VIRUS EVOLUTION

Virus Evolution, 2025, **11(1)**, veaf049
DOI: https://doi.org/10.1093/ve/veaf049
Advance Access Publication 00 Month 0000
Research Article

Surveillance of coronaviruses in wild aquatic birds in Hong Kong: expanded genetic diversity and discovery of novel subgenus in the *Deltacoronavirus*

Daisy Y. M. Ng[1], Wanying Sun[1], Thomas H. C. Sit[2], Christopher J. Brackman[2], Anne C.N. Tse[2], Christine H.T. Bui[2], Amy W. Y. Tang[1], Andrew N. C. Wong[1], Andrew T. L. Tsang[1], Joe C. T. Koo[1], Samuel M. S. Cheng[1], Malik Peiris[1,3], Dmitry V. Samborskiy[4], Alexander E. Gorbalenya[5], Alex W. H. Chin[1], Leo L. M. Poon[1,3,6,*]

[1]School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, G/F, Patrick Manson Building (North Wing), 7 Sassoon Road, Pokfulam, Hong Kong SAR, China
[2]Agriculture, Fisheries and Conservation Department, Government of the Hong Kong Special Administrative Region, 5/F, Cheung Sha Wan Government Offices, 303 Cheung Sha Wan Road, Sham Shui Po District, Kowloon, Hong Kong SAR, China
[3]Centre for Immunology and Infection, Unit 1111–1113, 11/F, Building 17W, 17 Science Park West Avenue, Hong Kong Science and Technology Park, Shatin, New Territories, Hong Kong SAR, China
[4]Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Building A, Leninskie Gory, Gagarinsky District, Moscow 119899, Russia
[5]Department of Medical Microbiology, Leiden University Center of Infectious Diseases, Leiden University Medical Center, Albinusdreef 2, P.O. Box 9600, E4-P, Leiden 2300 RC, South Holland, The Netherlands
[6]HKU-Pasteur Research Pole, School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, 7/F, Hong Kong Jockey Club Building for Interdisciplinary Research, 5 Sassoon Road, Pokfulam, Hong Kong SAR, China

*Corresponding author. L642A, Laboratory Block, School of Public Health, LKS Faculty of Medicine, The University of Hong Kong, 21 Sassoon Road, Pokfulam, Hong Kong SAR, China. E-mail: llmpoon@hku.hk

Abstract

Migratory birds may carry emerging viruses over long distances. Regular surveillance and metagenomic analysis were employed to explore the diversity of avian coronaviruses at Hong Kong's Mai Po Wetland. We tested a total of 3239 samples collected from 2018 to 2024, among which the prevalence rate of viruses of the genus *Gammacoronavirus* (64.4%) was higher than that of *Deltacoronavirus* (35.6%). The host species were identified for 79.8% of the coronavirus-positive samples. Two deltacoronaviruses with full-genome sequences and one nearly complete gammacoronavirus genome were identified in faecal samples of three bird species. We also predicted putative transcriptional regulatory sequences and 3CLpro and PLpro cleavage sites for these viruses. Results from our phylogenetic analysis and pairwise amino acid identity comparisons, using the International Committee on Taxonomy of Viruses classification criteria based on the DEmARC framework, indicate that black-faced spoonbill coronavirus (BSCoV, strain MP22-1474) prototypes a new subgenus. Great cormorant coronavirus (GCCoV, strain MP18-1070) and falcated duck coronavirus (FDCoV, strain MP22-196) belong to two previously known species while diverging most profoundly from known viruses of these species. Two recombination events may have contributed to the evolution of FDCoV MP22-196 in genome regions from ORF1b to the S gene and from the M gene to the N gene. The cophylogenetic analysis between avian hosts and coronaviruses provides evidence for a strong linkage between viruses of the genus *Gammacoronavirus* and the birds of order *Anseriformes*. This study highlights the importance of ongoing surveillance for coronaviruses in wild migratory birds.

Keywords: Gammacoronavirus; Deltacoronavirus; subgenus; virus surveillance; wild aquatic birds

Introduction

Coronaviruses are enveloped, positive-sense, single-stranded RNA viruses and are known to infect four groups of vertebrates, including mammals, birds, amphibians, and fish (Woo *et al.* 2023, Lauber *et al.* 2024). Numerous scholarly sources associate coronaviruses with incidences of respiratory, gastrointestinal, and neurological diseases (Weiss and Navas-Martin 2005, Alluwaimi *et al.* 2020, Zhou *et al.* 2021). In recent years, there has been a surge in global attention towards the Coronavirus Disease 2019 (COVID-19) pandemic caused by the zoonotic introduction of new severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) into humans. This development underscores how emerging coronaviruses with a high potential for transmission across species can significantly impact international public health (Coronaviridae Study Group of the International Committee on Taxonomy of 2020).

The International Committee on Taxonomy of Viruses (ICTV) *Coronaviridae* Study Group (CSG) is responsible for developing virus classification and taxonomic nomenclature for the family *Coronaviridae* (Coronaviridae Study Group of the International Committee on Taxonomy of 2020). According to the 2023 report, the family is divided into three subfamilies: *Orthocoronavirinae*, *Letovirinae*, and *Pitovirinae* (Woo *et al.* 2023). Currently, the *Orthocoronavirinae*

contains four genera: *Alphacoronavirus*, *Betacoronavirus*, *Deltacoronavirus*, and *Gammacoronavirus* (Woo *et al.* 2023). *Alphacoronavirus* and *Betacoronavirus* mainly infect mammals, while *Deltacoronavirus* and *Gammacoronavirus* predominantly infect birds but have occasionally been isolated from mammals (Woo *et al.* 2012). The classification criteria for subfamilies, genera, subgenus, and species are based on the application of the computational framework DEmARC (DivErsity pArtitioning by hieRarchical Clustering) to five universally conserved protein domains: chymotrypsin-like protease (3CLpro), nucleotidyl transferase (NiRAN), RNA-dependent RNA polymerase (RdRp) genes, zinc-binding domain (ZBD), and helicase superfamily 1 (HEL1) in viruses with the entire or nearly complete genome sequences (Lauber and Gorbalenya 2012a, Woo *et al.* 2023). The thresholds of classification ranks are derived within the DEmARC framework using analysis of weighted clustering cost (CC) for the pairwise patristic distances (PPDs), which are calculated from Bayesian or maximum likelihood phylogenetic trees (Lauber and Gorbalenya 2012a, Lauber *et al.* 2024). A hierarchical classification of coronaviruses provides an evolutionary framework for the comprehensive analysis of the obtained research results (Coronaviridae Study Group of the International Committee on Taxonomy of 2020, Lauber *et al.* 2024).

Birds are primary hosts of virus species in the genera *Gammacoronavirus* and *Deltacoronavirus*, whose infection may affect poultry production and lead to substantial economic losses (Cavanagh 2005). Infectious bronchitis virus (IBV), historically known as avian coronavirus, is associated with bronchial infections, nephritis, intestinal diseases, growth inhibition, and higher mortality in chicks (Cavanagh 2005, Weiss and Navas-Martin 2005, Hafez and Attia 2020). A recent study reported the detection of a virus similar to quail coronavirus UAE-HKU30 in Polish quails afflicted with enteritis (Domanska-Blicharz *et al.* 2019). A study from 2018 revealed that quail coronavirus UAE-HKU30 is classified within the same coronavirus species as porcine coronavirus HKU15 and sparrow coronavirus HKU17, suggesting interspecies transmission between birds and mammals (Lau *et al.* 2018, Vlasova *et al.* 2020). Of note, a case report in 2021 identified a *Deltacoronavirus* species infecting swine and birds in the plasma samples from three Haitian children with acute undifferentiated febrile illness (Lednicky *et al.* 2021). This discovery highlights the possibility that zoonotic coronaviruses might infect humans, subsequently undergoing evolutionary changes and adaptations. Accordingly, it is important to consider the potential for viruses of the *Gammacoronavirus* and *Deltacoronavirus* to spread among avian and mammalian hosts. Regular monitoring of these viruses within bird populations could facilitate understanding of the prevalence of these viruses and help assess the risk of transmission to mammalian hosts.

