






# A Novel Potentially Recombinant Rodent Coronavirus with a Polybasic Cleavage Site in the Spike Protein

Xin Li,<sup>a,b</sup> Liang Wang,<sup>c</sup> Peipei Liu,<sup>b</sup> Hongying Li,<sup>d</sup> Shuting Huo,<sup>a,b</sup> Kexin Zong,<sup>a,b</sup> Shiyan Zhu,<sup>a,b</sup> Yuanyuan Guo,<sup>b,e</sup> Libiao Zhang,<sup>f</sup> Ben Hu,<sup>g</sup> Yu Lan,<sup>b</sup>  Aleksei Chmura,<sup>d</sup> Guizhen Wu,<sup>b</sup> Peter Daszak,<sup>d</sup>  William J. Liu,<sup>a,b,h</sup>  George F. Gao<sup>a,b,c</sup>

<sup>a</sup>School of Laboratory Medicine and Life Sciences, Wenzhou Medical University, Wenzhou, China

<sup>b</sup>NHC Key Laboratory of Biosafety, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China

<sup>c</sup>CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China

<sup>d</sup>EcoHealth Alliance, New York, New York, USA

<sup>e</sup>School of Pharmaceutical Sciences, Nanjing Tech University, Nanjing, China

<sup>f</sup>Institute of Zoology, Guangdong Academy of Sciences, Guangzhou, China

<sup>g</sup>CAS Key Laboratory of Special Pathogens and Biosafety, Wuhan Institute of Virology, Chinese Academy of Sciences, Wuhan, China

<sup>h</sup>Research Units of Adaptive Evolution and Control of Emerging Viruses, Chinese Academy of Medical Sciences, Beijing, China

Xin Li, Liang Wang, Peipei Liu, and Hongying Li contributed equally to this work. Author order was determined by the most substantial contribution to the article draft.

**ABSTRACT** The emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has reignited global interest in animal coronaviruses and their potential for human transmission. While bats are thought to be the wildlife reservoir of SARS-CoV and SARS-CoV-2, the widespread human coronavirus OC43 is thought to have originated in rodents. Here, we sampled 297 rodents and shrews, representing eight species, from three municipalities of southern China. We report coronavirus prevalences of 23.3% and 0.7% in Guangzhou and Guilin, respectively, with samples from urban areas having significantly higher coronavirus prevalences than those from rural areas. We obtained three coronavirus genome sequences from *Rattus norvegicus*, including a *Betacoronavirus* (rat coronavirus [RCoV] GCCDC3), an *Alphacoronavirus* (RCoV-GCCDC5), and a novel *Betacoronavirus* (RCoV-GCCDC4). Recombination analysis suggests that there was a potential recombination event involving RCoV-GCCDC4, murine hepatitis virus (MHV), and Longquan RI rat coronavirus (LRLV). Furthermore, we uncovered a polybasic cleavage site, RARR, in the spike (S) protein of RCoV-GCCDC4, which is dominant in RCoV. These findings provide further information on the potential for interspecies transmission of coronaviruses and demonstrate the value of a One Health approach to virus discovery.

**IMPORTANCE** Surveillance of viruses among rodents in rural and urban areas of South China identified three rodent coronaviruses, RCoV-GCCDC3, RCoV-GCCDC4, and RCoV-GCCDC5, one of which was identified as a novel potentially recombinant coronavirus with a polybasic cleavage site in the spike (S) protein. Through reverse transcription-PCR (RT-PCR) screening of coronaviruses, we found that coronavirus prevalence in urban areas is much higher than that in rural areas. Subsequently, we obtained three coronavirus genome sequences by deep sequencing. After different method-based analyses, we found that RCoV-GCCDC4 was a novel potentially recombinant coronavirus with a polybasic cleavage site in the S protein, dominant in RCoV. This newly identified coronavirus RCoV-GCCDC4 with its potentially recombinant genome and polybasic cleavage site provides a new insight into the evolution of coronaviruses. Furthermore, our results provide further information on the potential for interspecies transmission of coronaviruses and demonstrate the necessity of a One Health approach for zoonotic disease surveillance.

**KEYWORDS** coronavirus, rodents, genomics, recombination, polybasic cleavage site

**Citation** Li X, Wang L, Liu P, Li H, Huo S, Zong K, Zhu S, Guo Y, Zhang L, Hu B, Lan Y, Chmura A, Wu G, Daszak P, Liu WJ, Gao GF. 2021. A novel potentially recombinant rodent coronavirus with a polybasic cleavage site in the spike protein. *J Virol* 95:e01173-21. <https://doi.org/10.1128/JVI.01173-21>.

**Editor** Colin R. Parrish, Cornell University

**Copyright** © 2021 American Society for Microbiology. All Rights Reserved.

Address correspondence to Peter Daszak, [daszak@ecohealthalliance.org](mailto:daszak@ecohealthalliance.org), William J. Liu, [liujun@ivdc.chinacc.cn](mailto:liujun@ivdc.chinacc.cn), or George F. Gao, [gaof@im.ac.cn](mailto:gaof@im.ac.cn).

**Received** 13 July 2021

**Accepted** 13 August 2021

**Accepted manuscript posted online**  
25 August 2021

**Published** 27 October 2021

Seven coronaviruses (CoVs) are known to cause respiratory disease in human populations, with symptoms ranging from mild to severe. These include human coronavirus (HCoV) HKU1, HCoV-NL63, HCoV-OC43, HCoV-229E (1–4), severe acute respiratory syndrome CoV (SARS-CoV), Middle East respiratory syndrome CoV (MERS-CoV), and severe acute respiratory syndrome CoV 2 (SARS-CoV-2). These viruses represent a significant threat to public health due to their contribution to seasonal respiratory and emerging pandemic diseases (5–7). Beyond this health burden, the emergences of SARS, MERS, and coronavirus disease 2019 (COVID-19) have also led to enormous economic costs at a global level. CoVs have been reported in domestic and wild mammal, bird, and reptile species (8–13), but all seven known human pathogens are thought to have their origins in mammals.

Rodents (order Rodentia), comprising the largest group of mammalian species (14), are widely distributed globally and are known to be the reservoirs of several zoonotic viruses (15). The prototypic murine hepatitis virus strain 1 (MHV-1) is a *Betacoronavirus* that was first identified and isolated in mice in 1949 (16), and it is one of the recognized animal models for multiple sclerosis (17). Although a number of diverse viruses have recently been described in rodents from China (18), most research on wildlife CoVs has focused on bat hosts due to their role as reservoirs of SARS-CoV, and likely also of MERS-CoV and SARS-CoV-2 (7, 19, 20).

Although the natural host of SARS-CoV-2 remains unidentified, evidence implies the origin of SARS-CoV-2 in wild animal reservoirs, such as pangolins (11) and bats (21), with the potential for introduction by contamination of imported cold-chain products (22). The HCoVs OC43 and HKU1 are believed to have originated from rodent-associated viruses (23). Given the close contact between humans and free-ranging rodents around homes and workplaces and the relative paucity of research on rodent CoVs, we conducted a One Health project to analyze the presence of potentially zoonotic CoVs in rodents and shrews across a range of habitats in southern China.

As an RNA virus with the largest genome at approximately 30 kb (24), CoV has been categorized by the International Committee on Taxonomy of Viruses (ICTV) into four genera, *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus*. Among these, betacoronaviruses are classified into lineages A, B, C, and D (25). HCoVs OC43 and HKU1 belong to lineage A, whereas SARS-CoV/SARS-CoV-2 and MERS-CoV are in lineages B and C, respectively. CoVs undergo complex replication and transcription processes. The RNA replication of CoVs is a continuous process that produces a genome-length template. In contrast, the transcription of CoV RNA relies on the discontinuous synthesis of negative RNA strands to produce a set of sub-genome-length mRNAs (26, 27). Approximately two-thirds of the 5′-proximal region of the CoV genome, divided into open reading frame 1a (ORF1a) and ORF1b, encodes nonstructural proteins (NSPs) involved in proteolysis and viral replication. The remaining one-third of the genome from 5′ to 3′ encodes four structural proteins, namely, spike (S), envelope (E), membrane (M), and nucleocapsid (N) (28). Some CoVs also contain other structural proteins, such as hemagglutinin esterase (HE) located upstream of the S gene in MHV-1 and in HCoVs OC43 and HKU1 in lineage A of *Betacoronavirus* (24). The S protein, which consists of approximately 1,300 amino acids, is divided into S1 and S2. S1 plays an important role in host cell receptor recognition, whereas S2 facilitates membrane fusion (29). There is a polybasic cleavage site between S1 and S2 in some CoVs, which enhances receptor binding efficacy and membrane fusion to form syncytia (30). In this way, the cleavage site can facilitate virus entry to cells and enable propagation. Moreover, this cleavage site is relatively conserved in some contagious viruses, representing a potentially ideal target to design universal drugs to protect from infection. For SARS-CoV-2, the cleavage site is used as a target for peptides and some antiviral compounds, which could reduce the binding of the receptor binding domain (RBD) to ACE2, inhibiting S2 production and syncytium formation (30, 31).

Here, we surveyed the prevalence of coronaviruses from rodents in southern China and identified three, namely, *Betacoronavirus* rat coronavirus (RCoV) GCCDC3 and

**TABLE 1** Prevalence of CoV in rodents and shrews in Guangdong Province and Guangxi Zhuang Autonomous Region, China

| Species                       | No. of CoV-positive samples/total no. of samples from <sup>a</sup> : |                   |       |      |                 | All locations |
|-------------------------------|--|-------------------|-------|------|-----------------|---------------|
|                               | Guilin (farm)  | Guangzhou         |       |      | Meizhou (field) |               |
| Hotel                         |  | Passenger station | Field |      |                 |               |
| <i>Rattus norvegicus</i>      | 0/10   | 11/29             | 17/47 | 5/59 | —               | 33/145        |
| <i>Rhizomys pruinosus</i>     | 1/121  | —                 | —     | —    | 0/2             | 1/123         |
| <i>Rattus pyctoris</i>        | —  | 0/4               | —     | 1/2  | —               | 1/6           |
| <i>Rattus tanezumi</i>        | 0/2  | 0/1               | —     | 0/3  | —               | 0/6           |
| <i>Niviventer confucianus</i> | —  | —                 | 0/1   | —    | 0/1             | 0/2           |
| <i>Niviventer huang</i>       | 0/1  | —                 | —     | —    | 0/2             | 0/3           |
| <i>Niviventer</i> sp. 1       | —  | —                 | —     | —    | 0/9             | 0/9           |
| <i>Crociodura tanakae</i>     | —  | —                 | —     | —    | 0/3             | 0/3           |
| All species                   | 1/134  | 11/34             | 17/48 | 6/64 | 0/17            | 35/297        |

<sup>a</sup>—, No rodents captured.

RCoV-GCCDC4 and *Alphacoronavirus* RCoV-GCCDC5. RCoV-GCCDC4 was recognized as a novel CoV with unique characteristics. Our study reveals the epidemiology of the coronaviruses among humans in contact with rodents that may serve as reservoirs for zoonotic CoVs.

