



LETTER

Genomic Characterization of a New Coronavirus from Migratory Birds in Jiangxi Province of China

Wentao Zhu¹ · Wentao Song² · Guoyin Fan² · Jing Yang^{1,3,4} · Shan Lu^{1,3,4} · Dong Jin^{1,3,4} · Xue-lian Luo¹ · Ji Pu¹ · Haiying Chen² · Jianguo Xu^{1,3,4}

Received: 25 January 2021 / Accepted: 29 March 2021 / Published online: 8 July 2021
© Wuhan Institute of Virology, CAS 2021

Dear Editor,

The emergence of COVID-19 since December 2019 has attracted great attention around the world and reminds the powerful pathogenic potential of viruses (Zhou *et al.* 2020). The subfamily *Orthocoronavirinae* of the family *Coronaviridae* contains four genera *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus*, which are the largest group of positive-sense, non-segmented, single-stranded, enveloped RNA viruses (Woo *et al.* 2012; Shi *et al.* 2016). The genus *Gammacoronavirus* currently has five species (<https://talk.ictvonline.org/taxonomy/>), which are primarily spread through birds (Woo *et al.* 2012). Infectious bronchitis virus (IBV) is the best-studied *Gammacoronavirus*, which can lead to economic losses to the global poultry industry by causing avian infectious bronchitis (a highly contagious viral respiratory

disease) (Jackwood and de Wit 2020). Turkey coronavirus (TCoV), also a member of *Gammacoronavirus*, causes acute infectious diarrhea in domestic turkeys (Lin *et al.* 2002). Recently, a novel species of genus *Gammacoronavirus* was identified in a mass die-off of Canada and snow geese (Papineau *et al.* 2019).

Birds, with more than 10,000 living species, are one of the most popular beings with global distribution (Prum *et al.* 2015), and are the reservoir of many emerging and re-emerging viruses, such as avian influenza viruses and the West Nile virus (Reed *et al.* 2003; Olsen *et al.* 2006). With the capacity to fly long distances, birds play an important role in disseminating these emerging viruses to animals and/or humans (Woo *et al.* 2012). In the previous study, we have identified a new deltacoronavirus from birds of Qinghai-Tibetan Plateau in China (Zhu *et al.* 2020).

In February–March 2019, migratory birds were live-captured and kept in cages by professionals of migratory bird protection station under the permission from Jiangxi Province Department of Forestry. Feces of migratory birds were collected using non-invasive ways respectively, and then released. A total of 48 fecal samples were collected from two regions (Nanchang City and Jiujiang City) around Poyang Lake in Jiangxi Province of China. Total RNA was extracted from these samples using the QIAamp Viral RNA Mini Kit. Then, 15 out of 48 total RNA were randomly selected and pooled in equal amounts. Then, total RNA was used to construct the RNA sequencing library using Truseq stranded total RNA library prep gold kit and sequenced by the Illumina HiSeq2000 platform. After quality control, a total of 75,448,386 clean reads were obtained. The *de novo* assembled contigs were obtained using Trinity v2.10.0 (Grabherr *et al.* 2011) and annotated against the NCBI nr database with an E-value cutoff of 1×10^{-5} .

We identified a 28,466 nucleotide (nt) long contig annotated as coronavirus (CoV). The contig was verified, and a total of 1,346 reads were mapped to the CoV

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12250-021-00402-x>.

✉ Jianguo Xu
xujianguo@icdc.cn

✉ Haiying Chen
nccdcchy@126.com

¹ State Key Laboratory of Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing 102206, China

² The Collaboration Unit for Field Epidemiology of State Key Laboratory of Infectious Disease Prevention and Control, Jiangxi Provincial Key Laboratory of Animal-Origin and Vector-Borne Diseases, Nanchang Center for Disease Control and Prevention, Nanchang 330038, China

³ Shanghai Public Health Clinical Center, Fudan University, Shanghai 201508, China

⁴ Research Units of Discovery of Unknown Bacteria and Function, Chinese Academy of Medical Sciences, Beijing 100730, China

genome. The full-length genome was named as *Anser fabalis* coronavirus (AFCoV) NCN2 based on one of its host, and was submitted to the NCBI database (GenBank accession number MW436465). BLASTn result shows that AFCoV shares 86.0% nt identity to a recently identified *Gammacoronavirus* [Canada goose coronavirus (CGCoV) CB17] and even lower nucleotide identity (< 78.8%) with other coronaviruses. The seven concatenated conserved domains (ADRP, 3CL^{pro}, RdRp, Hel, ExoN, NendoU and O-MT) between AFCoV and CGCoV show greater than 90% amino acid (aa) identities (Table 1). These analyses identified AFCoV as a new member of the species of *Goose coronavirus CB17*.