Birds are reservoirs for major emerging viruses. Their swarming behaviour and migratory capacities can pose a risk for long-distance virus transmission. Hong Kong's Mai Po Wetland (22°29′56″N 114°02′45″E) is located on the East Asian-Australasian Flyway, where a large number of migratory birds gather from October to April (winter) every year. When our team tested faecal samples from aquatic wild birds in Hong Kong in 2011, we observed frequent transmission of gammacoronaviruses among duck species (Chu *et al.* 2011). In this study, we report the findings of coronavirus surveillance conducted on faecal samples of the Mai Po aquatic wild birds in recent years. Based on the results of comparative genome and phylogenetic analysis of the complete gene sequences and one nearly complete genome, we identified two deltacoronaviruses and a gammacoronavirus. They were named great cormorant coronavirus (GCCoV) MP18-1070, black-faced spoonbill coronavirus (BSCoV) MP22-1474, and falcated duck coronavirus (FDCoV) MP22-196, respectively. Of these, BSCoV MP22-1474 was classified into a putative novel subgenus of genus *Deltacoronavirus*.

Materials and methods

Sample collection and processing

Randomly collected faecal swab samples from various sites within the Mai Po marshes in Hong Kong during the winters of 2018–19 ($N = 94$) and 2020–21 ($N = 94$). Subsequent sample collections were conducted between November 2022 and April 2023 ($N = 1524$) and October 2023 to April 2024 ($N = 1527$). Sterile swabs were used to sample fresh bird droppings found in a pond. The samples were preserved in individual vials containing 2 ml of the viral transport medium (VTM), kept in dry ice for transportation, and immediately processed for extraction once delivered to the laboratory. The components of the in-house prepared VTM included 25 g penicillin, 0.1 g ofloxacin, 0.2 g nystatin, 3.1 g polymycin B sulfate, 100 ml gentamycin, 2 g sulfamethoxazole, 0.4 g NaOH, and 19 g Medium 199 mixed with 2000 ml deionized water, which was then adjusted to pH 7.2 (Leung *et al.* 2007). The VTM samples were centrifuged at 8000 rpm for 10 min at a temperature of 4°C to separate the viral components. Before sample extraction, a process known as viral enrichment was performed. One hundred microlitres of the supernatant was mixed with phosphate-buffered saline (PBS) to reach 1000 μl and filtered through a 0.22 μm pore-size Millex-GP filter to eliminate cells and debris. Subsequently, 200 μl of the filtrate underwent treatment with 2 μl of RNase A at room temperature for 15 min, followed by treatment for 45 min at a temperature of 37°C with a mixture of nucleases, including 4 μl of Turbo DNase, 30 μl of 1× Dnase buffer, and 2 μl of Benzonase. Finally, the samples were promptly extracted using a NucliSENS® easyMag® (bioMérieux), an automated system designed for the extraction of total nucleic acids, following the manufacturer's instructions. The RNA was eluted in 50 μl of RNase-free water and was used as the template for reverse transcription-polymerase chain reaction (RT-PCR).

Screening of coronaviruses

Coronavirus screening was performed by amplifying a 440 bp fragment of the RdRp gene using a set of pan-coronavirus nested RT-PCR primers described previously (Chu *et al.* 2011). DNA was generated from RNA using a PrimeScript™ RT reagent Kit (Perfect Real Time) (TaKaRa) in a 10 μl reaction, followed by the manufacturer's protocol. First-round PCR was performed using Ex Taq® DNA Polymerase Hot-Start Version (TaKaRa). A 25 μl reaction mixture was composed of 2.5 μl 10× Buffer, 2 μl dNTP mixture (2.5 mM), 0.5 μl forward primer (100 μM, 5′-GGKTGGGAYTAYCCKAARTG-3′), 0.5 μl reverse primer (100 μM, 5′-TGYTGTS WRCARAAYTCRTG-3′), 0.125 μl TaKaRa Ex Taq HS, and 2 μl cDNA product, and sterilized distilled water was added up to 25 μl. Initial incubation of mixtures at 94°C for 1 min, then amplified by 40 cycles with 94°C for 30 s, 50°C for 30 s, and 72°C for 50 s, and final extension at 72°C for 5 min. Second-round PCR was performed using TaKaRa Taq™ DNA Polymerase Hot Start Version (TaKaRa). A 25 μl reaction mixture was composed of 2.5 μl 10× Buffer, 2 μl dNTP mixture (2.5 mM), 0.1 μl forward primer (100 μM, 5′-GGTTGGGACTATCCTAAGTGTGA-3′), 0.1 μl reverse primer (100 μM, 5′-CCATCATCAGATAGAATCATCAT-3′), 0.125 μl TaKaRa Taq HS, and 1 μl first PCR product, and sterilized distilled

water was added up to 25 μl. Initial incubation of mixtures at 94°C for 5 min, then amplified by 40 cycles with 94°C for 30 s, 50°C for 30 s, and 72°C for 50 s, and a final extension at 72°C for 7 min. Positive samples with a 440 bp amplicon, a region commonly used for identifying and classifying coronaviruses, were studied through Sanger sequencing. These sequences were compared with known reference sequences in the GenBank database.

Host identification

To confirm the bird species, coronavirus-positive samples subsequently underwent a DNA barcoding technology as previously outlined for host identification (Cheung *et al.* 2009, Chu *et al.* 2011). A nested PCR utilizing specific primers designed for avian mitochondrially encoded cytochrome c oxidase subunit I (COX1) (first round, forward primer 5′-GAYATRGCKTTYCCKCGKATRAA-3′ and reverse primer 5′-ATKGCYCAKACYATKCCYATRTA-3′; second round, forward primer 5′-GCKTTTCCKCGKATRAAYAAYAT-3′ and reverse primer 5′-CCYATRTAKCCRAAKGGYTCYTT-3′). The Ex Taq® DNA Polymerase Hot-Start Version was employed for the two-round PCR following the manufacturer's protocol. The PCR amplicon with an expected 670 bp product size was confirmed by Sanger sequencing and identified by Basic Local Alignment Search Tool (BLAST) from the National Center for Biotechnology Information (NCBI) database.

Complete genome sequencing

Complete genomes were amplified and sequenced using the RNA extracted from the original swab specimens as templates. Extracted RNAs were subjected to meta-transcriptome sequencing with the NovaSeq 6000 platform (Illumina, paired-end 150 bp) by Novogene (HK) Company Limited (Hong Kong, China). The process of constructing the meta-transcriptome sequencing library commenced with removing rRNAs using the Ribo-Zero™ Magnetic Gold Kit (Epidemiology) and the Ribo-Zero™ Magnetic Kit (Epicentre). Ribosomal RNAs from eukaryotes were depleted from the total RNA samples. The remaining RNAs were fragmented into fragments of ∼250–300 bp. These fragments were then reverse-transcribed into double-stranded cDNAs using random primers. End repair (150 bp), A-tailing, and adapter ligation steps were performed using the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (New England Biolabs). Subsequently, fragment size selection and PCR amplification were carried out to prepare the metatranscriptome library for sequencing. The library was quantified using Qubit and real-time PCR, and the size distribution was checked with a Bioanalyzer. Sequences of the 5′ and 3′ genomic ends were obtained by SMARTer® RACE 5′/3′ Kit (Takara), respectively. The expected size of PCR products was purified by gel and directly sequenced. Sequences were assembled to obtain the full-length genome sequences.

Phylogenetic and genomic analysis

Multiple sequence alignment was performed by using MAFFT v7.490 (Katoh and Standley 2013). Sanger sequence results were aligned with coronavirus sequences downloaded from the NCBI database. The raw data from meta-transcriptome sequencing were processed to remove adapters and low-quality reads using fastp v0.20.1 (Chen *et al.* 2018). *De novo* assembly was subsequently conducted using metaSPAdes (Nurk *et al.* 2017). To identify coronavirus-associated contigs, the assembled contigs were examined using BLASTN and BLASTX (Camacho *et al.* 2009), with references to the entire non-redundant protein (nr) and nucleotide (nt) databases, as well as to the coronavirus nt and nr databases acquired from NCBI (https://www.ncbi.nlm.nih.gov/). Geneious Prime 2022.2.2 (https://www.geneious.com/) was utilized to visualize and analyse the contigs. Genome polishing was performed using Pilon and BEDTools (Quinlan and Hall 2010, Walker *et al.* 2014). Coronavirus nucleotide alignments were used to build the phylogenetic tree using IQ-TREE v2.3.5 with the GTR + I + G model and bootstrap 1000 times (Nguyen *et al.* 2015, Kalyaanamoorthy *et al.* 2017, Hoang *et al.* 2018, Minh *et al.* 2020). The GTR + I + G model is a comprehensive framework that accounts for different nucleotide substitution rates and includes a proportion of invariable sites plus a discrete Gamma model for accurate phylogenetic analysis (Gu *et al.* 1995, Nguyen *et al.* 2015, Kalyaanamoorthy *et al.* 2017, Minh *et al.* 2020).