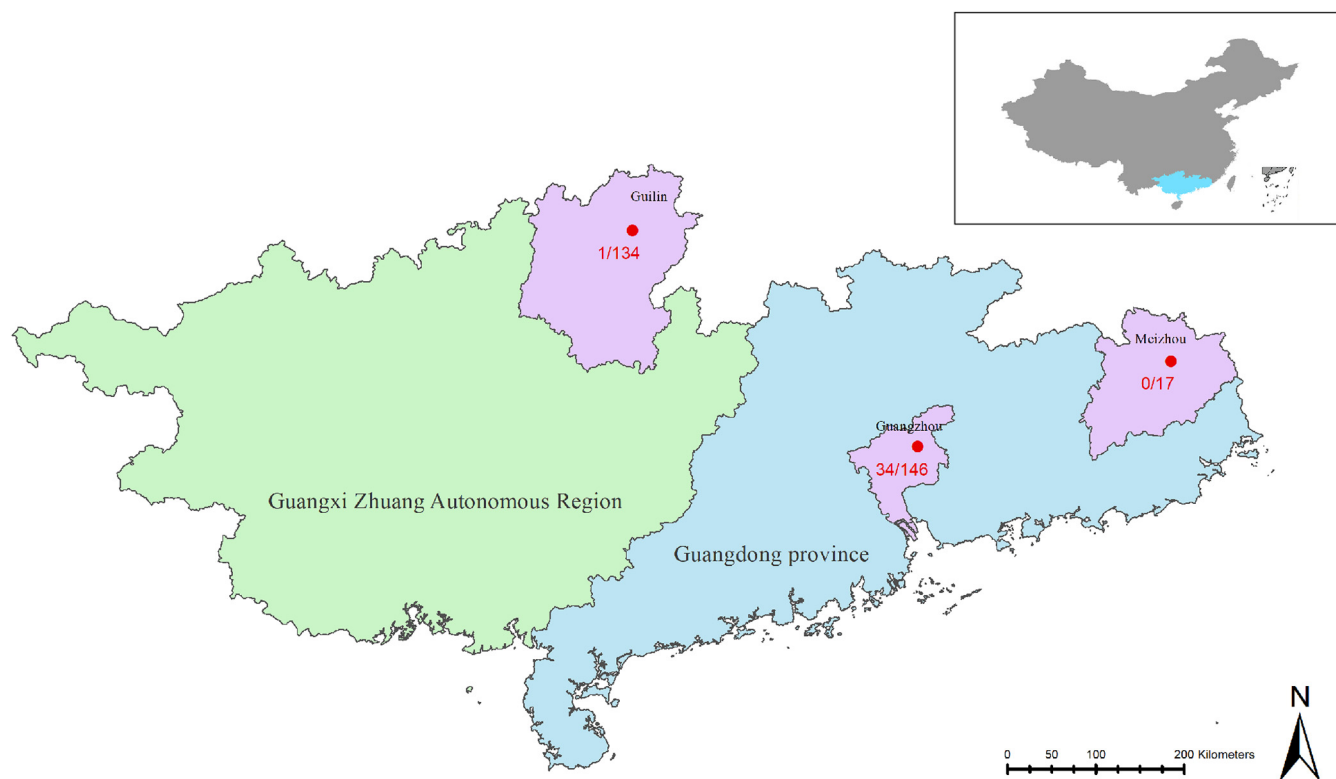
## RESULTS

**CoV prevalence in rodents and shrews.** We collected samples from 297 rodents representing eight species in three municipalities in Guangdong Province and the Guangxi Zhuang Autonomous Region of China, comprising 134 rodents from Guilin and Guangxi, 146 from Guangzhou and Guangdong, and 17 from Meizhou, Guangdong (Table 1 and Fig. 1). All captured animals belonged to three families, namely, *Muridae*, *Rhizomyidae*, and *Soricidae*.

Viral RNA was extracted from oral and anal swabs to screen for CoVs using reverse transcription-PCR (RT-PCR) with universal degenerate primers (32). The partial RNA-dependent RNA polymerase (RdRp) sequences obtained from RT-PCR were searched against the National Center for Biotechnology Information (NCBI) database. In total, 35 rodents were identified as CoV positive, including 33 of 145 (22.8%) *Rattus norvegicus* individuals, one of six (16.7%) *Rattus pyctoris* individuals, and one of 123 (0.8%) *Rhizomys pruinosus* individuals. Of the CoV-positive rodents, 97% (34 of 35), were from populated places such as hotels (11/35; 31%) and passenger stations (17/35; 49%) in Guangzhou; only one was from a bamboo rat farm (1/35; 3%) (Table 1 and Fig. 1). Using a chi-squared test, the prevalence of CoVs was found to be highly significantly greater in populated sites compared to that in rural regions ( $\chi^2 = 36.55$ ;  $P < 0.005$ ).

**Virus isolation.** We attempted to isolate virus from CoV-positive samples using African green monkey kidney cells (Vero-E6) and human epithelial colorectal adenocarcinoma (Caco-2) cells. Cell lines were inoculated with CoV-positive samples and cultivated in a constant-temperature incubator. Each sample was cultivated in three blind passages with or without trypsin treatment. No cytopathic effect was observed in any passage of any sample. RNA extraction from the culture supernatant of each cell line passage was tested by RT-PCR, and supernatants and cell lysates were analyzed with an electron microscope. Virus isolation was unsuccessful.

**Identification of three CoVs in rodents.** CoV-positive samples were identified through RT-PCR with universal degenerate primers (32). RNA of corresponding samples was analyzed using next-generation sequencing. When we calculated the coverage ratio by mapping the clean reads to assembled virus genomes of the three viruses, the coverage ratios for the genomes of RCoV-GCCDC3, RCoV-GCCDC4, and RCoV-GCCDC5 were all 100%, with good sequencing depths (Fig. 2). We obtained the full-length genomic sequence of RCoV-GCCDC3 with the 5' and 3' ends determined (Fig. 3) and the nearly full-length genome sequences of RCoV-GCCDC4 and RCoV-GCCDC5. Of the three viruses, RCoV-GCCDC3 and RCoV-GCCDC5 displayed high similarity to known CoVs, according to a Basic Local Alignment Search Tool (BLAST) search. The genomic



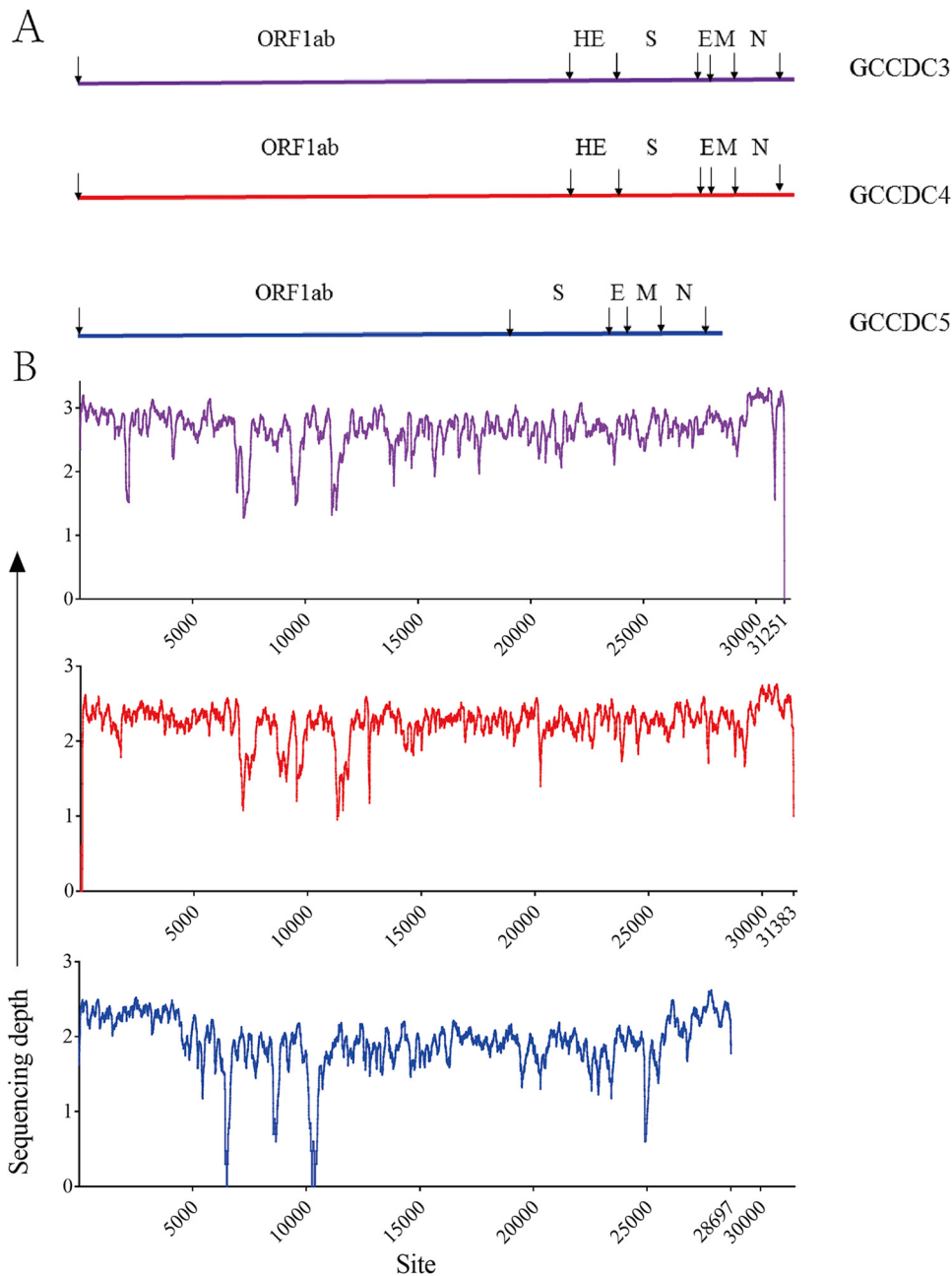
**FIG 1** Sampling locations and positivity rates of rat coronaviruses (RCoVs). Green, Guangxi Zhuang Autonomous Region; blue, Guangdong Province; purple, the cities of Guilin, Guangzhou, and Meizhou; red, positive proportions.

sequence of RCoV-GCCDC4 exhibited low similarity to known CoVs, suggesting that it is a novel CoV. According to the BLAST analysis, the whole-genome sequence of RCoV-GCCDC3 had a 96% similarity to that of China *Rattus* CoV HKU24 (HKU24), which supports a murine origin of *Betacoronavirus* 1. RCoV-GCCDC5 was 97% identical to Lucheng Rn rat CoV (LRNV), a recombinant CoV identified in Zhejiang Province.

**Phylogenetic analysis of RCoVs.** We next determined the phylogenetic relationships between RCoV-GCCDC3, RCoV-GCCDC4, RCoV-GCCDC5, and other existing CoVs. We constructed phylogenetic trees based on the nucleotide sequences encoding polyproteins 1a and 1b and the S, E, M, and N proteins. The phylogenetic trees of these six proteins showed that RCoV-GCCDC3 and RCoV-GCCDC4 belonged to *Betacoronavirus* genus lineage A, whereas RCoV-GCCDC5 belonged to the *Alphacoronavirus* genus. RCoV-GCCDC3 and RCoV-GCCDC4 were located in two separate clades within lineage A of the *Betacoronavirus* genus. Notably, in the S, E, M, and N gene trees, RCoV-GCCDC4 formed a monophyletic group with Longquan RI rat coronavirus (LRLV). However, in 1a and 1b gene trees, RCoV-GCCDC4 clustered with some rodent-derived viruses, such as MHV-1 and rat CoV. RCoV-GCCDC5 and LRNV formed a distinct clade in the *Alphacoronavirus* genus, which includes CoVs discovered in humans, pigs, and other animals (Fig. 4).

