 22, 2019	Dec 22, 2019	Dec 16, 2019	Dec 20, 2019
Date of admission	Dec 18, 2019	Dec 27, 2019	Dec 29, 2019	Dec 22, 2019	Dec 27, 2019
Admission to ICU	Dec 22, 2019	Dec 29, 2019	Dec 29, 2019	Dec 30, 2019	Dec 28, 2019
Sampling date	Dec 24, 2019	Dec 30, 2019	Dec 30, 2019	Dec 30, 2019	Jan 1, 2020
Signs and Symptoms					
Cough	Yes	Yes	Yes	Yes	Yes
Sputum production	Yes	No	No	No	No
Dyspnea (onset date)	Dec 31, 2019	Dec 27, 2019	No	Dec 22, 2019	Dec 27, 2019
Fever	Yes	Yes	Yes	Yes	Yes
Muscle pain/Fatigue	No	No	Yes	Yes	Yes
Headache	No	No	Yes	No	No
Diarrhea	No	No	No	No	No
Highest temperature (°C)	39.3	38.5	37.5	39	N.A.
Laboratory tests on admission					
Number of white blood cells, $\times 10^9$ /L	11.9	8.3	2.4	6.6	17.9
Number of neutrophils, $\times 10^9$ /L	11.6	7.6	2.0	5.0	16.2

Number of lymphocytes, $\times 10^9$ /L	0.2	0.44	0.3	0.98	1.2
Number of platelets, $\times 10^9$ /L	92	273	140	129	315
Prothrombin time, s	12	12.5	12.3	11.2	28.4
APTT, s	25.5	25.2	28.5	26.9	105.2
D-dimer, mg/L	40.7	0.21	0.65	1.31	20.6
ALT, U/L	50	45	19	29	35
Serum potassium, mmol/L	5.0	3.8	4.0	5.4	4.8
Serum creatine, μmol/L	53.1	42.6	54.5	85.1	106.6
Procalcitonin, ng/mL	1.46	<0.05	< 0.05	<0.05	0.69
Pathogen testing					
Coronavirus reads proportion, %	39.9	73.1	13.6	1.6	80.3
Imaging					
Bilateral GGO	Yes	Yes	Yes	Yes	Yes
Consolidation	Yes	Yes	No	Yes	Yes
Treatment					
Oxygen therapy	Mechanical ventilation	HFNC	Nasal cannula	Non-invasive ventilation	Mechanical ventilation
ECMO (initiation date)	Jan 6, 2020	No	No	No	Jan 2, 2020
Antibiotic therapy	Yes	Yes	Yes	Yes	Yes
CRRT	No	No	No	No	Yes
Complications					
ARDS (onset date)	Dec 22, 2019	Dec 29, 2019	No	Dec 23, 2019	Dec 27, 2019
Septic shock (onset date)	No	No	No	No	Dec 31, 2019
AKI (onset date)	No	No	No	No	Dec 31, 2019

Secondary infections	Yes	No	No	No	Yes
Outcome*	Hospitalized	Hospitalized	Discharged	Hospitalized	Death

*Outcome data as of January 11, 2020.

ICU: Intensive Care Unit; WBC: White blood cell; APTT: Activated partial thromboplastin time; ALT: Alanine aminotransferase; HFNC: High-flow nasal cannula; GGO: Ground-glass opacity; ECMO: Extracorporeal membrane oxygenation; CRRT: Continuous renal replacement therapy; ARDS: Acute respiratory distress syndrome; AKI: Acute kidney injury.

Table 2: Location and size of the putative proteins of the representative strain, IPBCAMS-WH-01/2019.