Open reading frame (ORF) predictions were performed using NCBI Open Reading Frame Finder (https://www.ncbi.nlm.nih.gov/orffinder/). Conserved domains and protein family analysis were identified using NCBI Conserved Domain Search (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) and InterProScan (https://www.ebi.ac.uk/interpro/) using the integrated protein databases CATH-Gene3D, CDD, HAMAP, NCBIfam, PANTHER, Pfam, PIRSF, PRINTS, PROSITE profiles and patterns, SFLD, SMART, and SUPERFAMILY. The putative cleavage sites of the nonstructural proteins encoded by ORF1ab were predicted by InterProScan and 3CLP web servers (http://www.computationalbiology.cn/3CLPHost/home.html) (Nikolskaya *et al.* 2007, Attwood *et al.* 2012, Sigrist *et al.* 2013, Akiva *et al.* 2014, Oates *et al.* 2015, Pedruzzi *et al.* 2015, Letunic *et al.* 2021, Mistry *et al.* 2021, Sillitoe *et al.* 2021, Chen *et al.* 2022, Thomas *et al.* 2022, Wang *et al.* 2023, Haft *et al.* 2024, Blum *et al.* 2025). The phylogenetic tree of 3CL-pro, NIRAN, RdRp, ZBD, HEL1, S, and N protein was also constructed by IQ-TREE v2.3.5 with bootstraps 1000 times and Shimodaira–Hasegawa-like approximate likelihood ratio test (SH-aLRT) 1000 times. The best-fit substitution model of amino acid was calculated using ModelFinder according to the Bayesian information criterion method (Guindon *et al.* 2010, Nguyen *et al.* 2015, Kalyaanamoorthy *et al.* 2017, Hoang *et al.* 2018, Minh *et al.* 2020). The Coronavirus ORF and Regulatory Sequence Identifier (CORSID) algorithm predicted transcription regulatory sequences (TRSs) in the coronavirus genome with high genome coverage, identifying them as optimal core sequences (Zhang *et al.* 2022).

Comparative sequence analysis and virus classification

Classification of the three discovered avian coronaviruses was conducted as described for SARS-CoV-2 (Coronaviridae Study Group of the International Committee on Taxonomy of 2020). Using the Viralis platform, the previously reported multiple protein sequence alignment of the five most conserved replicative proteins, 3CLpro, NiRAN, RdRp, ZBD, and HEL1, was extended for the subfamily *Orthocoronavirinae* (Gorbalenya *et al.* 2010). In total, 2111 coronaviruses were analysed, including three discovered avian viruses. Phylogenetic analysis using this multiple-sequence alignment was conducted by FastTree v2.1.8 (double precision, no SSE3) with the WAG+CAT model and gamma correction for branch lengths (Price *et al.* 2010). The PPDs were extracted from the phylogeny and used as input for CC-based analysis by DEmARC v1.5 to classify three avian viruses within established or tentative taxa in the family *Coronaviridae* (Lauber *et al.* 2012, Lauber and Gorbalenya 2012a, Lauber and Gorbalenya 2012b, Coronaviridae Study Group of the International Committee on Taxonomy of 2020, Brinton *et al.* 2021). Local minima in CC distribution serve as thresholds for PPD that define levels of hierarchical classification. Viruses with PPD below these thresholds are grouped into clusters that correspond to taxa of different ranks, starting from species, and depending on the dataset analysed. Previously established PPD thresholds for species, subgenus, and

genus ranks of the family were used for taxa delineation in genera *Deltacoronavirus* and *Gammacoronavirus*; they are 0.101, 0.212, and 0.883, respectively. Note that they are proxies for the respective local CC minima and are dependent on the maximum likelihood program used for phylogenetic analysis.

Recombinant analysis

The recombination events and breakpoints were identified by Recombination Detection Program 5 (RDP5) v5.61 with seven different scanning methods, namely, RDP, GENECONV, Chimaera, MaxChi, BootScan, SiScan, and 3Seq (Smith 1992, Padidam *et al.* 1999, Gibbs *et al.* 2000, Posada and Crandall 2001, Martin *et al.* 2005, Boni *et al.* 2007, Martin *et al.* 2021). Aligned nucleotide sequences were treated as linear, the window size was set at 1000 nt, and other parameters were set to default. Putative recombinant events were considered significant when a *P*-value equal to or smaller than 0.01 was observed for the same event using five or more algorithms and a recombination region length larger than 1000 nt. The region between the recombination breakpoints was further verified by similarity plot analysis with the SimPlot++ Program v1.3 using the nucleotide sequence alignment of the provisional major and minor parents and their recombinant (Samson *et al.* 2022). The sequence similarity was assessed using the Kimura distance model with a sliding window of 1000 nt and moving in 200 nt steps with 1000 bootstrap replicates.

Coevolutionary analyses of virus and host

ParaFit and PACo functions in the R program (R version 3.6.0) were used to analyse the evolutionary relationships between two phylogenetic trees (R Core Team 2024). ParaFit was employed to assess the congruence between the phylogenetic trees of hosts and viruses and identify which host–virus associations are significant in their coevolutionary relationship (Legendre *et al.* 2002). Function Parafit tested the null hypothesis of coevolution between a clade of hosts and a clade of viruses that are associated randomly. Significance is determined by randomizing the host–virus association matrix. Virus (RdRp gene, 440 nt) and host (COX1 gene, 670 bp) phylogenetic trees were built using IQ-TREE v2.3.5 with the GTR + I + G model and bootstrap 1000 times (Nguyen *et al.* 2015, Kalyaanamoorthy *et al.* 2017, Hoang *et al.* 2018, Minh *et al.* 2020). All coronavirus strains without host records were excluded from this analysis. The 'Cophenetic' function from the package ape in R calculated the matrices of patristic distances used in the global-fit analysis (Paradis *et al.* 2004). The Procrustean Approach to Cophylogeny (PACo) serves as a tool for evaluating the interdependence between topologies of two phylogenies and produces a Procrustes superimposition plot, which visually assesses the congruence of the virus phylogeny with the host phylogeny (Balbuena *et al.* 2013). Additionally, PACo provides a goodness-of-fit statistic based on 10 000 randomizations and was used to assess the significance of dependence between trees. The 'Cophylo' function from the package phytools 2.0 in R creates a co-phylogenetic plot (Revell 2024).

Results

Prevalence of coronaviruses

A total of 3239 wild bird faecal samples were collected across the winter seasons of 2018–19, 2020–21, 2022–23 (November to April), and 2023–24 (October to April) in the Hong Kong Mai Po Marshlands and tested for coronaviruses using RT-PCR. Coronavirus-positive rates in the four periods were 16.0%, 6.4%, 9.4%, and 5.4%, respectively (Table 1). In total, 247 samples were positive, of which 159 (64.4%) were gammacoronaviruses and 88 (35.6%) were deltacoronaviruses (Table 1). This study found that the prevalence of gammacoronaviruses was statistically significantly higher than that of deltacoronaviruses in the periods 2020–21 (*P*-value = .0105), 2022–23 (*P*-value < .0001), and 2023–24 (*P*-value = .0144), as determined by logistic regression with a binomial family and logit link, followed by Wald tests, which were used to compare proportions for each period relative to 2010. Host species identification by amplification of the COX1 gene was successful in 79.8% of the positive samples, with the majority identified as *Spatula clypeata* and *Mareca falcata* in the *Anseriformes* order, as well as *Recurvirostra avosetta* in the *Charadriiformes* order (Table 2).

Preliminary phylogenetic analysis of the recovered partial coronavirus RdRp sequences (*N* = 247), together with sequences of the already known coronaviruses obtained from GenBank, suggested that these coronaviruses belong to the genera *Gammacoronavirus* and *Deltacoronavirus* (Fig. 1, Table S1). We further analysed some of the tree clusters. One cluster of 42 viruses is situated between the Canada goose coronavirus (NC_046965) and Duck coronavirus isolate DK/GD/27 (DuCoV 2 714, NC_048214) within the *Gammacoronavirus* lineage. BLAST results show that they are closely related to the Avian coronavirus strain OH38 RdRp gene (KJ741881), with nucleotide identities ranging from 98.86% to 96.34%. Within the *Deltacoronavirus* lineage, a branch with 28 samples exhibits high similarity with the Avian coronavirus isolate J0569 RdRp gene (JN788788), showing nucleotide identities ranging from 93.00% to 93.96%. In addition, four samples from another branch have the nucleotide identity of 98.61%–97.92% with the Avian coronavirus isolate J1175 RdRp gene (JN788813), which is most closely related to the Night-heron coronavirus HKU19 (NHCoV HKU19, NC_016994) (Fig. 1).