**Recombination analysis of RCoV-GCCDC4.** We conduct a detailed sequence analysis to investigate possible virus recombination events. The genomes of RCoV-GCCDC4, MHV-1, and LRLV were aligned and examined for recombination using SimPlot and RDP4. Similarity plots indicated that the putative recombination site was located at nucleotide position “20, 170,” which separated the genome of RCoV-GCCDC4 into two parts. The similarity of LRLV to RCoV-GCCDC4 in the region of the structural and accessory genes was significantly higher, at 95.3%, than that of MHV-1 to RCoV-GCCDC4, at 81.6% (Fig. 5A). The same recombination event was confirmed by RDP4 (Fig. 5B).

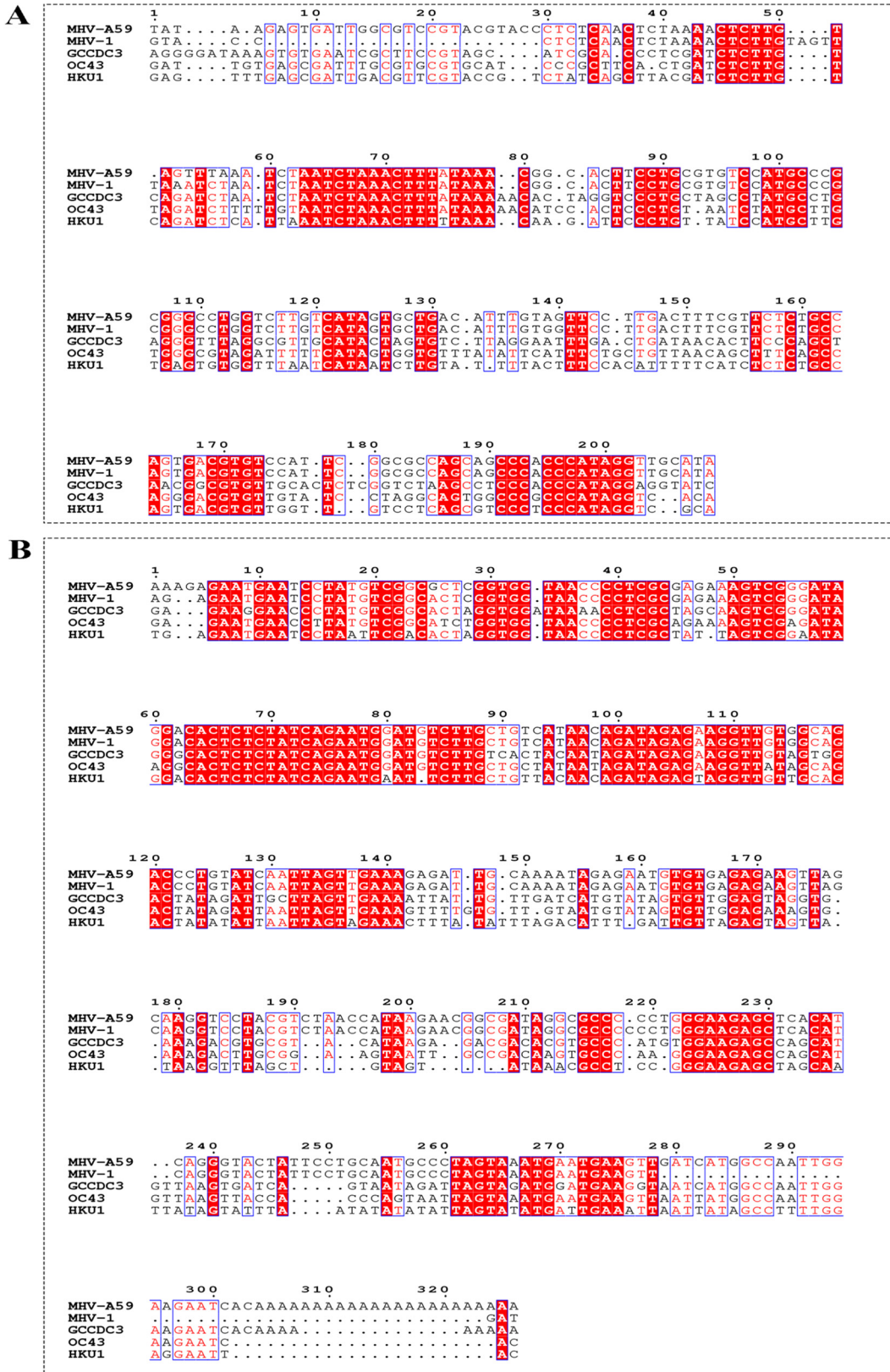
**Genomic organization of RCoV-GCCDC4.** The size of the RCoV-GCCDC4 genome was determined to be 31,383 bp, with a G+C content of 40.13%. Like other A lineage betacoronaviruses, the genome of RCoV-GCCDC4 had the same gene order, that is,

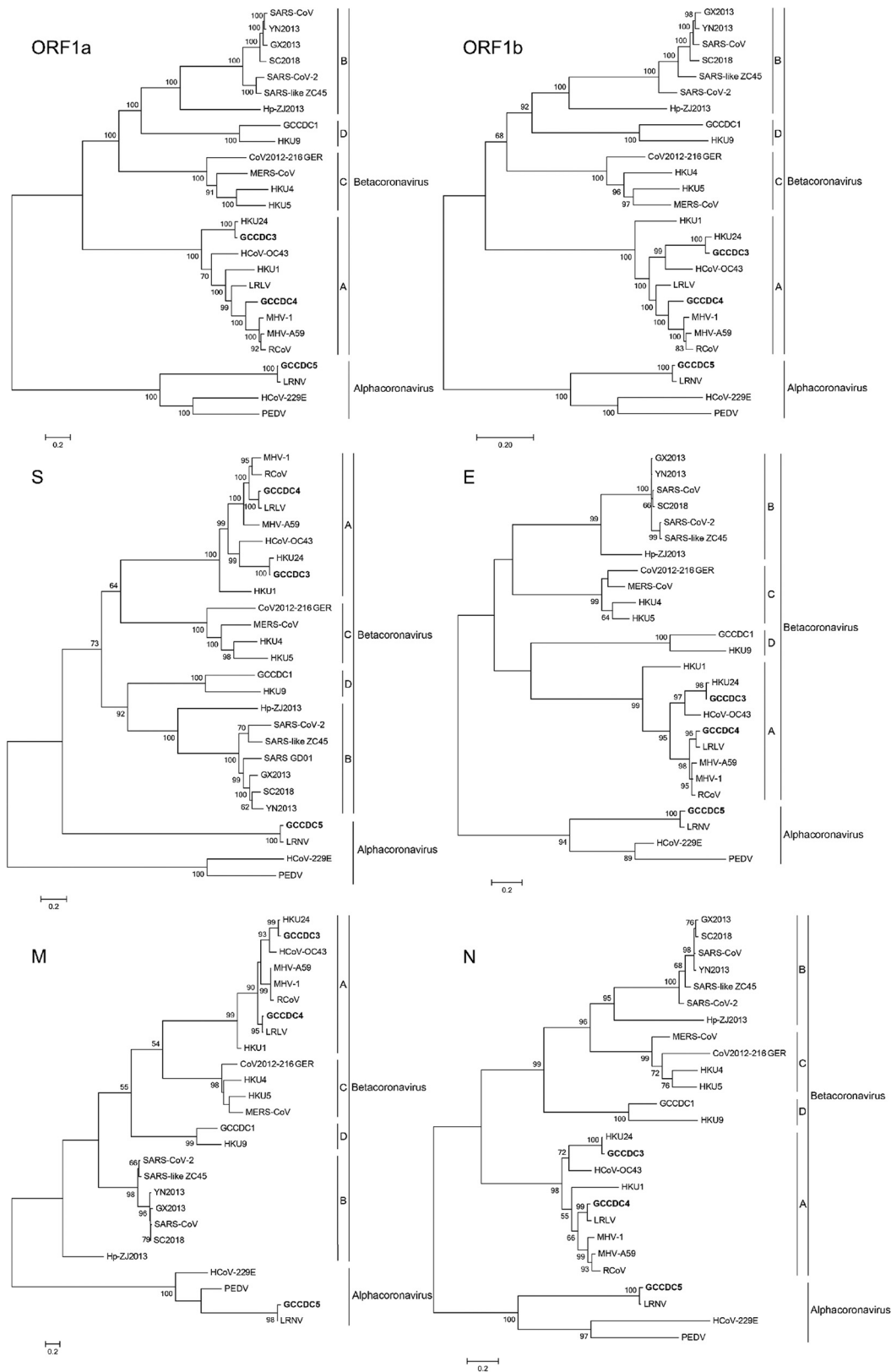


**FIG 2** Sequencing depths of three RCoV genomes identified in this study. The ordinate value represents the index. (A) Schematic diagrams for the genomes of RCoV-GCCDC3, RCoV-GCCDC4, and RCoV-GCCDC5. (B) Sequencing depths of RCoV-GCCDC3, RCoV-GCCDC4, and RCoV-GCCDC5. Purple, red, and blue represent RCoV-GCCDC3, RCoV-GCCDC4, and RCoV-GCCDC5, respectively.

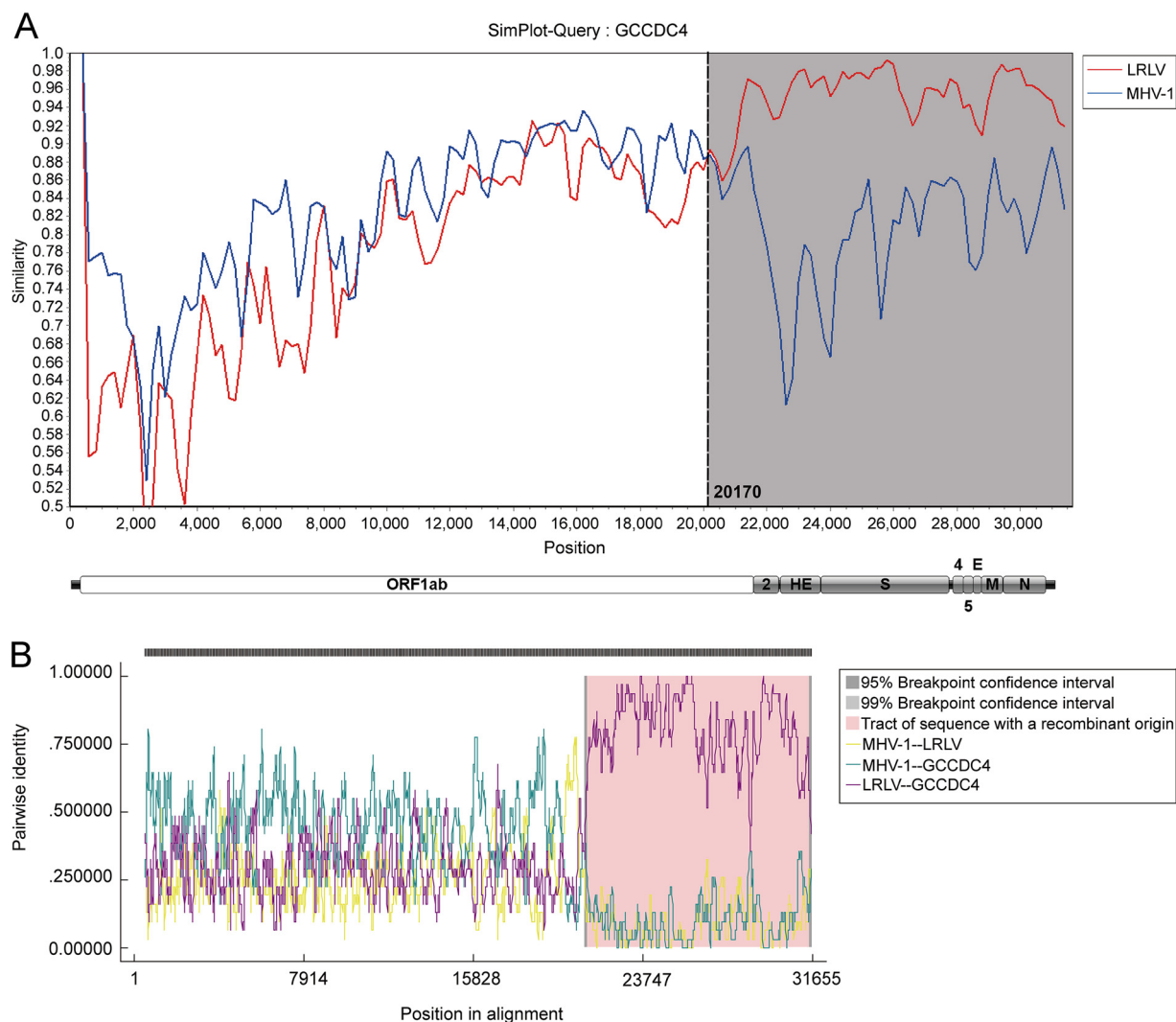
5'-ORF1ab-HE-S-E-M-N-3'. Furthermore, we found several genes encoding accessory proteins such as NS2a, NS5a, and NS5b (Table 2 and Fig. 6A). The amino acid sequence of the most conserved M protein was 37.99% to 86.40% identical to those of known betacoronaviruses. Conversely, the similarity of the N protein to that in representative betacoronaviruses ranged from 27.09% to 86.18%.