Gene region	Location*	Size, aa	GC content, %	Similarity, % (nt/aa)	
				SARS-Cov	bat-SL-ZC45
Whole genome	NA	NA	38.02	79.0	87.7
5′UTR	1–265	NA	44.53	88.7	93.5
nsp1	266–805	180	48.52	82.2/84.4	93.1/95.5
nsp2	806–2719	638	39.86	68.3/68.3	92.2/95.2
nsp3 (PLpro)	2720–8554	1945	35.94	72.7/75.8	91/93.8
nsp4	8555–10,054	500	36.47	74.8/79.8	90.2/96.6
nsp5 (3CLpro)	10,055–10,972	306	38.13	83.6/96.0	92.1/99.0
nsp6	10,973–11,842	290	36.32	79/87.2	91.7/97.9
nsp7	11,843–12,091	83	38.15	79.9/98.7	82.7/100
nsp8	12,092–12,685	198	38.38	86.1/97.4	87.2/97.4
nsp9	12,686–13,024	113	40.12	84.9/97.3	86.7/97.3
nsp10	13,025–13,441	139	42.44	88.2/97.1	90.8/97.1
nsp12 (RdRp)	13,442–13,468 13,468–16,236	932	37.28	88.5/96.3	86.5/95.9
nsp13 (Hel)	16,237–18,039	601	38.16	88.4/99.8	87.9/99.3
nsp14 (ExoN)	18,040–19,620	527	38.27	83.3/95.0	82.9/94.4
nsp15	19,621–20,658	346	34.01	82/88.7	82.7/89.0
nsp16	20,659–21,552	298	35.46	85.1/93.2	89.5/97.9
S	21,563–25,384	1274	37.31	72.3/75.5	75.9/80.4
3	25,393–26,220	276	39.49	75.3/72†	87.8/90.9
E	26,245–26,472	76	38.16	93.5/94.7	98.6/100

M	26,523–27,191	223	42.6	85.2/90.5	93.4/98.6
6	27,202–27,387	62	27.96	73.9/66.6	95.1/93.4
7a	27,394–27,759	122	38.25	82.1/85.2	88.7/87.6[‡]
7b	27,756–27,887	43	31.06	82.0/79.5	NA
8	27,894–28,259	122	35.79	NA	88.5/94.2
N	28,274–29,533	420	47.2	88.1/90.5	91.1/94.2
9b	28,284–28,577	97	49.66	88.5/72.4	89.1/73.1
3′UTR	29,534–29,870		40.36	95.2	93.2

[*]Representative strain, IPBCAMS-WH-01/2019, no. EPI_ISL_402123. [†]Compared with 3a of SARS-CoV. [‡]Compared with 7a of bat-SL-ZC45. aa: amino acids; nt: nucleotide; SARS-CoV: severe acute respiratory syndrome coronavirus; SL: SARS-like; UTR: untranslated regions; nsp: non structural protein; NA: not applicable.

Supplementary Table 1. Primer sequences for whole genome amplification.

Primer code	Sequence (5'->3')	Start	Stop	Targeted gene region
5'RACE-1	CGATCATCAGCACATCTAGG	218	237	UTR
5'RACE-2	CTGTCGTTGACAGGACACGAGTAA	147	170	
5'RACE-3	AGTGCACTCACGCAGTAT	113	130	
B1	F: TGTCGTTGACAGGACACGAG	148	167	ORF1ab
	R: TAACAAAATCGCCCGTCTGC	1277	1258	
B2	F: GTCTATCCAGTTGCGTCACCA	1172	1192	
	R: TCCACCGACAATTTCACAAGC	2251	2231	
B3	F: AACCCGTCCTTGATTGGCTT	2139	2158	
	R: TGATTGTCCTCACTGCCGTC	3258	3239	
B4	F: GCCACTTCTGCTGCTCTTCA	3155	3174	
	R: TTTCAGTAGTGCCACCAGCC	4199	4180	
B5	F: GGGTGATGTTGTTCAAGAGGG	4126	4146	
	R: AGCCTCAACACGTAGAGTGTC	5158	5138	
B6	F: GGACAACAGTTTGGTCCAACTTA	5036	5058	
	R: AGCTTGCGTTTGGATATGGTTG	6038	6017	
B7	F: CAAAGTCCTCAGAATACAAAGGTCC	5826	5850	
	R: GCCTCTAGACAAAATTTACCGACA	6903	6880	
B8	F: ACGCGCAGGGAATGGATAAT	6360	6379	
	R: ACCATAGCTGAAATCGGGGC	7407	7388	
B9	F: TGGATTGGCTGCAATCATGC	7288	7307	
	R: ATGACGCGCACTACAGTCAA	8359	8340	
B10	F: TGCAACTGCAGAAGCTGAAC	8104	8123	
	R: TTTCACAAGTGCCGTGCCTA	9236	9217	
B11	F: TCAGCTTGTGTTTTGGCTGC	8996	9015	
	R: CGTCATCAAGCCAAAGACCG	10157	10138	
B12	F: GCTGCTTGTTGTCATCTCGC	9950	9969	
	R: CAACCAGTGGTGTGTACCCT	11014	10995	
B13	F: CCTCTTTCTGCTCAAACTGGAA	10808	10829	
	R: GTGTAACTGGACACATTGAGCC	11950	11929	
B14	F: TCACAGGGACTACTCCCACC	11738	11757	
	R: TGTAGTACCGGCAGCACAAG	12742	12723	
B15	F: AGGGCCAATTCTGCTGTCAA	12659	12678	
	R: AGACGAGGTCTGCCATTGTG	13825	13806	
B16	F: GTGGGGGACAACCAATCACT	13122	13141	
	R: CAGGAACTCCACTACCTGGC	14137	14118	
B17	F: TGATGCCATGCGAAATGCTG	14019	14038	
	R: GCTACGGTGCGAGCTCTATT	15113	15094	
B18	F: ACCAAGTCATCGTCAACAACCT	14912	14933	
	R: ATCTACAAAACAGCCGGCCC	15975	15956	
B19	F: TGTTGGACTGAGACTGACCTT	15835	15855	