Sequencing and genomic characterization of three coronaviruses

This study determined complete genome sequences of two deltacoronaviruses (GCCoV MP18-1070 and BSCoV MP22-1474) and a nearly complete genome sequence of one gammacoronavirus (FDCoV MP22-196). The deduced genome lengths are 26 306, 26 220, and 28 490 nt, respectively. The G + C content of the three viruses ranges from 39.4%, 37.9%, to 39.8%. The genome architecture of these three viruses is similar to other known coronaviruses, with typical genomic organization including 5′ untranslated region (UTR), replicase polyprotein ORF1ab, spike (S), envelope (E), membrane (M), and nucleocapsid (N) genes, and 3' UTR (Fig. 2, Table 3). The conserved coronavirus heptameric sequence 'UUUAAC' was identified in all three genomes. It is located in the overlap between ORF1a and ORF1b and induces a −1 ribosomal frameshift during translation in characterized coronaviruses (Kelly *et al.* 2021). In addition, putative accessory proteins gene NS6 and NS7ab located downstream of the S gene were identified in GCCoV MP18-1070 and BSCoV MP22-1474, while NS3a-c, NS6a-c, and NS7 were identified in FDCoV MP22-196 (Fig. 2, Table 3).

Coronavirus capsid and accessory genes are expressed using discontinuous transcription that is regulated by TRS at the intergene junctions (Zhang *et al.* 2022). Previous research has shown that the TRS of *Deltacoronavirus* and *Gammacoronavirus* have the unique sequences of 5'-ACAACCA-3′ and 5'-CUUAACAA-3′, respectively (Woo *et al.* 2009, Lau *et al.* 2018). These TRSs were identified in all three genome sequences determined in this study and according to the respective coronavirus genus. They preceded the start codon adenine-uracil-guanine (AUG) of ORF1ab and S gene and some accessory protein genes (NS6 and NS7) at a distance varying between 110 and 201 nt (Table 3).

Computational analysis of the ORF1ab-encoded polyprotein of these three coronaviruses by InterProScan and 3CLP servers

Table 1. Detection of coronaviruses in wild bird samples in 2018–19, 2020–21, 2022–23, and 2023–24 in Mai Po, Hong Kong

Year/Month for sample collection	Total no. of samples	No. of CoV-positive samples (% of total)	No. of CoV-positive samples for	
			Gammacoronavirus	Deltacoronavirus
2018–19	94	15 (16.0)	9	6
2020–21	94	6 (6.4)	5	1
2022–23	1 524	144 (9.4)	97	47
November 2022	216	11(5.1)	11	0
December 2022	288	40 (13.9)	38	2
January 2023	288	13 (4.5)	12	1
February 2023	300	39 (13)	19	20
March 2023	288	33 (11.5)	17	16
April 2023	144	8 (5.6)	0	8
2023–24	1 527	82 (5.4)	48	34
October 2023	143	1 (0.7)	1	0
November 2023	288	3 (1)	2	1
December 2023	288	12 (4.2)	12	0
January 2024	288	21 (7.3)	9	12
February 2024	288	24 (8.3)	21	3
March 2024	216	21 (9.7)	3	18
April 2024	16	0 (0)	0	0
Total	3 239	247	159	88

Table 2. Overview of host identification for coronavirus-positive samples in 2018–19, 2020–21, 2022–23, and 2023–24 in Mai Po, Hong Kong

Avian (order/family)	Common name (species)	No. of CoV-positive samples (% of total)	No. of CoV-positive samples in different year			
			2018–19	2020–21	2022–23	2023–24
Anseriformes/Anatidae	Common teal (*Anas crecca*)	22 (8.9)	1	1	17	3
	Eurasian wigeon (*Mareca penelope*)	7(2.8)	0	0	7	0
	Falcated duck (*Mareca falcata*)	30 (12.2)	3	1	25	1
	Gadwall (*Mareca strepera*)	1 (0.4)	0	0	1	0
	Northern shoveler (*Spatula clypeata*)	45 (18.2)	5	0	22	18
	Northern pintail (*Anas acuta*)	23 (9.3)	0	2	21	0
Charadriiformes/ Charadriidae	Grey plover (*Pluvialis squatarola*)	1 (0.4)	0	0	0	1
	Lesser sand plover (*Charadrius mongolus*)	1 (0.4)	0	0	1	0
	Long-billed plover (*Charadrius placidus*)	2 (0.8)	0	0	0	2
	Pacific golden plover (*Pluvialis fulva*)	2 (0.8)	0	1	1	0
Charadriiformes/ Recurvirostridae	Pied avocet (*Recurvirostra avosetta*)	44 (17.8)	0	0	11	33
Charadriiformes/ Scolopacidae	Common greenshank (*Tringa nebularia*)	1 (0.4)	0	0	1	0
Pelecaniformes/ Arteidae	Eastern great egret (*Ardea modesta*)	3 (1.2)	0	0	3	0
	Grey heron (*Ardea cinerea*)	6 (2.4)	1	0	3	2
Pelecaniformes/ Threskiornithidae	Black-faced spoonbill (*Platalea minor*)	7(2.8)	0	0	7	0
Suliformes/ Phalacrocoracidae	Great cormorant (*Phalacrocorax carbo*)	2 (0.8)	2	0	0	0
	Unidentified host	50 (20.2)	3	1	24	22
	Total	247	15	6	144	82

predicted 15 putative nonstructural proteins (nsp2–nsp16) that are separated by PLpro and 3CLpro cleavage sites (Table 4). These proteins comprise the five most conserved replication enzymes and cofactors, including 3CLpro found in nsp5, NiRAN and RdRp found in nsp12, and ZBD and HEL1 found in nsp13 (Fig. 2). Notably, we found that the P2P1 dipeptides in the putative PLpro cleavage sites at nsp2/nsp3 and nsp3/nsp4 junctions of GCCoV MP18-1070 and BSCoV MP22-1474 are VG, TG, and KG, which deviate from the AG amino acids common in other coronaviruses (Table 4). The P2P1 dipeptides of the putative 3CLpro cleavage sites at nsp4/nsp5 to nsp15/nsp16 are commonly VQ and LQ in these three viruses. The only exceptions are the CQ and TQ amino acids at nsp8/nsp9 and nsp13/nsp14 in BSCoV MP22-1474 (Table 4).