Following the criteria for CoV species defined by the International Committee on Taxonomy of Viruses (ICTV) (24), we selected seven conserved replicase domains of RCoV-GCCDC3, RCoV-GCCDC4, and RCoV-GCCDC5 for further analysis (33, 34), namely, ADP-ribose 1-phosphatase, chymotrypsin-like protease, RdRp, helicase, exoribonuclease, nidoviral endoribonuclease specific for uridylylate, and ribose-2'-O-methyltransferase.





**FIG 4** Phylogenetic analyses of the nucleotide sequences of the three identified CoVs and other representative CoVs. The nucleotide sequences of open reading frame 1a (ORF1a), ORF1b, S, E, M, and N in RCoV-GCCDC3, RCoV-GCCDC4, RCoV-GCCDC5, and other representative CoVs were analyzed. All trees were constructed by the maximum-likelihood method in MEGA. Bootstrap values above 50% are shown. The three RCoV sequences acquired in this study are shown in bold. The accession numbers of CoVs used in this analysis are shown in Table 7.



**FIG 5** Recombination within the genome of RCoV-GCCDC4. (A) Recombination analysis through SimPlot. Curves display whole-genome similarities of murine hepatitis virus strain 1 (MHV-1) (blue) and Longquan RI rat coronavirus (LRLV) (red) with RCoV-GCCDC4 (“query”), respectively. The dashed line shows the recombination site, and the location is indicated at the bottom. The background color of sequence with recombination origin is gray. (B) Recombination analysis through RDP4. Curves display whole-genome pairwise identity of MHV-1-LRLV (yellow), MHV-1-GCCDC4 (cyan), and LRLV-GCCDC4 (purple). The pink background indicates the tract of sequence with a recombinant origin.

The amino acid sequences of the seven concatenated domains in RCoV-GCCDC4 were less than 90% similar to those in the known CoVs. According to the ICTV criteria of CoV species demarcation, RCoV-GCCDC4 likely represents a novel CoV in the *Betacoronavirus* genus (Table 3). However, RCoV-GCCDC3 and RCoV-GCCDC5 did not demonstrate sufficient sequence divergence to represent novel species.

**Critical sites within the genome of RCoV-GCCDC4.** We predicted the S protein of RCoV-GCCDC4, consisting of 1,361 amino acids, to be a class I viral fusion protein, including primarily an N-terminal domain (NTD), a C-terminal domain (CTD), fusion peptide (FP), and two heptad repeats (HR1/HRN and HR2/HRC). A notable feature of RCoV-GCCDC4 was its polybasic cleavage site (RARR) within the S protein. The cleavage site between S1 and S2 was in residue 756, contributing to protease activation (Fig. 6B). The corresponding amino acid sequence in MHV-A59 is RADR. However, furin-like protease prefers K/R in the third position. Therefore, the S protein of RCoV-GCCDC4 possesses higher cleavage activity and efficiency (35).

The putative leader and body transcription regulatory sequences (TRS) and their genomic locations in RCoV-GCCDC4 were predicted by comparing them with those of



**TABLE 2** Similarities between the genes of RCoV-GCCDC4 and those of other representative CoVs

| Coronaviruses                    | Genome feature |                 | Pairwise amino acid sequence identities between GCCDC4 and other coronaviruses (%) <sup>a</sup> |       |       |       |       |       |       |       |
|----------------------------------|----------------|-----------------|---|-------|-------|-------|-------|-------|-------|-------|
|                                  | Size (nt)      | G+C content (%) | 3CLpro  | RdRp  | HEL   | HE    | S     | E     | M     | N     |
| <i>Betacoronavirus</i> lineage A |                |                 |   |       |       |       |       |       |       |       |
| HCoV-HKU1                        | 29,926         | 32.06           | 85.48   | 91.27 | 82.87 | 49.53 | 64.10 | 53.41 | 78.95 | 69.67 |
| HCoV-OC43                        | 30,741         | 36.69           | 83.83   | 90.41 | 84.10 | 57.70 | 64.37 | 64.44 | 83.55 | 71.77 |
| BCoV                             | 31,028         | 37.02           | 84.82   | 91.06 | 84.40 | 59.08 | 65.07 | 63.33 | 83.12 | 71.83 |
| ECoV                             | 30,992         | 37.23           | 84.82   | 91.06 | 85.02 | 57.60 | 66.88 | 63.33 | 83.55 | 70.83 |
| CRCoV                            | 31,028         | 37.01           | 84.49   | 90.84 | 84.25 | 59.08 | 64.64 | 62.22 | 83.98 | 71.33 |
| MHV-A59                          | 31,357         | 41.78           | 94.72   | 96.55 | 88.38 | 59.11 | 78.42 | 78.41 | 86.40 | 84.68 |
| RCoV                             | 31,250         | 41.26           | 93.40   | 96.44 | 88.38 | 80.41 | 86.56 | 90.91 | 84.65 | 86.18 |
| LRLV                             | 30,821         | 39.06           | 91.42   | 95.04 | 85.02 | 93.88 | 95.61 | 96.59 | 98.69 | 94.74 |
| RabbitCoV-HKU14                  | 31,100         | 37.64           | 88.12   | 90.95 | 84.86 | 58.39 | 65.55 | 67.78 | 84.42 | 72.43 |
| RatCoV-HKU24                     | 31,249         | 40.07           | 83.17   | 89.98 | 83.33 | 55.73 | 63.28 | 61.36 | 83.98 | 68.71 |
| DcCoV-HKU23                      | 31,052         | 36.96           | 84.82   | 90.95 | 84.10 | 58.85 | 65.08 | 63.33 | 84.42 | 71.62 |
| <i>Betacoronavirus</i> lineage B |                |                 |   |       |       |       |       |       |       |       |
| SARS-CoV                         | 29,751         | 40.80           | 50.33   | 65.59 | 61.98 |       | 27.27 | 25.00 | 37.99 | 30.00 |
| SARS-CoV-2                       | 29,903         | 37.97           | 50.33   | 65.70 | 61.98 |       | 26.46 | 25.00 | 39.30 | 29.72 |
| BatCoV-HKU3                      | 29,728         | 41.15           | 50.00   | 65.45 | 62.14 |       | 26.74 | 25.00 | 38.43 | 30.00 |
| <i>Betacoronavirus</i> lineage C |                |                 |   |       |       |       |       |       |       |       |
| MERS-CoV                         | 30,119         | 41.18           | 53.59   | 68.20 | 61.68 |       | 29.29 | 26.14 | 42.11 | 30.89 |
| Ty-BatCoV HKU4                   | 30,286         | 37.82           | 53.27   | 68.09 | 62.14 |       | 28.89 | 26.14 | 42.11 | 29.74 |
| <i>Betacoronavirus</i> lineage D |                |                 |   |       |       |       |       |       |       |       |
| Ro-BatCoV HKU9                   | 29,114         | 41.05           | 49.67   | 65.77 | 62.29 |       | 26.47 | 22.73 | 41.67 | 27.09 |
| Ro-BatCoV GCCDC1                 | 30,161         | 45.40           | 47.06   | 65.77 | 61.07 |       | 26.67 | 22.73 | 39.47 | 28.97 |

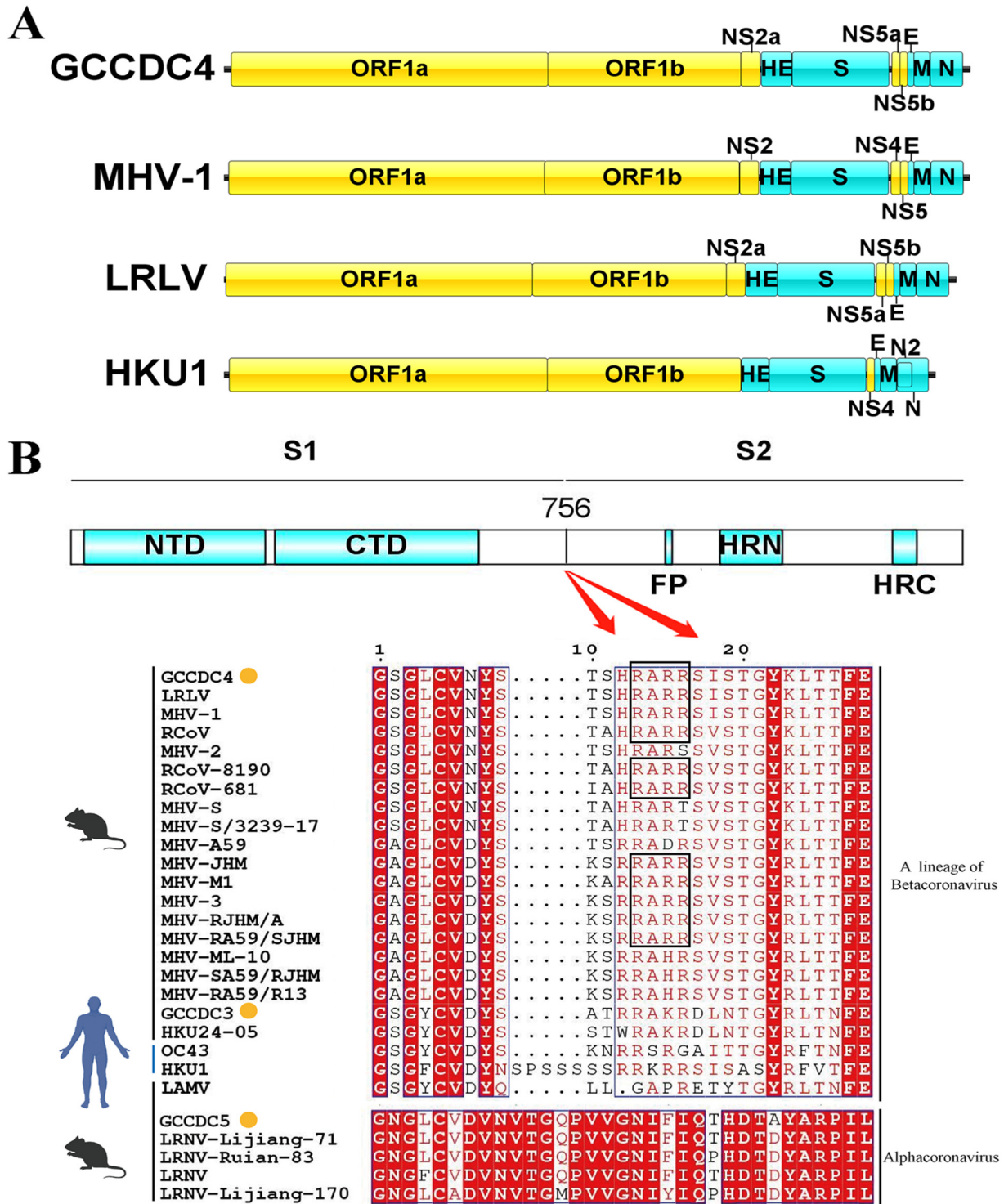
<sup>a</sup>3CLpro, 3-chymotrypsin-like protease; RdRp, RNA-dependent RNA polymerase.

other *Betacoronavirus*. A putative TRS motif, 5'-CUAAAC-3', was found upstream of each ORF, except for the NS2a, HE, E, and M genes in the genome of RCoV-GCCDC4. Moreover, the NS2a, HE, E, and M genes had variant TRS, namely, 5'-GUAAAG-3', 5'-CUAAAA-3', 5'-CUAAAA-3', and 5'-AUAAUC-3', respectively (Table 4).