The genome of AFCoV has a 38.5% G + C content and typical CoV genomic organization consisting of 5' UTR (untranslated region), replicase ORF1ab, spike (S), envelope (E), membrane (M), nucleocapsid (N), several accessory proteins, and 3' UTR (Fig. 1A). The 5' UTR of AFCoV is 519 nt in length and shares the highest nucleotide identity (94.9%) with that of CGCoV. Like other coronaviruses, the *ORF1ab* gene occupies a large portion of the AFCoV genome and has a frameshift site at position 11,940 based on the conserved heptamer 'UUUAAAC'. The polyprotein ORF1ab of AFCoV is cleaved into 16 non-structural proteins (NSPs) by two viral proteases, with almost the same putative protease cleavage sites as those of CGCoV (Supplementary Table S1). The AFCoV genome contains nine predicted accessory proteins, one less than those in CGCoV (Papineau *et al.* 2019) (Supplementary

Table S2). The 3' UTR (excluding the poly(A) tail) of AFCoV (255 nt in length) shows 98.8% nt identity with that of CGCoV.

The protein ORF1ab of AFCoV (6500 aa) shares 90.0% aa identity with that of CGCoV. The deduced S protein (1186 aa) of AFCoV shares a low identity (68.5%) with the S protein of CGCoV (1184 aa), and even lower identities with that of other coronaviruses (< 58.1%). The deduced E, M, and N proteins of AFCoV have 100%, 97.8%, and 94.9% aa identities with those of CGCoV respectively, and relatively low identities with the proteins of other Gammacoronaviruses ranging from 29.7 to 62.4% for E protein, 30.5–65.8% for M protein, and 36.4%–60.8% for N protein.

To assess the phylogenetic relationship of AFCoV with other coronaviruses, the full-length genome sequences of alphacoronaviruses, betacoronaviruses, gammacoronaviruses, and deltacoronaviruses were aligned with MAFFT and used to reconstruct the phylogenetic tree by the maximum-likelihood method with JC best-fit model. The phylogenetic tree (Fig. 1B) reveals that AFCoV clusters in the genus *Gammacoronaviruses* and groups closely with CGCoV. Phylogenetic trees based on proteins ORF1ab (LG model), and S (Dayhoff model) were also constructed, and show similar topologies as that of the genomes, indicating that AFCoV is closely related to CGCoV (Supplementary Fig. S1–S2).

To determine the prevalence of the new Gammacoronavirus in bird samples, we designed primers (F: 5'-CTTTTGGTCTCTACCCTGTTC-3', R: 5'-CCA-TAAAAATCCAGGACTTGTT-3') to specifically amplify an 836 bp segment located in the replicase *ORF1ab* of AFCoV genome. All fecal samples were tested using the One-Step RT-PCR Kit Ver.2 (Takara, Japan). Five out of 48 (10.4%) samples were found to be positive for coronavirus and were also verified by Sanger sequencing. Result of alignment showed that four of the five sequences were absolutely the same, and the remaining one sequence has seven nucleotide differences with the consensus sequence. The five hosts of those coronavirus positive samples belong to four species, all from family *Anatidae*, including *Cygnus columbianus* (one host) and *Anser albifrons* (one host) from Nanchang City, *Anas zonorhyncha* (two hosts), and *Anser fabalis* (one host) from Jiujiang City.

Identifying viruses that have ability to spillover from wildlife into other animals and/or humans promotes our understanding of viral families that are very likely to be new pathogens in the future (Anthony *et al.* 2015). Here, we detected a new Gammacoronavirus (AFCoV) in four different bird species from two sites with a high (10.4%) detection rate. The ORF1ab, E, M, and N proteins of AFCoV had > 90% aa identities to those of CGCoV which caused a die-off of Canada and snow geese in a site near

Table 1 Comparison of amino acid identities between AFCoV and closely related CoV genome sequences.

AFCoV proteins	Amino acid identity (%)		TCoV	SW1
	CGCoV	IBV		
ADRP	88.9	51.6	50.4	37.5
3CL ^{pro}	86.9	57.7	59.0	49.5
RdRp	97.4	77.1	76.9	65.7
Hel	100	88.5	88.3	75.0
ExoN	99.8	78.2	76.6	57.4
NendoU	99.7	59.0	58.7	41.8
O-MT	99.3	73.5	70.9	62.8
S	68.5	52.0	31.8	26.5
E	100	62.4	68.0	29.7
M	97.8	65.8	64.9	30.5
N	94.9	60.8	60.6	36.4

CGCoV: Canada goose coronavirus (NC_046965).