	R: CGGTAAACAACAGCATCACCA	16871	16851	
B20	F: TGCAGCAGAAACGCTCAAAG	16635	16654	
	R: CTCCAAGCAGGGTTACGTGT	17756	17737	
B21	F: TGTTCCTCGGAACTTGTCGG	17543	17562	
	R: CAACTCAAAGCCATGTGCCC	18615	18596	
B22	F: TAGTGCTAAACCACCGCCTG	18447	18466	
	R: TGTCTACAGACAGCACCACC	19496	19477	
B23	F: CATGCATTCCACACACCAGC	19309	19328	
	R: TGAACCTGTTTGCGCATCTG	20481	20462	
B24	F: AGGCTATGCCTTCGAACATATC	20304	20325	
	R: CGCGTGGTTTGCCAAGATAA	21307	21288	
B25	F: CTTGGAGGTTCCGTGGCTAT	21145	21164	S
	R: AAACCCTGAGGGAGATCACG	22221	22202	
B26	F: TATCTTGGCAAACCACGCGA	21289	21308	S
	R: ACCAGCTGTCCAACCTGAAG	22345	22326	
B27	F: CCCTCAGGGTTTTTCGGCTT	22210	22229	S
	R: CTGTGGATCACGGACAGCAT	23302	23283	
B28	F: CCAGCAACTGTTTGTGGACC	23123	23142	S
	R: GTGGCAAAACAGTAAGGCCG	24149	24130	
B29	F: ACTTGCAGATGCTGGCTTCA	24043	24062	S
	R: CTCATTGAGGCGGTCAATTTCT	25126	25105	
B30	F: TGATTTAGGTGACATCTCTGGCA	25054	25076	S
	R: ACAACTCCGGATGAACCGTC	26159	26140	
B31	F: GCTGGCTTGATTGCCATAGT	25226	25245	S, ORF3
	R: ACAATCGAAGCGCAGTAAGG	26362	26343	
B32	F: ACGGTTCATCCGGAGTTGTT	26141	26160	E
	R: TAGTACCGTTGGAATCTGCCAT	26544	26523	
B33	F: TCGATTGTGTGCGTACTGCT	26355	26374	M
	R: AGCTCACAAGTAGCGAGTGT	27443	27424	
B34	F: ACAGTCGCTACAGGATTGGC	27109	27128	M, ORF6, 7, 8
	R: CCCACTGCGTTCTCCATTCT	28375	28356	
B35	F: GAATTGTGCGTGGATGAGGC	28068	28087	N
	R: CGTTCCCGAAGGTGTGACTT	29260	29241	
B36	F: CTACGCAGAAGGGAGCAGAG	28786	28805	N, ORF9b
	R: TAGGCAGCTCTCCCTAGCAT	29790	29771	
3’RACE-1	GTCTACTCTTGTGCAGAATG	29601	29620	UTR
3’RACE-2	GAGAGCTGCCTATATGGAAG	29779	29798	

*The annealing temperature is 50℃.