We also predicted NSPs in ORF1ab of RCoV-GCCDC4 encoding the replicase. The protein sizes, amino acid positions, and cleavage sites of NSP1 to NSP16 were determined (Table 5). Previous studies reported that the P1 position cleavage sites in the peptide sequence of 3-chymotrypsin-like protease (3CLpro) were exclusively occupied by glutamine (Q) (36, 37). However, at nucleobase positions 18338 to 18340, the codon was ATA, resulting in a Q-to-I mutation. Additionally, there was no Q residue in the sequence from -318 to +153 around the site. Therefore, LI|SL may be the potential cleavage site by 3CLpro between NSP13 and NSP14 (Table 5).

## DISCUSSION

In this study, we tested nearly 300 rodents collected from urban and rural sites in southern China. We found that CoV prevalence in rodents from urban areas was much higher than that in rural regions and even higher than those reported in Hong Kong and in Zhejiang Province (<5%) (13, 38). Remarkably, the urban sites with the highest CoV-positive proportions of rodents were passenger stations and hotels with high population density and mobility in Guangzhou. Therefore, there is a potential risk of CoV transmission to humans via close contact. In 2018, swine acute diarrhea syndrome CoV (SADS-CoV) originated from bats, infected piglets, and induced a series of fatal swine disease outbreaks (39). During the current COVID-19 pandemic, transmission of SARS-CoV-2 has emerged between humans and minks (40). Notably, a recent study reported the isolation of a novel canine-feline recombinant *Alphacoronavirus* from a child with pneumonia (41). Thus, considering the increasing transspecies transmissions of CoVs in recent years, we cannot underestimate the potential risk of CoV spillover from animals in close contact with



**FIG 6** Comparison of genomic organization between RCoV-GCCDC4, MHV-1, LRLV, and HKU1, and predicted domain and polybasic cleavage site of the S gene. (A) Predicted structural genes are indicated by blue boxes, and ORFs and nonstructural genes are indicated by yellow boxes. (B) Amino acid sequences of the polybasic cleavage site in different CoVs are presented. Multiple-sequence alignments within the S protein. Residues in red are completely conserved, residues boxed in blue are represent 80% conserved, and common amino acids are shown in black. Black boxes indicate the major cleavage site in lineage A of *Betacoronavirus*. Orange dots represent CoVs identified in this study. The multiple-sequence alignments were generated by ClustalW and ESPrpt. NTD, N-terminal domain; CTD, C-terminal domain; FP, fusion peptide; HRN, heptad repeat N; HRC, heptad repeat C.

**TABLE 3** Pairwise comparison of the conserved domains in replicase polyprotein between RCoV-GCCDC4 and other betacoronaviruses

| RCoV-GCCDC4 <sup>a</sup> | nt <sup>b</sup> positions (start–end) | No. of amino acids | Amino acid identity (%) |       |         |         |        |        |
|--------------------------|---------------------------------------|--------------------|-------------------------|-------|---------|---------|--------|--------|
|                          |                                       |                    | LRLV                    | MHV-1 | MHV-A59 | Rat CoV | GCCDC3 | GCCDC5 |
| NSP3 (ADRP)              | 2784–8738                             | 1,985              | 71.2                    | 79.7  | 77.9    | 78.2    | 54.3   | 12.7   |
| NSP5 (3CLpro)            | 10227–11135                           | 303                | 91.4                    | 93.7  | 94.7    | 93.4    | 83.5   | 24.4   |
| NSP12 (RdRp)             | 13596–16378                           | 928                | 95.0                    | 95.8  | 96.6    | 96.4    | 90.3   | 55.2   |
| NSP13 (HEL, NTPase)      | 16379–18340                           | 654                | 85.0                    | 88.2  | 88.4    | 88.4    | 83.6   | 53.7   |
| NSP14 (ExoN, NMT)        | 18341–19741                           | 467                | 79.3                    | 85.4  | 85.0    | 85.4    | 76.2   | 46.9   |
| NSP15 (NendoU)           | 19742–20863                           | 374                | 89.6                    | 88.2  | 89.3    | 87.4    | 77.3   | 40.6   |
| NSP16 (O-MT)             | 20864–21760                           | 299                | 97.0                    | 90.6  | 92.3    | 91.0    | 86.1   | 57.6   |
| Concatenated domains     |                                       |                    | 84.7                    | 88.3  | 87.9    | 87.8    | 73.8   | 34.4   |

<sup>a</sup>ADRP, ADP-ribose 1-phosphatase; 3CLpro, 3-chymotrypsin-like protease; RdRp, RNA-dependent RNA polymerase; HEL, helicase; ExoN, exoribonuclease; NendoU, nidoviral endoribonuclease specific for uridylylate; O-MT, ribose-2'-O-methyltransferase.

<sup>b</sup>nt, nucleotide.

humans, in addition to that from wildlife. Therefore, effective surveillance and early warning of zoonotic viral diseases is vitally necessary. Daily cleaning, including sterilization, disinfection, and deratization, in passenger stations and hotels, is crucial and necessary to prevent the spillover of CoVs from rodents to humans. Conversely, we only detected one CoV-positive bamboo rat sample from the tested bamboo rat farms that supplied rats for food. Hence, the limited contact between humans and other animals in bamboo rat farms may restrict the transmission of CoVs from breeding bamboo rats compared with that from rodents in urban settings.

To the best of our knowledge, according to the current literature, just one species of rodent CoV was reported before 2014 (42). However, in recent years, an increasing number of rodent-associated CoVs have been detected or isolated (18), suggesting that wild rodents represent a potential reservoir for the spillover of CoVs into people. This is particularly important for species like *Rattus norvegicus*, which inhabits residential and indoor areas and has a high human contact potential. Our findings indicate the significance of virus surveillance in animals in close contact with humans in addition to wildlife, which is important and effective supplementary work for the prevention of emerging infectious diseases.

According to the criteria defined by the ICTV, we have identified a novel CoV, RCoV-GCCDC4. It belongs to lineage A of the *Betacoronavirus* genus, which is most closely related to MHV-1. Notably, recombination analysis shows that there was a potential recombination event between MHV-1 and LRLV, thus implying a more diverse RCoV genome in RCoV-GCCDC4. Although we cannot determine the recombination ancestor within this potential event, the phenomenon implies an important issue, namely, that the RCoVs could exchange one-third of the 3'-proximal region of the genome. This region covers the receptor-binding S protein and also the polybasic cleavage site within it. Thus, RCoV-GCCDC4 might acquire/provide its S protein and also the polybasic cleavage site from/to other CoVs through such recombination to present novel virulence features. As reported

**TABLE 4** Locations of the predicted ORFs in the genome of RCoV-GCCDC4

| ORF  | Location (nt positions) | Length (nt/aa) <sup>b</sup> | Frame(s) | TRS location | TRS sequence(s)<br>(distance to ATG) <sup>a</sup> |
|------|-------------------------|-----------------------------|----------|--------------|---|
| 1ab  | 288–21763               | 21,477/7,157                | +3, +2   | 142          | <b>CUAAAC</b> (140)ATG                            |
| NS2a | 21753–22565             | 813/270                     | +3       | 21720        | <b>GUAAAG</b> (27)ATG                             |
| HE   | 22610–23896             | 1,287/428                   | +2       | 22464        | <b>CUAAAA</b> (140)ATG                            |
| S    | 23905–27990             | 4,086/1,361                 | +1       | 23899        | <b>CUAAAC</b> ATG                                 |
| NS4a | 28110–28463             | 354/117                     | +3       | 28016        | <b>CUAAAC</b> (88)ATG                             |
| NS5a | 28445–28768             | 324/107                     | +2       | 28372        | <b>CUAAAC</b> (67)ATG                             |
| E    | 28761–29027             | 267/88                      | +3       | 28619        | <b>CUAAAA</b> (136)ATG                            |
| M    | 29024–29710             | 687/228                     | +2       | 29010        | <b>AUAAUC</b> (8)ATG                              |
| N    | 29726–31084             | 1,359/452                   | +2       | 29713        | <b>CUAAAC</b> (7)ATG                              |

<sup>a</sup>Boldface indicates putative transcription regulatory sequences (TRS).

<sup>b</sup>aa, amino acid.

**TABLE 5** Prediction of NSP1 to NSP16 and the cleavage sites of polyprotein 1ab of the RCoV-GCCDC4

| NSP   | First and last amino acid positions  | Protein size (aa) | Cleavage site <sup>a</sup> | Function                                      |
|-------|--------------------------------------|-------------------|----------------------------|---|
| NSP1  | M <sup>1</sup> –G <sup>247</sup>     | 247               | LKGYRG VKPILF              | Unknown                                       |
| NSP2  | V <sup>248</sup> –A <sup>832</sup>   | 585               | WRFPCA GKKVGF              | Unknown                                       |
| NSP3  | G <sup>833</sup> –G <sup>2817</sup>  | 1,985             | FSLKGG AILSNL              | ADRP, PL1 <sup>Pro</sup> , PL2 <sup>Pro</sup> |
| NSP4  | A <sup>2818</sup> –Q <sup>3313</sup> | 496               | SFLQ SGIV                  | Hydrophobic domain                            |
| NSP5  | S <sup>3314</sup> –Q <sup>3616</sup> | 303               | VKLQ SKRT                  | 3CLpro  |
| NSP6  | S <sup>3617</sup> –Q <sup>3903</sup> | 287               | SQIQ SRLT                  | Hydrophobic domain                            |
| NSP7  | S <sup>3904</sup> –Q <sup>3992</sup> | 89                | TVLQ ALQS                  | Unknown                                       |
| NSP8  | A <sup>3993</sup> –Q <sup>4189</sup> | 197               | AVLQ NNEL                  | Unknown                                       |
| NSP9  | N <sup>4190</sup> –Q <sup>4299</sup> | 110               | VRLQ AGVA                  | Unknown                                       |
| NSP10 | A <sup>4300</sup> –Q <sup>4436</sup> | 137               | TQFQ SKDT                  | Growth-factor-like                            |
| NSP11 | S <sup>4437</sup> –L <sup>4443</sup> | 7                 | —                          | Unknown (short peptide at the end of ORF1a)   |
| NSP12 | S <sup>4437</sup> –Q <sup>5364</sup> | 928               | AVMQ SVGA                  | RdRp  |
| NSP13 | S <sup>5365</sup> –I <sup>6018</sup> | 654               | SRLI SLMG                  | NTPase/helicase                               |
| NSP14 | S <sup>6019</sup> –Q <sup>6485</sup> | 467               | TKLQ SLEN                  | ExoN  |
| NSP15 | S <sup>6486</sup> –Q <sup>6859</sup> | 374               | PRLQ ASAE                  | NendoU  |
| NSP16 | A <sup>6860</sup> –K <sup>7158</sup> | 299               | —                          | Methyltransferase                             |

<sup>a</sup>Boldface indicates a potential cleavage site. "—" indicates that no cleavage site was found.

previously by Menachery et al., a recombinant bat CoV can replicate in human airway cells and produce obvious pathological changes in the mouse lung, thus showing a potential risk for human infection via recombination (43). Although recombination is common in similar CoV species, it occurs rarely among different CoV species to generate novel viruses. Here, the potential recombination event between RCoV-GCCDC4, MHV, and LRLV may indicate an uncommon recombination between different CoV species. Meanwhile, according to the phylogenetic analysis, LRLV is closely related to human CoVs HKU1 and OC43. The recombination event between LRLV and RCoV-GCCDC4 may increase the potential of cross-species transmission of RCoVs.