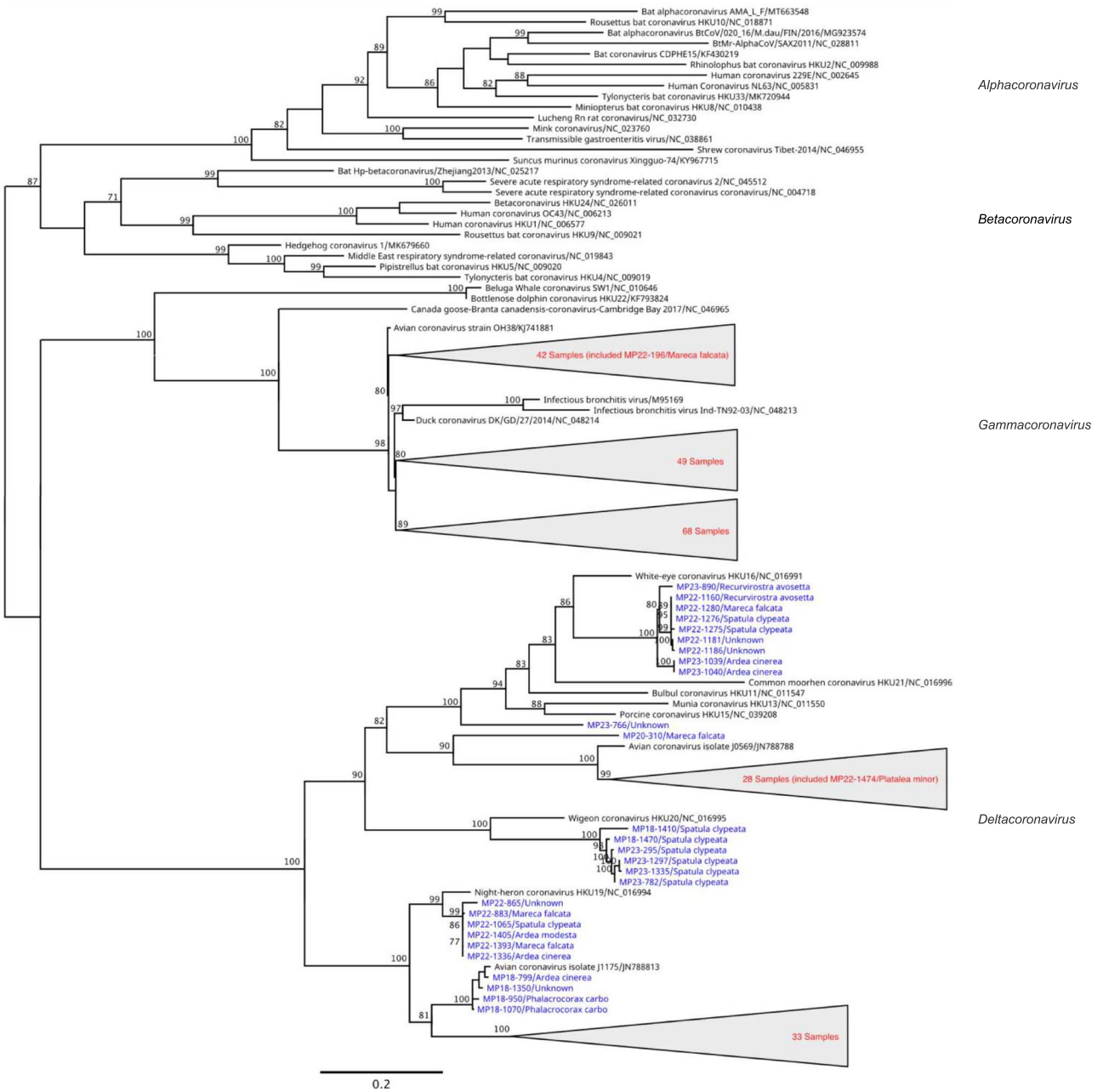

Figure 1. Phylogenetic analysis on the RdRp region (360 bp) of coronaviruses. The trees are constructed by IQ-TREE v2.3.5 with the GTR + I + G substitution model. The tree is midpoint rooted. Bootstrap values are set up with 1000 replicates as statistical support. The branch values are bootstrap supports (%). Coronaviruses detected in this study ($N = 247$) are highlighted in colour and named with a sample number, followed by the corresponding identical host species. Samples related to the same groups of similar nodes have been collapsed and indicated by the isosceles triangles figure, and the total number of sequences in each group is shown next to the triangles. The GenBank accession numbers of retrieved genes are indicated at the end of the reference name. Four main monophyletic clusters corresponding to the genera *Alphacoronavirus*, *Betacoronavirus*, *Deltacoronavirus*, and *Gammacoronavirus* within the subfamily *Orthocoronavirinae* are indicated on the right side of this figure.

Phylogenetic analysis of coronaviruses

We then determined the phylogenetic relationship of the three coronaviruses with reference sequences of the *Coronaviridae* ($N = 41$, Table S2) using the amino acid sequences of the seven conserved domains (3CLpro, NiRAN, RdRp, ZBD, HEL1, S, and N) as shown in Fig. 3. The corresponding pairwise amino acid identities are shown in Table 5. All seven phylogenetic trees showed that GCCoV MP18-1070 is closely related to the NHCoV HKU19 (NC_016994) (Fig. 3). In contrast, BSCoV MP22-1474 forms a separate branch and clusters within the deltacoronavirus lineage, except for the N gene–based tree, in which it is most closely related to Wigeon coronavirus HKU20 (NC_016995) (Fig. 3). Phylogenetic analysis also showed that FDCoV MP22-196 clusters within the gammacoronavirus lineage and is most closely related to DuCoV 2714 (NC_048214) for RdRp, ZBD, HEL1, and N phylogenetic trees (Fig. 3).

Comparisons of the pairwise nucleotide identity of full genomes and the pairwise amino acid identities of the seven conserved domains showed that GCCoV MP18-1070 is most similar to NHCoV HKU19 (NC_016994): 73.5% nt identity and (98.1%–66.3%) aa identity (Table 5). BSCoV MP22-1474 shows the highest nucleotide identity to Bulbul coronavirus HKU11 (NC_011547) at 54.8%,

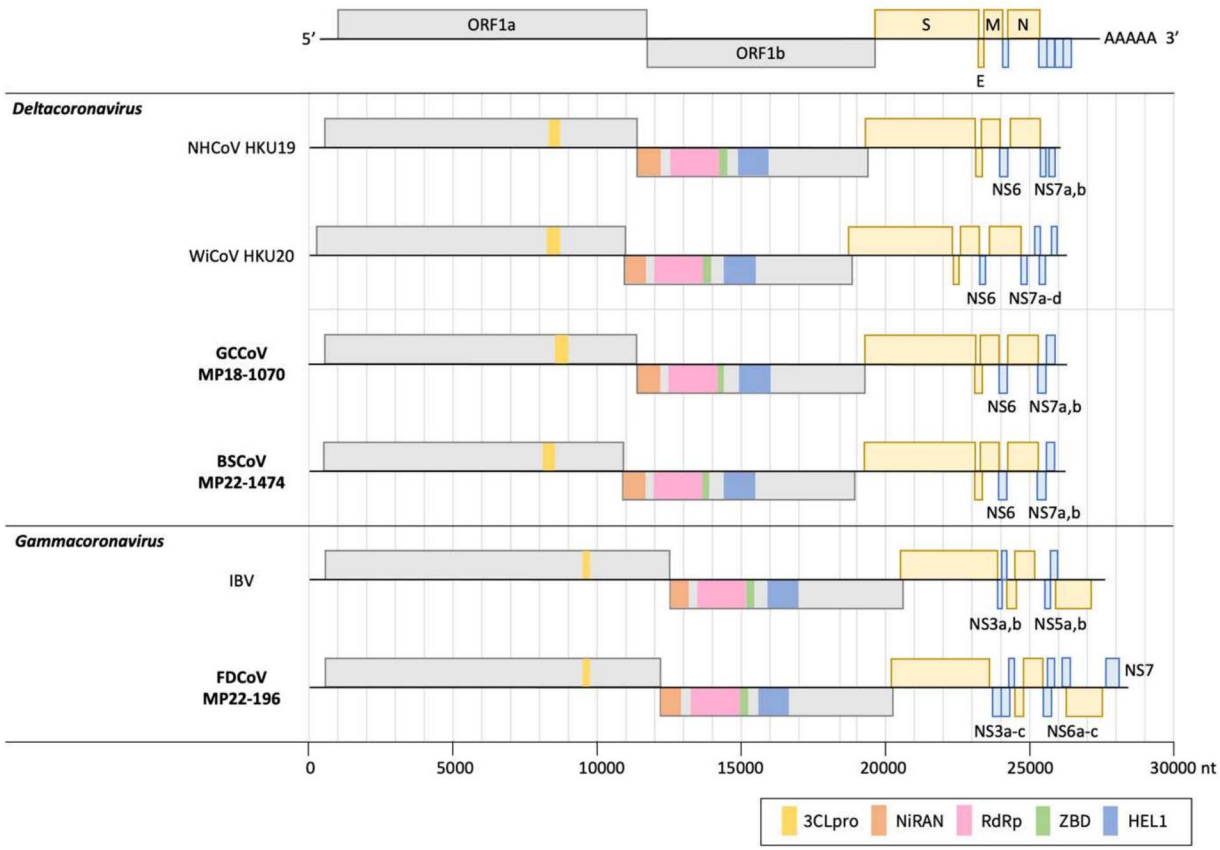

Figure 2. Genome and domain organization of coronaviruses discovered in this study with night-heron coronavirus HKU19 (NHCoV HKU19, NC_016994), Wigeon coronavirus HKU20 (WiCoV HKU20, NC_016995), and infectious bronchitis virus (IBV, NC_048213). Three coronaviruses are shown in bold. The upper indicated ORF1a and ORF1b regions and downstream of the S, E, M, and N genes are represented by labeled boxes. Putative accessory ORFs located before and after the N gene region are represented by outlined boxes. 3CLpro, NIRAN, RdRp, ZBD, and HEL 1 are annotated within the ORF1a and ORF1b regions. These domain combinations are used for phylogenetic and DEmARC analysis of the family *Coronaviridae*.

while its seven conserved domains show the highest amino acid identities to various deltacoronaviruses at 57.7%–88.9% (Table 5). For FDCoV MP22-196, the highest pairwise nucleotide identity at 71.8%, along with the highest amino acid identities surpassing 80% across most conserved domains (NiRAN, RdRp, ZBD, HEL1, and N), was found in comparison with DuCoV 2714 (NC_048214) (Table 5). The results of these phylogenetic tree analyses and pairwise amino acid differences, GCCoV MP18-1070 and BSCoV MP22-1474 belong to the genus *Deltacoronavirus*, while FDCoV MP22-196 belongs to the genus *Gammacoronavirus*. We further refined the classification of these three viruses, as described below.