IBV: Avian infectious bronchitis virus (NC_001451).

TCoV: Turkey coronavirus (NC_010800).

SW1: Beluga Whale coronavirus SW1 (NC_010646).

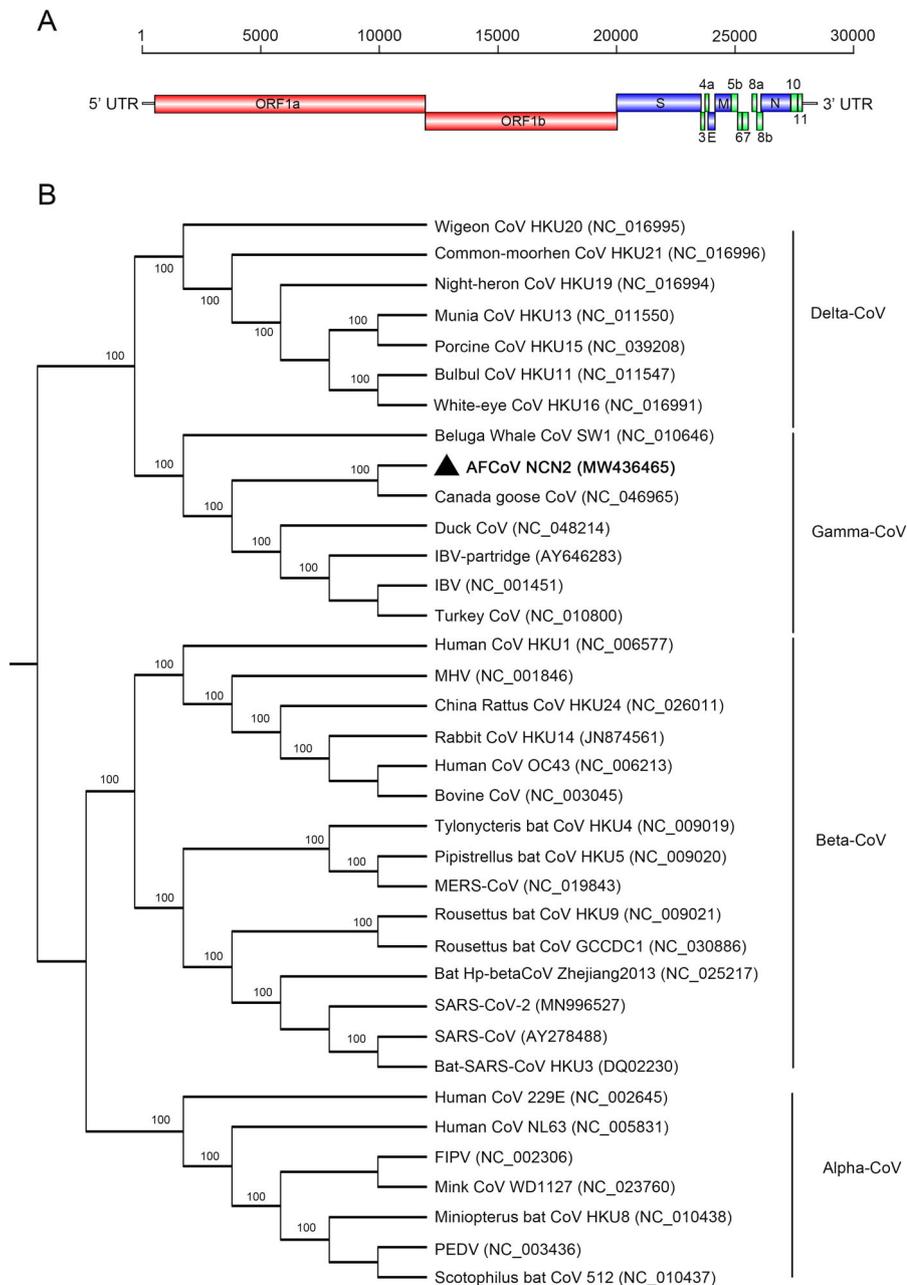


Fig. 1 **A** Genome organization of AFCoV. **B** The phylogenetic tree was constructed using genome sequences of different coronaviruses. The maximum-likelihood tree was constructed using PhyML 3.0, and

only bootstrap values > 70% are shown. The viral sequence obtained in this study is marked in bold type.