Moreover, insertion of a furin-like protease cleavage site at the S1/S2 site of SARS-CoV may enhance virus entry and syncytium formation (44). Efficient cleavage of the MERS-CoV S protein enables the MERS-like CoV to adapt from infecting bats to infecting human cells (45). In avian influenza virus, acquisition of a polybasic site by the hemagglutinin (HA) protein through genomic insertion and recombination enables the virus to change from low to high pathogenicity (46). Notably, SARS-CoV-2 is currently the only CoV in *Betacoronavirus* lineage B that possesses this polybasic cleavage site (47); this characteristic may be the reason that SARS-CoV-2 has higher infectivity and pathogenicity than other identified human-infecting CoVs. Thus, the highly polybasic residue site within the S protein of RCoV-GCCDC4 may be crucial for virulence and the potential expansion of the host range to include humans.

We found that the amino acid sequence of RCoV-GCCDC4 located at the polybasic cleavage site is RARR, whereas the corresponding sequence for MHV-A59 is RADR. For furin-like protease, the preference of amino acid sequence motif RXK/RR shows that RCoV-GCCDC4 may demonstrate more efficient cleavage of the S protein (35). Although this polybasic cleavage site is diverse in different RCoVs, the bi-arginine motif RXXR is important for inducing syncytium formation and production of S2. The side chains of these two arginines at this site were on the protein surface, thus facilitating interactions with proteases for efficient cleavage. Furthermore, this bi-arginine motif is conserved and present in other highly contagious viruses, such as HIV, respiratory syncytial virus (RSV), and some avian influenza viruses, indicating potential pathogenicity and infectiousness of RCoV-GCCDC4 (30). Moreover, this cleavage site is relatively conserved in some contagious viruses and may be an ideal target to design universal drugs to protect from infection. Recently, it has been reported that the polybasic cleavage site can be bound by tetrapeptide EELE to reduce the strength of SARS-CoV-2 RBD binding to ACE2 (31).

During the replicative cycle of MHV, NSP1 is the first mature protein after MHV enters the host cells, and MHV cannot efficiently infect cells following NSP1 deletion in the coding region (48). Moreover, NSP1 is a virulence factor and contributes to host

**TABLE 6** Primers used in this study

| Application of primer | Sequence                             |
|-----------------------|--------------------------------------|
| Barcoding             |                                      |
| CytB_F                | 5'-GAGGMCAAATATCATTCTGAGG-3'         |
| CytB_R                | 5'-TAGGGCVAGGACTCCTCCTAGT-3'         |
| CO1_long_F            | 5'-AACCACAAAGACATTGGCAC-3'           |
| CO1_long_R            | 5'-AAGAATCAGAATARGTGTG-3'            |
| CO1_short_F           | 5'-GCAGGAACAGGWTGAACCG-3'            |
| CO1_short_R           | 5'-AATCAGAAAYAGGTGTTGGTATAG-3'       |
| RT-PCR                |                                      |
| CoV-FWD1              | 5'-CGTTGGIACWAAAYBTCCWYTICARBTRGG-3' |
| CoV-RVS1              | 5'-GGTCATKATAGCRTCAVMASWWGCNACATG-3' |
| CoV-FWD2              | 5'-GGCWCCWCHGGNGARCAATT-3'           |
| CoV-RVS2              | 5'-GGWAWCCCAAYTGYTGWAYRTC-3'         |
| CoV-FWD3              | 5'-GGTTGGGAYTAYCCHAARTGTGA-3'        |
| CoV-RVS3              | 5'-CCATCATCASWYRAATCATCATA-3'        |
| CoV-FWD4              | 5'-GAYTAYCCHAARTGTGAYAGAGC-3'        |

liver pathogenesis during infection (49). It was reported previously that the LLRKxGxKG region (L191-G199) within NSP1 is involved in virulence, innate immune response, and inhibition of cellular gene expression (50). Meanwhile, this pathogenicity factor can be targeted for design of attenuated viral vaccines with low virulence but sound T-cell responses (51). In our study, through multiple-sequence alignments, RCoV-GCCDC3 and RCoV-GCCDC4 NSP1 proteins were found to be relatively conserved in the LLRKxGxKG region. This may indicate potential pathogenicity of these two CoVs and a threat to human health via the wildlife-livestock-human chain.

Our study revealed that distinct alphacoronaviruses and betacoronaviruses circulate in *Rattus norvegicus* populations in urban areas of Guangdong Province, China, especially in places with dense or highly mobile human populations. These findings demonstrate the potential value of a One Health approaches to virus discovery focusing on human-livestock-wildlife interfaces. Monitoring these interfaces through targeted viral discovery efforts may help prevent the emergence and reemergence of viruses with zoonotic potential (52, 53). The newly identified CoV RCoV-GCCDC4, with its features of potential recombination and an efficient cleavage site, provides new insight into the evolution of CoVs and highlights the value of One Health surveillance of potential zoonotic pathogens.

## MATERIALS AND METHODS

**Sample collection.** The study was conducted in three municipalities of Guangzhou, Guilin, and Meizhou in southern China across different environmental contexts, including both urban and rural areas, as well as domesticated wildlife farms where interactions between humans and rodents had been observed. Free-ranging rodents and shrews were captured using metal box traps, set open at sunset and checked at sunrise at selected locations recorded by Global Positioning System (GPS) coordinates and place names. Captured animals were removed from the traps for specimen collection and identified morphologically by experienced biologists, with follow-up confirmation by DNA barcoding. Rectal and oral swabs were collected from all captured individuals and placed into a cryogenic vial with viral transport medium (VTM). Small intestine specimens were collected in the case of accidental death before or during the sampling, or when dead animals were available opportunistically for sampling, and were placed in cryogenic vials with VTM. All samples were temporarily stored in liquid nitrogen before being transported to the laboratory and stored at  $-80^{\circ}\text{C}$ .

**Ethics statement.** Field animal capture and sampling protocols were reviewed and approved by the Institutional Animal Care and Use Committee of UC Davis (protocol no. 19300) and by the Institutional Review Board of the Wuhan Institute of Virology, Chinese Academy of Sciences (approval no. WIVA05201705). No rodent species in the study was a protected species according to the Law of the People's Republic of China on the Protection of Wildlife, nor considered endangered by the Red List of China's Vertebrates or the International Union for Conservation of Nature Red List of Threatened Species.

**Barcoding.** The identity of each species in the study was confirmed using DNA barcoding. DNA was extracted from swabs and small intestine samples using the Ex-DNA kit (Tianlong, Xi'an, China). Nested

**TABLE 7** Selected CoVs and corresponding accession numbers used in this phylogenetic analysis

| Virus strain  | Abbreviation    | GenBank accession no. |
|---|-----------------|-----------------------|
| Severe acute respiratory syndrome coronavirus 2 isolate Wuhan-Hu-1            | SARS-CoV-2      | NC_045512             |
| Betacoronavirus Erinaceus/VMC/DEU/2012 isolate ErinaceusCoV/2012-216/GER/2012 | CoV2012-216 GER | KC545386              |
| Rousettus bat coronavirus isolate GCCDC1 356                                  | GCCDC1          | KU762338              |
| BtRs-BetaCoV/GX2013   | GX2013          | KJ473815              |
| Human coronavirus HKU1 strain SC2521  | HKU1            | MK167038              |
| Bat coronavirus HKU4  | HKU4            | EF065505              |
| Bat coronavirus HKU5  | HKU5            | EF065509              |
| Bat coronavirus HKU9  | HKU9            | EF065513              |
| Betacoronavirus HKU24 strain HKU24-R050101                                    | HKU24           | KM349744              |
| Bat Hp-betacoronavirus/Zhejiang2013   | Hp-ZJ2013       | KF636752              |
| Middle East respiratory syndrome coronavirus                                  | MERS-CoV        | NC_019843             |
| Murine coronavirus MHV-1  | MHV-1           | FJ647223              |
| Mouse hepatitis virus strain MHV-A59 C12 mutant                               | MHV-A59         | NC_001846             |
| Rat coronavirus Parker  | RCoV            | NC_012936             |
| Human coronavirus OC43  | HCoV-OC43       | AY391777              |
| SARS coronavirus GD01   | SARS-CoV        | AY278489              |
| Bat SARS-like coronavirus isolate bat-SL-CoVZC45                              | SARS-like ZC45  | MG772933              |
| Coronavirus BtRI-BetaCoV/SC2018   | SC2018          | MK211374              |
| BtRs-BetaCoV/YN2013   | YN2013          | KJ473816              |
| Human coronavirus 229E  | HCoV-229E       | NC_002645             |
| Lucheng Rn rat coronavirus isolate Lucheng-19                                 | LRNV            | KF294380              |
| Porcine epidemic diarrhea virus   | PEDV            | NC_003436             |

PCR was performed using primers (Table 6) to amplify the target gene CytB (CO1). The expected amplicon using CytB\_F and CytB\_R primers was a 457-bp product. The expected amplicon lengths of two rounds of nested PCR were 663 bp (using primers CO1\_long\_F and CO1\_long\_R) and 324 bp (using primers CO1\_short\_F and CO1\_short\_R). Purified DNA products were sequenced bidirectionally with primers we provided on an ABI Prism 3730XL DNA analyzer (Applied Biosystems, USA). To identify species, sequences were searched against the Basic Local Alignment Search Tool (BLAST) database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>).

**RNA extraction.** Viral RNA was extracted from 140  $\mu$ l of VTM suspension of each oral and anal swab using a QIAamp viral RNA minikit (Qiagen, Germany) according to the manufacturer's protocol. RNA was eluted with 60  $\mu$ l of AVE buffer, 20  $\mu$ l of which was used as the reverse transcription-PCR (RT-PCR) template; the remainder was stored at  $-80^{\circ}\text{C}$ .