Classification of three newly reported avian coronaviruses below genus rank

To quantify the relation of the newly identified avian coronaviruses with other coronaviruses below genus rank, we used DEmARC, which employs a cost-based function for PPD clustering and was previously used to establish a framework for the taxonomy classification of nidoviruses (Fig. 4). Accordingly, BSCoV MP22-1474 prototypes a new subgenus in the genus *Deltacoronavirus* since its most closely related coronavirus, Common-moorhen coronavirus HKU21 (NC_016996), was separated by 0.306 PPD, exceeding the subgenus threshold of 0.212. In contrast, GCCoV MP18-1070 belongs to a species that also includes NHCoV HKU19 (NC_016994) of the genus *Deltacoronavirus*, and FDCoV MP22-196 belongs to a species that also includes DuCoV 2714 (NC_048214) of the genus *Gammacoronavirus*. Both GCCoV MP18-1070 and FDCoV MP22-196 diverged considerably from the single known virus and two known viruses of these species, respectively, indicating their novelty. As shown in Table 5, the S proteins of these viruses have lower amino acid homology, with pairwise identities of 66.3% for GCCoV MP18-1070 and NHCoV HKU19 (NC_016994), and 29.4% for FDCoV MP22-196 and DuCoV 2714 (NC_048214).

Recombination analysis

The full-length genome sequences of three coronaviruses identified in this study were screened for potential recombination events using RDP5 software and a set of the most closely related coronaviruses, which were considered as putative parents (Table S3). Due to the relatively low pairwise nucleotide identity between most of the genomes involved in this analysis (putative parents and query sequence), the closest available sequences were considered parental proxies for inferring recombination events. No evidence of recombination was found for GCCoV MP18-1070 and BSCoV MP22-1474. In contrast, two putative recombination events might have contributed to the evolution of FDCoV MP22-196, according to all seven methods used (Fig. 5, Table S4). The first recombination breakpoint is located in a region encompassing the junction of the ORF1b and S gene (18 650–23 044 nt) and was supported with *P*-values ranging from 2.565×10^{-317} to 6.847×10^{-21} (Fig. 5A, Table S4). Bird gammacoronavirus AnasCN24 PP845446 and PP845443 detected in Shanghai were

Table 3. Coding potentials and putative transcription regulatory sequences of two deltacoronavirus genomes and one gammacoronavirus genome

Virus	ORF	Location (nt)			Length (nt)	Length (aa)	Frame(s)	TRS location (nt)	TRS sequence(s) (distance in bases to AUG)[a]
GCCoV	1ab	496	-	19 412	18 917	6306	+1, +3	65	ACACCG(425)AUG
MP18-1070	S	19 394	-	23 245	3852	1283	+2	19 245	ACACCG(143)AUG
	E	23 245	-	23 502	258	85	+1	23 197	ACACAC(42)AUG
	M	23 495	-	24 145	651	216	+2	23 387	ACACCG(102)AUG
	NS6	24 145	-	24 423	279	92	+1	23 987	ACACCG(152)AUG
	N	24 434	-	25 489	1056	351	+2	24 416	ACACCU(12)AUG
	NS7a	25 489	-	25 785	297	98	+1	25 461	ACACCG(22)AUG
	NS7b	25 793	-	26 089	297	98	+2	25 736	ACAUCA(51)AUG
BSCoV	1ab	457	-	18 917	18 460	6154	+1, +3	63	ACACCA(388)AUG
MP22-1474	S	18 869	-	22 627	3759	1252	+2	18 753	ACACCA(110)AUG
	E	22 627	-	22 881	255	84	+1	22 580	ACACCG(34)AUG
	M	22 884	-	23 534	651	216	+3	22 848	ACACCA(30)AUG
	NS6	23 534	-	23 806	273	90	+2	23 374	ACACCA(152)AUG
	N	23 820	-	24 872	1053	350	+3	23 806	ACACCA(8)AUG
	NS7a	24 655	-	25 038	384	127	+1	24 525	ACACCU(124)AUG
	NS7b	25 007	-	25 375	369	122	+2	24 916	ACACUU(85)AUG
	NS7c	25 539	-	25 814	276	91	+3	25 388	ACACCA(145)AUG
FDCoV	1ab	445	-	20 207	19 762	6588	+1, +3	36	CUUUAGCG(359)AUG
MP22-196	S	20 158	-	23 577	3420	1139	+1	20 014	AUUUGCAA(136)AUG
	NS3a	23 641	-	23 970	330	109	+1	23 546	GUGAACAA(87)AUG
	NS3b	23 974	-	24 279	306	101	+1	23 952	UUUAGUAC(14)AUG
	NS3c	24 293	-	24 505	213	70	+2	24 251	CUUCAGAG(7)AUG
	E	24 505	-	24 807	303	100	+1	24 445	CCUUUCAA(52)AUG
	M	24 804	-	25 475	672	223	+3	24 775	AAUAACAA(21)AUG
	NS6a	25 476		25 760	285	94	+3	25 365	GAUAACAA(103)AUG
	NS6b	25 660	-	25 905	246	81	+1	25 612	GUAAACAA(40)AUG
	NS6c	26 096	-	26 365	270	89	+2	25 887	CUUAACAA(201)AUG
	N	26 275	-	27 531	1257	418	+1	26 258	AAUAACAA(9)AUG
	NS7	27 735	-	28 187	453	150	+3	27 594	CUUAGCAC(133)AUG

[a]The nucleotide variations from 'ACACCA' and 'CUUAACAA' are underlined.

Table 4. Putative PLpro and 3CLpro cleavage sites at the junctions between nonstructural proteins in polyproteins encoded by two deltacoronavirus genomes and one gammacoronavirus genome

	Putative cleavage site		
nsp pair	GCCoV MP18-1070	BSCoV MP22-1474	FDCoV MP22-196
nsp2/nsp3	YKVG↓GL	YKVG↓GS	VKAG↓GK
nsp3/nsp4	NKTG↓GS	SKKG↓GA	KKAG↓GF
nsp4/nsp5	TVVQ↓AG	AVVQ↓AG	SRLQ↓SG
nsp5/nsp6	INLQ↓ST	VNLQ↓SY	VVLQ↓SS
nsp6/nsp7	STVQ↓NK	STVQ↓NK	AVVQ↓SK
nsp7/nsp8	PVLQ↓VV	PVLQ↓VV	TVLQ↓SV
nsp8/nsp9	IKLQ↓NN	VVCQ↓NN	ASLQ↓NN
nsp9/nsp10	VTLQ↓AS	VTLQ↓AN	VVLQ↓SK
nsp10/nsp11	NKLQ↓SV	GKLQ↓FN	ATVQ↓GT
nsp12/nsp13	PVLQ↓AT	PTLQ↓AT	TTLQ↓SC
nsp13/nsp14	PLVQ↓SL	VKTQ↓CS	VALQ↓ST
nsp14/nsp15	TNLQ↓TL	TNLQ↓TL	SSLQ↓SI
nsp15/nsp16	PIVQ↓AL	PVVQ↓AL	PQLQ↓SA

identified as the putative major and minor parents of these recombination events, respectively (Fig. 5A and B). The second recombination breakpoint is located over the junction between the M and N genes (25 259–27 472 nt) and was supported with *P*-values ranging from 3.971×10^{-99} to 6.762×10^{-29} (Fig. 5C, Table S4). Bird gammacoronavirus AnasCN24 PP845387 was identified as the putative major parent, while Avian coronavirus MK204411 was identified as the minor parent in the second recombination event (Fig. 5C and D). Further, to characterize the putative recombination events, we used SimPlot++ to compare the nucleotide sequence similarity of FDCoV MP22-196 with the major and minor parental sequences identified by RDP5. The similarity plots showed higher sequence identity between FDCoV MP22-196 and Bird gammacoronavirus AnasCN24 PP845443 at the

Table 5. Comparison of genomic features and percentage nucleotide and amino acid identity of GCCoV MP18-1070, BSCoV MP22-1474, FDCoV MP22-196, and representative coronaviruses used as references