the arctic, indicating that AFCoV may also be pathogenic to animals. Birds of the family *Anatidae* may be the natural host of *Goose coronavirus CBI7* species (including AFCoV and CGCoV), because all the reservoirs of AFCoV and CGCoV belong to the family *Anatidae* (Papineau *et al.* 2019). The S protein of AFCoV has a significantly low aa identity (68.5%) with that of CGCoV, which may explain the broader diversity of AFCoV reservoirs (*Cygnus columbianus*, *Anser albifrons*, *Anas zonorhyncha*, and

Anser fabalis) than CGCoV reservoirs (*Branta canadensis* and *Anser caerulescens*) (Papineau *et al.* 2019). The sampling dates and sites of five coronavirus positive birds *Cygnus columbianus* (February 27, 2019; Nangchang City), *Anser albifrons* (March 5, 2019; Nangchang city), *Anas zonorhyncha* (March 10, 2019; Jiujiang city), and *Anser fabalis* (On March 10, 2019; Jiujiang city) were different, indicating the spread of AFCoV among birds of family *Anatidae* because different migratory birds share the

same habitat. Bootscan analysis using genomes of avian coronavirus, IBV, CGCoV, and AFCoV (as the query sequence) indicated the absence of any obvious recombination event (Supplementary Fig. S3). To summarize, this study describes a new member of *Goose coronavirus CB17* species in migratory birds from China and promotes our understanding of the diversity of gammacoronaviruses. We did not have the permission to collect tissue samples, and the isolation of virus was unsuccessful. Therefore, the specific pathogenicity of AFCoV still needs further study.

Acknowledgements This work was supported by grants from National Science and Technology Major Project of China (2018ZX10712001-018), National Key R&D Program of China (2019YFC1200500 and 2019YFC1200505), National Natural Science Foundation of China (81802017) and Research Units of Discovery of Unknown Bacteria and Function (2018RU010).

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Animal and Human Rights Statement The fecal samples of migratory birds were collected according the Chinese law and regulations and authorized by Forestry Bureau of Jiangxi Province of China. The study practices were approved by Nanchang Center for Diseases Prevention and Control.

References

- Anthony SJ, St Leger JA, Liang E, Hicks AL, Sanchez-Leon MD, Jain K, Lefkowitz JH, Navarrete-Macias I, Knowles N, Goldstein T, Puglianes K, Ip HS, Rowles T, Lipkin WI (2015) Discovery of a novel hepatovirus (phopivirus of seals) related to human hepatitis a virus. *mBio* 6:e01180-15
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652
- Jackwood MW, de Wit S (2020) Infectious bronchitis. In: Diseases of poultry, 3th edn. Wiley. <https://doi.org/10.1002/9781119371199.ch4>
- Lin TL, Loa CC, Tsai SC, Wu CC, Bryan TA, Thacker HL, Hooper T, Schrader D (2002) Characterization of turkey coronavirus from turkey poults with acute enteritis. *Vet Microbiol* 84:179–186
- Olsen B, Munster VJ, Wallensten A, Waldenström J, Osterhaus AD, Fouchier RA (2006) Global patterns of influenza a virus in wild birds. *Science* 312:384–388
- Papineau A, Berhane Y, Wylie TN, Wylie KM, Sharpe S, Lung O (2019) Genome organization of Canada goose coronavirus, a novel species identified in a mass die-off of canada geese. *Sci Rep* 9:5954
- Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR (2015) A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* 526:569–573
- Reed KD, Meece JK, Henkel JS, Shukla SK (2003) Birds, migration and emerging zoonoses: west Nile virus, Lyme disease, influenza A and enteropathogens. *Clin Med Res* 1:5–12
- Shi ZL, Guo D, Rottier PJ (2016) Coronavirus: epidemiology, genome replication and the interactions with their hosts. *Virology* 511:1–2
- Woo PC, Lau SK, Lam CS, Lau CC, Tsang AK, Lau JH, Bai R, Teng JL, Tsang CC, Wang M, Zheng BJ, Chan KH, Yuen KY (2012) Discovery of seven novel mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. *J Virol* 86:3995–4008
- 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
- Zhu W, Yang J, Lu S, Lan R, Jin D, Luo XL, Pu J, Wu S, Xu J (2020) Beta- and novel delta-coronaviruses are identified from wild animals in the Qinghai-Tibetan Plateau, China. *Virology* 511:1–2. <https://doi.org/10.1007/s12250-020-00325-z>