**RT-PCR screening of CoVs.** Viral RNA was extracted from oral and anal swabs to screen for CoVs using RT-PCR with universal degenerate primers, as previously reported (32). The primers used (Table 6) targeted two nonoverlapping regions of the sequence encoding the RNA-dependent RNA polymerase (RdRp) in ORF1b. Both regions were used for phylogenetic discrimination. After RNA reverse transcription (Toyobo, Japan), the cDNA was used for seminested or nested PCR. The expected amplicon lengths of two rounds were 520 bp (using primers CoV-FWD1 and CoV-RVS1), 328 bp (using primers CoV-FWD2 and CoV-RVS2), 440 bp (using primers CoV-FWD3 and CoV-RVS3), and 434 bp (using primers CoV-FWD4 and CoV-RVS3). Purified DNA products were sequenced bidirectionally with the primers we provided (Table 6) on an ABI Prism 3730XL DNA analyzer.

**Genomic sequencing.** A volume of 560  $\mu$ l of CoV-positive sample was used to extract RNA for deep sequencing. RNA was subjected to next-generation sequencing using the HiSeq 150-bp paired-end (PE150) platform (Illumina). The original sequencing reads were assembled *de novo* using MEGAHIT to produce three nearly full-length genomic sequences, named RCoV-GCCDC3, RCoV-GCCDC4, and RCoV-GCCDC5, respectively.

**Rapid amplification of cDNA ends.** As described previously (33), according to the genome sequences of RCoV-GCCDC3, RCoV-GCCDC4, and RCoV-GCCDC5, we designed gene-specific primers for further rapid amplification of cDNA ends (RACE) experiments. A volume of 200  $\mu$ l of corresponding sample was used to extract total RNA with TRIzol. First-strand cDNA was synthesized to generate RACE-ready cDNA. Subsequently, rapid amplification of cDNA ends was performed with the 5'- and 3'-Full RACE kit (TaKaRa, Japan). Purified DNA products were sequenced bidirectionally with the primers on an ABI Prism 3730XL DNA analyzer.

**Sequence analysis.** The genome sequence of RCoV-GCCDC4 was compared with the sequences of other known representative CoVs in the BLAST database (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>). The ORFs, cleavage sites, and amino acid sequences of RCoV-GCCDC4 were predicted by ZCURVE\_CoV 2.0 (36). Sequence alignment and editing were performed using DNAMAN, EditSeq, and MegAlign. The leader and body transcription regulatory sequences (TRS) were acquired by comparison with the sequence of HKU24 (13). Phylogenetic analysis of polyproteins 1a and 1b and the S, E, M, and N proteins was conducted using the maximum-likelihood method in MEGA 7 (54), with bootstrap values calculated from 1,000 replicate trees.

**Cell culture and virus isolation.** African green monkey kidney (Vero-E6) cells and human epithelial colorectal adenocarcinoma (Caco-2) cells (33) were cultured in Dulbecco's modified Eagle medium with 10% and 20% fetal bovine serum, respectively, in 5% CO<sub>2</sub> at 37°C. Virus isolation with the 35 CoV-positive samples was performed under trypsin concentration gradients of 0, 1, and 2 μg/ml for Vero-E6 and 0, 1, 2, 3, 4, and 5 μg/ml for Caco-2 cell lines. Subsequently, samples were cultured in Vero-E6 cells with 2 μg/ml and in Caco-2 cells with 5 μg/ml trypsin. After adding samples and incubating for 1 h, maintenance solution containing the corresponding trypsin concentration was added. Each sample was cultivated in three blind passages with or without trypsin treatment. The cultured supernatant of each cell line passage was harvested for RT-PCR.

**Recombination analysis.** Genomic sequences of RCoV-GCCDC4, MHV-1, and Longquan RI rat coronavirus (LRLV) were aligned using SimPlot (55) and RDP4 (56) to detect possible recombination events. The parameters were set with a window size of 400 bp and a step size of 200 bp.

**Data availability.** The complete genome sequences of RCoV-GCCDC3, RCoV-GCCDC4, and RCoV-GCCDC5 have been deposited in GenBank and assigned accession numbers [MW802581](https://doi.org/10.1093/nar/gkz025), [MW773844](https://doi.org/10.1093/nar/gkz026), and [MW802582](https://doi.org/10.1093/nar/gkz027), respectively.

## ACKNOWLEDGMENTS

We thank Tao Jin for helpful suggestions and discussions during the preparation of the manuscript.

This work was supported by grants from the Major Special Projects for Infectious Disease Research of China (grant 2018ZX10101002-005) and the Research Units of Adaptive Evolution and Control of Emerging Viruses (grant 2018RU009), Chinese Academy of Medical Sciences. Field animal sampling was supported by the United States Agency for International Development (USAID) Emerging Pandemic Threats PREDICT project (cooperative agreement no. AID-OAA-A-14-00102). P.L. was supported by the Young Scientists Research Fund of Academician Yunde Hou. W.J.L. is supported by the Excellent Young Scientist Program of the Natural Science Foundation of China (81822040) and by the National Youth Talent Support Program. G.F.G. is a leading principal investigator of the National Natural Science Foundation of China Innovative Research Group (81621091).

We declare that we have no conflicts of interest.

## REFERENCES

- Hamre D, Procknow JJ. 1962. A new virus isolated from the human respiratory tract. *Proc Soc Exp Biol Med* 121:190–193.
- Hendley JO, Fishburne HB, Gwaltney JM, Jr. 1972. Coronavirus infections in working adults. Eight-year study with 229 E and OC 43. *Am Rev Respir Dis* 105:805–811.
- van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJ, Wolthers KC, Wertheim-van Dillen PM, Kaandorp J, Spaargaren J, Berkhout B. 2004. Identification of a new human coronavirus. *Nat Med* 10:368–373. <https://doi.org/10.1038/nm1024>.
- Woo PC, Lau SK, Chu CM, Chan KH, Tsoi HW, Huang Y, Wong BH, Poon RW, Cai JJ, Luk WK, Poon LL, Wong SS, Guan Y, Peiris JS, Yuen KY. 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol* 79:884–895. <https://doi.org/10.1128/JVI.79.2.884-895.2005>.
- Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W, Rollin PE, Dowell SF, Ling AE, Humphrey CD, Shieh WJ, Guarner J, Paddock CD, Rota P, Fields B, DeRisi J, Yang JY, Cox N, Hughes JM, LeDuc JW, Bellini WJ, Anderson LJ, SARS Working Group. 2003. A novel coronavirus associated with severe acute respiratory syndrome. *N Engl J Med* 348:1953–1966. <https://doi.org/10.1056/NEJMoa030781>.
- Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 367:1814–1820. <https://doi.org/10.1056/NEJMoa1211721>.
- Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W, China Novel Coronavirus Investigating and Research Team. 2020. A novel coronavirus from patients with pneumonia in China, 2019. *N Engl J Med* 382:727–733. <https://doi.org/10.1056/NEJMoa2001017>.
- Woo PC, Lau SK, Wernery U, Wong EY, Tsang AK, Johnson B, Yip CC, Lau CC, Sivakumar S, Cai JP, Fan RY, Chan KH, Mareena R, Yuen KY. 2014. Novel betacoronavirus in dromedaries of the Middle East, 2013. *Emerg Infect Dis* 20:560–572. <https://doi.org/10.3201/eid2004.131769>.
- Zhang J, Guy JS, Snijder EJ, Denniston DA, Timoney PJ, Balasuriya UB. 2007. Genomic characterization of equine coronavirus. *Virology* 369:92–104. <https://doi.org/10.1016/j.virol.2007.06.035>.
- Lim SI, Choi S, Lim JA, Jeoung HY, Song JY, Dela Pena RC, An DJ. 2013. Complete genome analysis of canine respiratory coronavirus. *Genome Announc* 1:e00093-12. <https://doi.org/10.1128/genomeA.00093-12>.
- Xiao K, Zhai J, Feng Y, Zhou N, Zhang X, Zou JJ, Li N, Guo Y, Li X, Shen X, Zhang Z, Shu F, Huang W, Li Y, Zhang Z, Chen RA, Wu YJ, Peng SM, Huang M, Xie WJ, Cai QH, Hou FH, Chen W, Xiao L, Shen Y. 2020. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. *Nature* 583:286–289. <https://doi.org/10.1038/s41586-020-2313-x>.
- Obameso JO, Li H, Jia H, Han M, Zhu S, Huang C, Zhao Y, Zhao M, Bai Y, Yuan F, Zhao H, Peng X, Xu W, Tan W, Zhao Y, Yuen KY, Liu WJ, Lu L, Gao GF. 2017. The persistent prevalence and evolution of cross-family recombinant coronavirus GCCDC1 among a bat population: a two-year follow-up. *Sci China Life Sci* 60:1357–1363. <https://doi.org/10.1007/s11427-017-9263-6>.
- Lau SK, Woo PC, Li KS, Tsang AK, Fan RY, Luk HK, Cai JP, Chan KH, Zheng BJ, Wang M, Yuen KY. 2015. Discovery of a novel coronavirus, China Rattus coronavirus HKU24, from Norway rats supports the murine origin of *Beta-coronavirus* 1 and has implications for the ancestor of *Betacoronavirus* lineage A. *J Virol* 89:3076–3092. <https://doi.org/10.1128/JVI.02420-14>.
- Wilson DE, Reeder DM. 2005. *Mammal species of the world: a taxonomic and geographic reference*, 3rd ed. Johns Hopkins University Press, Baltimore, MD.
- Olival KJ, Hosseini PR, Zambrana-Torrel C, Ross N, Bogich TL, Daszak P. 2017. Host and viral traits predict zoonotic spillover from mammals. *Nature* 546:646–650. <https://doi.org/10.1038/nature22975>.
- Cheever FS, Daniels JB. 1949. A murine virus (JHM) causing disseminated encephalomyelitis with extensive destruction of myelin. *J Exp Med* 90:181–210. <https://doi.org/10.1084/jem.90.3.181>.
- Matthews AE, Weiss SR, Paterson YJ. 2002. Murine hepatitis virus—a model for virus-induced CNS demyelination. *J Neurovirol* 8:76–85. <https://doi.org/10.1080/13550280290049534>.