						Pairwise identity (%)															
						GCCoV MP18-1070								BSCoV MP22-1474							
Subfamily	Genus	Subgenus	Virus name	Accession no.	Sequence Length (nt)	Genome (nt)	3CLpro (aa)	NiRAN (aa)	RdRp (aa)	ZBD (aa)	HEL1 (aa)	S (aa)	N (aa)	Genome (nt)	3CLpro (aa)	NiRAN (aa)	RdRp (aa)	ZBD (aa)	HEL1 (aa)	S (aa)	N (aa)
Letovirinae	*Alphaletovirus*	*Milecovirus*	Microhyla letovirus 1	GECV01031551	22 303	24.6	15.8	27.2	43.8	31.8	37.3	8.80	10.8	24.8	14.2	25.8	45.2	31.8	37.8	8.50	11.8
Pitovirinae	*Alphapironavirus*	*Samovirus*	Pacific salmon nidovirus	MK611985	36 652	26.7	18.6	20.8	44.6	23.1	32.8	9.00	7.20	27.2	18.6	21.0	44.1	24.8	33.3	7.90	8.00
Orthocoronavirus	*Alphacoronavirus*	*Amalacovirus*	Bat alphacoronavirus AMA_L_F	MT663548	29 097	29.7	28.6	37.2	53.2	49.1	46.6	31.1	14.5	30.0	33.0	37.2	53.7	52.8	46.3	31.2	16.2
		Colacovirus	Bat coronavirus CDPHE15	KF430219	28 035	30.7	30.4	35.8	52.5	46.3	46.3	32.9	14.3	31.2	34.8	36.8	55.0	50.0	46.1	32.3	15.4
		Decacovirus	Rousettus bat coronavirus HKU10	NC_018871	28 494	31.1	32.1	35.1	53.0	50.0	46.3	33.0	12.7	31.2	32.1	35.8	53.7	52.8	46.1	33.7	13.9
		Duvinacovirus	Human coronavirus 229E	NC_002645	27 317	31.7	31.3	35.1	53.2	51.9	45.8	37.2	14.7	31.5	33.0	36.1	54.8	51.9	46.3	38.3	14.4
		Luchacovirus	Lucheng Rn rat coronavirus	NC_032730	28 763	29.0	26.8	35.8	54.3	47.2	47.5	18.1	16.2	29.0	32.1	36.8	54.4	50.0	47.1	17.3	15.1
		Minacovirus	Mink coronavirus	NC_023760	28 941	30.5	25.9	35.8	54.6	49.1	47.5	32.9	15.9	30.7	31.3	36.5	55.1	49.1	47.5	32.4	15.3
		Minunacovirus	Miniopterus bat coronavirus HKU8	NC_010438	28 773	30.9	30.4	36.5	54.3	46.3	46.1	32.1	13.8	30.6	31.3	36.1	55.0	50.9	46.3	31.7	14.8
		Myotacovirus	BtMr-AlphaCoV/SAX2011	NC_028811	27 935	30.7	33.0	37.2	53.6	45.4	46.8	32.6	12.8	31.1	34.8	36.8	55.1	49.1	46.6	32.0	13.5
		Nyctacovirus	Tylonycteris bat coronavirus HKU33	MK720944	27 636	31.4	26.8	36.5	53.6	48.1	47.1	30.8	14.6	31.5	33.0	36.5	55.3	50.9	45.6	30.3	13.3
		Pedacovirus	Bat alphacoronavirus BtCoVBtCoV/020_16/M.dau/FIN/2016	MG923574	28 045	31.2	33.9	34.0	52.7	46.3	44.4	33.0	15.2	31.0	32.1	34.7	53.6	50.0	44.9	33.5	15.0
		Rhinacovirus	Rhinolophus bat coronavirus HKU2	NC_009988	27 165	29.7	25.0	35.1	54.1	50.9	48.0	17.2	15.5	29.9	30.4	36.1	54.6	50.9	47.3	17.0	16.2
		Setracovirus	Human Coronavirus NL63	NC_005831	27 553	31.7	30.4	35.1	52.5	47.2	45.8	33.3	16.7	31.7	35.7	35.4	53.2	50.0	46.3	33.1	16.4
		Soracovirus	Shrew coronavirus Tibet-2014	NC_046955	27 102	29.0	28.3	35.5	53.2	54.6	47.3	17.6	15.3	29.2	29.2	35.8	54.8	53.7	47.8	17.8	15.3
		Sunacovirus	*Suncus murinus* coronavirus Xingguo-74	KY967715	25 984	28.7	27.7	36.9	53.8	50.0	45.9	17.0	15.1	29.0	27.7	36.8	54.2	54.6	45.7	16.5	15.6
		Tegacovirus	Transmissible gastroenteritis virus	NC_038861	28 586	30.9	28.6	28.4	54.4	53.7	47.1	32.9	14.6	30.8	32.1	31.9	54.6	54.6	47.3	32.4	14.7
	Betacoronavirus	*Embecovirus*	Human coronavirus OC43	NC_006213	30 741	27.7	31.3	40.5	57.7	49.1	45.6	17.3	17.0	28.0	28.6	39.4	57.0	48.1	46.8	17.3	16.8
		Embecovirus	Human coronavirus HKU1	NC_006577	29 926	29.3	29.5	42.3	57.5	50.0	44.4	16.9	16.8	29.8	31.3	40.4	56.4	50.9	45.8	16.9	16.2
		Embecovirus	Betacoronavirus HKU24	NC_026011	31 249	27.4	31.3	39.1	57.5	49.1	44.5	16.8	16.7	27.6	31.3	39.0	57.3	48.1	46.4	16.7	15.9
		Hibecovirus	Bat Hp-betacoronavirus/Zhejiang2013	NC_025217	31 491	34.3	22.1	38.1	57.3	49.1	47.5	17.8	15.4	34.6	25.7	37.7	58.7	51.9	48.0	17.3	18.3
		Merbecovirus	MERS coronavirus	NC_019843	30 119	28.3	25.9	39.2	54.9	47.2	47.8	18.1	17.3	28.6	28.6	34.6	56.1	48.1	46.8	18.8	18.0
		Merbecovirus	Hedgehog coronavirus 1	MK679660	30 172	28.5	25.9	39.1	54.4	47.2	48.0	18.2	16.6	28.4	29.5	35.5	56.3	50.0	45.8	18.2	17.3
		Merbecovirus	Pipistrellus bat coronavirus HKU5	NC_009020	30 482	33.6	25.0	39.1	55.4	47.2	47.3	18.8	16.1	33.0	28.6	35.5	56.1	48.1	47.5	18.9	17.2
		Merbecovirus	Tylonycteris bat coronavirus HKU4	NC_009019	30 286	28.6	26.8	39.1	55.7	48.1	47.3	19.8	16.1	28.5	30.4	35.2	56.4	49.1	46.8	19.8	16.4
		Nobecovirus	Rousettus bat coronavirus HKU9	NC_009021	29 114	34.2	24.8	37.2	57.1	48.1	48.9	17.8	16.1	33.6	32.7	37.2	57.7	50.0	47.4	17.7	18.2
		Sarbecovirus	SARS coronavirus Tor2	NC_004718	29 751	33.1	23.9	36.5	55.9	51.9	47.1	18.6	17.1	33.1	26.5	34.4	57.8	51.9	48.3	18.1	17.6
		Sarbecovirus	SARS 2	NC_045512	29 903	33.5	23.0	38.7	56.3	51.9	46.8	18.1	17.4	33.7	26.5	36.8	58.4	51.9	48.3	18.0	17.9
	Deltacoronavirus	*Andecovirus*	Wigeon coronavirus HKU20	NC_016995	26 227	33.9	56.8	52.8	76.7	81.5	70.4	37.9	52.5	35.2	**57.7**	52.2	78.6	85.2	72.1	38.2	**62.4**
		Buldecovirus	Common moorhen coronavirus HKU21	NC_016996	26 223	33.4	45.0	60.0	77.9	82.4	72.6	39.2	52.1	34.7	54.1	61.5	82.5	86.1	**77.1**	38.3	55.1
		Buldecovirus	Munia coronavirus HKU13	NC_011550	26 552	50.2	45.0	57.4	79.9	83.3	72.3	37.7	48.1	53.4	50.5	57.9	**82.6**	87.0	74.3	38.5	53.2
		Buldecovirus	Bulbul coronavirus HKU11	NC_011547	26 487	51.6	43.2	58.5	78.8	82.4	72.3	36.9	47.7	**54.8**	53.2	60.1	80.7	87.0	74.5	37.7	50.3
		Buldecovirus	Porcine coronavirus HKU15	NC_039208	25 425	51.1	45.0	58.5	79.0	83.3	71.4	37.4	47.8	53.2	50.5	60.1	81.2	**88.9**	74.8	38.1	52.4
		Buldecovirus	White-eye coronavirus HKU16	NC_016991	26 041	51.1	45.9	58.1	77.8	83.3	71.8	36.4	45.0	54.7	55.9	58.6	81.2	87.0	75.9	37.7	50.4
		Herdecovirus	Night-heron coronavirus HKU19	NC_016994	26 077	**73.5** [a]	**82.0**	**82.6**	**96.3**	**98.1**	**91.6**	**66.3**	**78.9**	54.1	49.5	**64.7**	79.9	86.1	75.1	**59.4**	57.3
	Gammacoronavirus	*Brangacovirus*	Canada goose coronavirus	NC_046965	28 539	29.8	35.3	39.2	59.3	55.6	53.7	22.8	25.3	30.1	37.1	38.8	60.2	60.2	54.9	23.0	26.5
		Cegacovirus	Beluga Whale coronavirus SW1	NC_010646	31 686	26.9	29.6	38.8	57.3	58.3	52.7	20.5	24.2	27.2	27.8	39.9	58.0	62.0	50.2	20.1	27.6
		Cegacovirus	Bottlenose dolphin coronavirus HKU22	KF793824	31 769	26.9	29.6	39.2	57.7	58.3	52.7	20.4	23.5	27.1	27.8	40.2	58.5	63.0	50.2	19.4	26.7
		Igacovirus	Duck coronavirus isolate DK/GD/27	NC_048214	27 754	30.7	35.3	37.8	59.3	54.6	54.5	25.9	24.4	30.7	38.8	38.2	59.9	59.3	54.3	24.8	27.1
		Igacovirus	Infectious bronchitis virus	M95169	27 608	30.9	36.2	31.8	58.8	57.4	53.3	22.7	23.4	31.0	37.1	31.5	59.9	62.0	53.3	22.0	25.2
		Igacovirus	Infectious bronchitis virus Ind-TN92–03	NC_048213	27 464	30.4	34.5	33.7	56.0	56.5	47.4	22.9	23.6	30.4	37.1	34.0	56.5	61.1	47.9	22.3	25.0