18. Wu Z, Lu L, Du J, Yang L, Ren X, Liu B, Jiang J, Yang J, Dong J, Sun L, Zhu Y, Li Y, Zheng D, Zhang C, Su H, Zheng Y, Zhou H, Zhu G, Li H, Chmura A, Yang F, Daszak P, Wang J, Liu Q, Jin Q. 2018. Comparative analysis of rodent and small mammal viromes to better understand the wildlife origin of emerging infectious diseases. *Microbiome* 6:178. <https://doi.org/10.1186/s40168-018-0554-9>.
19. Latifin A, Hu B, Olival KJ, Zhu G, Zhang L, Li H, Chmura AA, Field HE, Zambrana-Torrel C, Epstein JH, Li B, Zhang W, Wang LF, Shi ZL, Daszak P. 2020. Origin and cross-species transmission of bat coronaviruses in China. *Nat Commun* 11:4235. <https://doi.org/10.1038/s41467-020-17687-3>.
20. Memish ZA, Mishra N, Olival KJ, Fagbo SF, Kapoor V, Epstein JH, Alhakeem R, Durosinloun A, Al Asmari M, Islam A, Kapoor A, Briese T, Daszak P, Al Rabeeah AA, Lipkin WI. 2013. Middle East respiratory syndrome coronavirus in bats, Saudi Arabia. *Emerg Infect Dis* 19:1819–1823.
21. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Si HR, Zhu Y, Li B, Huang CL, Chen HD, Chen J, Luo Y, Guo H, Jiang RD, Liu MQ, Chen Y, Shen XR, Wang X, Zheng XS, Zhao K, Chen QJ, Deng F, Liu LL, Yan B, Zhan FX, Wang YY, Xiao GF, Shi ZL. 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579:270–273. <https://doi.org/10.1038/s41586-020-2012-7>.
22. Liu P, Yang M, Zhao X, Guo Y, Wang L, Zhang J, Lei W, Han W, Jiang F, Liu WJ, Gao GF, Wu G. 2020. Cold-chain transportation in the frozen food industry may have caused a recurrence of COVID-19 cases in destination: successful isolation of SARS-CoV-2 virus from the imported frozen food package surface. *Biosaf Health* 2:199–201. <https://doi.org/10.1016/j.bshealth.2020.11.003>.
23. Corman VM, Muth D, Niemeyer D, Drosten C. 2018. Hosts and sources of endemic human coronaviruses. *Adv Virus Res* 100:163–188. <https://doi.org/10.1016/bs.aivir.2018.01.001>.
24. Masters PS, Perlman S. 2013. *Coronaviridae*, p 825–879. In Knipe DM, Howley PM (ed), *Fields virology*, 6th ed. Lippincott Williams & Wilkins, Philadelphia, PA.
25. King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ. 2012. Virus taxonomy: classification and nomenclature of viruses: ninth report of the International Committee on Taxonomy of Viruses. Academic Press, London, United Kingdom.
26. Sawicki SG, Sawicki DL, Siddell SG. 2007. A contemporary view of coronavirus transcription. *J Virol* 81:20–29. <https://doi.org/10.1128/JVI.01358-06>.
27. Pasternak AO, Spaan WJM, Snijder EJ. 2006. Nidovirus transcription: how to make sense...? *J Gen Virol* 87:1403–1421. <https://doi.org/10.1099/vir.0.81611-0>.
28. Brian DA, Baric RS. 2005. Coronavirus genome structure and replication. *Curr Top Microbiol Immunol* 287:1–30. [https://doi.org/10.1007/3-540-26765-4\\_1](https://doi.org/10.1007/3-540-26765-4_1).
29. Peng G, Sun D, Rajashankar KR, Qian Z, Holmes KV, Li F. 2011. Crystal structure of mouse coronavirus receptor-binding domain complexed with its murine receptor. *Proc Natl Acad Sci U S A* 108:10696–10701. <https://doi.org/10.1073/pnas.1104306108>.
30. Zhang Z, Zheng Y, Niu Z, Zhang B, Wang C, Yao X, Peng H, Franca DN, Wang Y, Zhu Y, Su Y, Tang M, Jiang X, Ren H, He M, Wang Y, Gao L, Zhao P, Shi H, Chen Z, Wang X, Piacentini M, Bian X, Melino G, Liu L, Huang H, Sun Q. 2021. SARS-CoV-2 spike protein dictates syncytium-mediated lymphocyte elimination. *Cell Death Differ* 132:47–68.
31. Qiao B, Olvera de la Cruz M. 2020. Enhanced binding of SARS-CoV-2 spike protein to receptor by distal polybasic cleavage sites. *ACS Nano* 14:10616–10623. <https://doi.org/10.1021/acsnano.0c04798>.
32. Watanabe S, Masangkay JS, Nagata N, Morikawa S, Mizutani T, Fukushi S, Alviola P, Omatsu T, Ueda N, Iha K, Taniguchi S, Fujii H, Tsuda S, Endoh M, Kato K, Tohya Y, Kyuwa S, Yoshikawa Y, Akashi H. 2010. Bat coronaviruses and experimental infection of bats, the Philippines. *Emerg Infect Dis* 16:1217–1223. <https://doi.org/10.3201/eid1608.100208>.
33. Huang C, Liu WJ, Xu W, Jin T, Zhao Y, Song J, Shi Y, Ji W, Jia H, Zhou Y, Wen H, Zhao H, Liu H, Li H, Wang Q, Wu Y, Wang L, Liu D, Liu G, Yu H, Holmes EC, Lu L, Gao GF. 2016. A bat-derived putative cross-family recombinant coronavirus with a reovirus gene. *PLoS Pathog* 12:e1005883. <https://doi.org/10.1371/journal.ppat.1005883>.
34. Corman VM, Ithete NL, Richards LR, Schoeman MC, Preiser W, Drosten C, Drexler JF. 2014. Rooting the phylogenetic tree of Middle East respiratory syndrome coronavirus by characterization of a conspecific virus from an African bat. *J Virol* 88:11297–11303. <https://doi.org/10.1128/JVI.01498-14>.
35. Henrich S, Cameron A, Bourenkov GP, Kiefersauer R, Huber R, Lindberg I, Bode W, Than ME. 2003. The crystal structure of the proprotein processing proteinase furin explains its stringent specificity. *Nat Struct Biol* 10:520–526. <https://doi.org/10.1038/nsb941>.
36. Gao F, Ou HY, Chen LL, Zheng WX, Zhang CT. 2003. Prediction of proteinase cleavage sites in polyproteins of coronaviruses and its applications in analyzing SARS-CoV genomes. *FEBS Lett* 553:451–456. [https://doi.org/10.1016/s0014-5793\(03\)01091-3](https://doi.org/10.1016/s0014-5793(03)01091-3).
37. Fang S, Shen H, Wang J, Tay FP, Liu DX. 2010. Functional and genetic studies of the substrate specificity of coronavirus infectious bronchitis virus 3C-like proteinase. *J Virol* 84:7325–7336. <https://doi.org/10.1128/JVI.02490-09>.
38. Wang W, Lin XD, Guo WP, Zhou RH, Wang MR, Wang CQ, Ge S, Mei SH, Li MH, Shi M, Holmes EC, Zhang YZ. 2015. Discovery, diversity and evolution of novel coronaviruses sampled from rodents in China. *Virology* 474:19–27. <https://doi.org/10.1016/j.virol.2014.10.017>.
39. Zhou P, Fan H, Lan T, Yang XL, Shi WF, Zhang W, Zhu Y, Zhang YW, Xie QM, Mani S, Zheng XS, Li B, Li JM, Guo H, Pei GQ, An XP, Chen JW, Zhou L, Mai KJ, Wu ZX, Li D, Anderson DE, Zhang LB, Li SY, Mi ZQ, He TT, Cong F, Guo PJ, Huang R, Luo Y, Liu XL, Chen J, Huang Y, Sun Q, Zhang XL, Wang YY, Xing SZ, Chen YS, Sun Y, Li J, Daszak P, Wang LF, Shi ZL, Tong YG, Ma JY. 2018. Fatal swine acute diarrhoea syndrome caused by an HKU2-related coronavirus of bat origin. *Nature* 556:255–258. <https://doi.org/10.1038/s41586-018-0010-9>.
40. Oude Munnink BB, Sikkema RS, Nieuwenhuijse DF, Molenaar RJ, Munger E, Molenkamp R, van der Spek A, Tolma S, Rietveld A, Brouwer M, Bouwmeester-Vincken N, Harders F, Hakze-van der Honing R, Wegdam-Blans MCA, Bouwstra RJ, GeurtsvanKessel C, van der Eijk AA, Velkers FC, Smit LAM, Stegeman A, van der Poel WHM, Koopmans MPG. 2021. Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science* 371:172–177. <https://doi.org/10.1126/science.abe5901>.
41. Vlasova AN, Diaz A, Damtie D, Xiu L, Toh TH, Lee JS, Saif LJ, Gray GC. 2021. Novel canine coronavirus isolated from a hospitalized pneumonia patient, East Malaysia. *Clin Infect Dis* <https://doi.org/10.1093/cid/ciab456>.
42. de Groot RJ, Baker SC, Baric R, Enjuanes L, Gorbalenya AE, Holmes KV, Perlman S, Poon L, Rottier PJM, Talbot PJ, Woo PCY, Ziebuhr J. 2011. Family *Coronaviridae*, p 806–828. In King AMQ, Lefkowitz E, Adams MJ, Carstens EB (ed), *Virus taxonomy: 9th report of the international committee on taxonomy of viruses*. Elsevier, San Diego, CA.
43. Menachery VD, Yount BL, Jr, Debbink K, Agnihothram S, Gralinski LE, Plante JA, Graham RL, Scobey T, Ge XY, Donaldson EF, Randell SH, Lanzavecchia A, Marasco WA, Shi ZL, Baric RS. 2015. A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat Med* 21:1508–1513. <https://doi.org/10.1038/nm.3985>.
44. Follis KE, York J, Nunberg JH. 2006. Furin cleavage of the SARS coronavirus spike glycoprotein enhances cell-cell fusion but does not affect virion entry. *Virology* 350:358–369. <https://doi.org/10.1016/j.virol.2006.02.003>.
45. Menachery VD, Dinnon KH, 3rd, Yount BL, Jr, McAnarney ET, Gralinski LE, Hale A, Graham RL, Scobey T, Anthony SJ, Wang L, Graham B, Randell SH, Lipkin WI, Baric RS. 2020. Trypsin treatment unlocks barrier for zoonotic bat coronavirus infection. *J Virol* 94:e01774–19. <https://doi.org/10.1128/JVI.01774-19>.
46. Alexander DJ, Brown IH. 2009. History of highly pathogenic avian influenza. *Rev Sci Tech* 28:19–38. <https://doi.org/10.20506/rst.28.1.1856>.
47. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. 2020. The proximal origin of SARS-CoV-2. *Nat Med* 26:450–452. <https://doi.org/10.1038/s41591-020-0820-9>.
48. Brockway SM, Denison MR. 2005. Mutagenesis of the murine hepatitis virus nsp1-coding region identifies residues important for protein processing, viral RNA synthesis, and viral replication. *Virology* 340:209–223. <https://doi.org/10.1016/j.virol.2005.06.035>.
49. Zhang R, Li Y, Cowley TJ, Steinbrenner AD, Phillips JM, Yount BL, Baric RS, Weiss SR. 2015. The nsp1, nsp13, and M proteins contribute to the hepatotropism of murine coronavirus JHM.WU. *J Virol* 89:3598–3609. <https://doi.org/10.1128/JVI.03535-14>.
50. Lei L, Ying S, Baojun L, Yi Y, Xiang H, Wenli S, Zouan S, Deyin G, Qingyu Z, Jingmei L, Guohui C. 2013. Attenuation of mouse hepatitis virus by deletion of the LLRKxGxKG region of Nsp1. *PLoS One* 8:e61166. <https://doi.org/10.1371/journal.pone.0061166>.
51. Züst R, Cervantes-Barragán L, Kuri T, Blakqori G, Weber F, Ludewig B, Thiel V. 2007. Coronavirus non-structural protein 1 is a major pathogenicity factor: implications for the rational design of coronavirus vaccines. *PLoS Pathog* 3:e109. <https://doi.org/10.1371/journal.ppat.0030109>.
52. Wang P, Liu WJ. 2020. It's not just science: challenges for public health intervention in Ebola epidemics in the Democratic Republic of Congo.



- Sci China Life Sci 63:1079–1081. <https://doi.org/10.1007/s11427-019-1670-6>.
53. Liu WJ, Wu G. 2020. Convincing the confidence to conquer COVID-19: from epidemiological intervention to laboratory investigation. *Biosaf Health* 2:185–186. <https://doi.org/10.1016/j.bshealth.2020.11.005>.
54. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874. <https://doi.org/10.1093/molbev/msw054>.
55. Lole KS, Bollinger RC, Paranjape RS, Gadkari D, Kulkarni SS, Novak NG, Ingersoll R, Sheppard HW, Ray SC. 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 73:152–160. <https://doi.org/10.1128/JVI.73.1.152-160.1999>.
56. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. 2015. RDP4: detection and analysis of recombination patterns in virus genomes. *Virus Evol* 1:vev003. <https://doi.org/10.1093/ve/vev003>.