(continued)

Table 5. Continued.

Subfamily	Genus	Subgenus	Virus name	Accession no.	Sequence length (nt)	Pairwise identity (%) FDCoV MP22-196							
						Genome (nt)	3CLpro (aa)	NiRAN (aa)	RdRp (aa)	ZBD (aa)	HEL1 (aa)	S (aa)	N (aa)
Letovirinae	*Alphaletovirus*	*Milecovirus*	Microhyla letovirus 1	GECV01031551	22303	23.5	14.5	23.8	45.4	27.3	40.8	8.40	11.5
Pitovirinae	*Alphapironavirus*	*Samovirus*	Pacific salmon nidovirus	MK611985	36652	20.6	13.7	23.9	45.4	30.8	39.7	7.60	9.30
Orthocoronavirus	*Alphacoronavirus*	*Amalacovirus*	Bat alphacoronavirus AMA_L_F	MT663548	29097	31.7	37.5	43.6	61.7	60.2	54.1	18.8	18.6
		Colacovirus	Bat coronavirus CDPHE15	KF430219	28035	33.0	31.3	44.9	61.6	60.2	55.5	19.6	18.8
		Decacovirus	Rousettus bat coronavirus HKU10	NC_018871	28494	33.1	31.3	45.6	63.0	60.2	54.3	20.6	17.5
		Duvinacovirus	Human coronavirus 229E	NC_002645	27317	33.6	35.7	45.3	61.2	59.3	55.4	22.7	18.3
		Luchacovirus	Lucheng Rn rat coronavirus	NC_032730	28763	31.2	33.0	42.2	61.4	60.2	55.5	19.9	20.8
		Minacovirus	Mink coronavirus	NC_023760	28941	32.0	32.1	44.6	62.6	55.6	56.8	19.0	18.6
		Minunacovirus	Miniopterus bat coronavirus HKU8	NC_010438	28773	32.8	32.1	44.6	63.7	57.4	55.1	19.5	17.1
		Myotacovirus	BtMr-AlphaCoV/SAX2011	NC_028811	27935	32.7	31.3	45.6	63.8	55.6	55.1	18.7	14.5
		Nyctacovirus	Tylonycteris bat coronavirus HKU33	MK720944	27636	33.0	30.4	44.9	63.7	59.3	53.3	19.2	17.3
		Pedacovirus	Bat alphacoronavirus BtCoVBtCoV/020_16/M.dau/FIN/2016	MG923574	28045	32.7	33.9	41.8	62.6	62.0	55.6	19.9	17.7
		Rhinacovirus	Rhinolophus bat coronavirus HKU2	NC_009988	27165	32.7	35.7	45.3	63.5	55.6	56.5	19.0	18.6
		Setracovirus	Human Coronavirus NL63	NC_005831	27553	34.0	32.1	43.9	61.7	59.3	55.4	21.7	18.6
		Soracovirus	Shrew coronavirus Tibet-2014	NC_046955	27102	32.3	31.9	47.0	60.5	55.6	58.0	18.4	17.2
		Sunacovirus	*Suncus murinus* coronavirus Xingguo-74	KY967715	25984	31.3	25.0	47.0	63.2	62.0	55.3	19.7	17.3
		Tegacovirus	Transmissible gastroenteritis virus	NC_038861	28586	32.4	33.9	39.0	62.1	59.3	56.0	19.0	20.3
	Betacoronavirus	*Embecovirus*	Human coronavirus OC43	NC_006213	30741	30.5	33.0	41.1	67.1	64.8	55.8	16.8	18.6
		Embecovirus	Human coronavirus HKU1	NC_006577	29926	32.8	32.1	42.2	65.7	65.7	53.7	17.6	17.3
		Embecovirus	Betacoronavirus HKU24	NC_026011	31249	34.8	34.8	44.3	67.2	64.8	55.6	17.3	19.1
		Hibecovirus	Bat Hp-betacoronavirus/Zhejiang2013	NC_025217	31491	28.5	22.6	43.7	65.3	63.9	57.5	17.6	21.8
		Merbecovirus	MERS coronavirus	NC_019843	30119	35.2	34.8	45.7	65.9	67.6	57.8	16.8	21.0
		Merbecovirus	Hedgehog coronavirus 1	MK679660	30172	36.4	33.0	46.7	67.1	70.4	57.0	18.3	20.0
		Merbecovirus	Pipistrellus bat coronavirus HKU5	NC_009020	30482	29.9	29.5	45.0	65.2	67.6	57.6	17.2	20.3
		Merbecovirus	Tylonycteris bat coronavirus HKU4	NC_009019	30286	35.9	31.3	45.7	66.0	64.8	57.3	17.9	19.2
		Nobecovirus	Rousettus bat coronavirus HKU9	NC_009021	29114	29.0	28.7	48.1	66.2	63.9	58.7	20.3	21.3
		Sarbecovirus	SARS coronavirus Tor2	NC_004718	29751	29.2	25.2	45.3	65.7	63.9	55.4	18.4	20.6
		Sarbecovirus	SARS 2	NC_045512	29903	29.7	25.2	47.4	65.9	63.9	55.2	18.2	20.3
	Deltacoronavirus	*Andecovirus*	Wigeon coronavirus HKU20	NC_016995	26227	35.4	33.6	36.1	57.8	57.4	53.1	21.2	24.6
		Buldecovirus	Common moorhen coronavirus HKU21	NC_016996	26223	35.6	34.5	39.7	58.5	58.3	53.3	23.1	25.2
		Buldecovirus	Munia coronavirus HKU13	NC_011550	26552	29.8	36.2	39.3	59.9	57.4	59.2	22.7	24.7
		Buldecovirus	Bulbul coronavirus HKU11	NC_011547	26487	29.8	31.9	37.6	58.6	57.4	53.3	22.0	24.5
		Buldecovirus	Porcine coronavirus HKU15	NC_039208	25425	29.8	33.6	40.0	59.9	55.6	58.9	22.6	25.3
		Buldecovirus	White-eye coronavirus HKU16	NC_016991	26041	29.9	37.1	38.6	58.8	57.4	52.8	22.7	24.4
		Herdecovirus	Night-heron coronavirus HKU19	NC_016994	26077	30.0	32.8	36.8	59.9	54.6	53.1	23.0	23.9
	Gammacoronavirus	*Brangacovirus*	Canada goose coronavirus	NC_046965	28539	56.1	57.1	68.1	90.6	91.7	91.5	32.2	64.2
		Cegacovirus	Beluga Whale coronavirus SW1	NC_010646	31686	37.5	32.2	55.8	74.9	76.9	74.3	22.1	28.7
		Cegacovirus	Bottlenose dolphin coronavirus HKU22	KF793824	31769	37.7	32.2	56.2	75.3	76.9	74.3	22.2	28.7
		Igacovirus	Duck coronavirus isolate DK/GD/27	NC_048214	27754	**71.8** [a]	63.4	**89.7**	**99.1**	**99.1**	**99.2**	29.4	**83.5**
		Igacovirus	Infectious bronchitis virus	M95169	27608	64.4	**70.5**	66.3	91.6	91.7	91.5	**32.4**	67.3
		Igacovirus	Infectious bronchitis virus Ind-TN92–03	NC_048213	27464	63.2	64.3	67.9	87.1	89.8	82.5	32.0	67.1

[a]Bolded value is the highest percentage identity (%) of each genomic feature

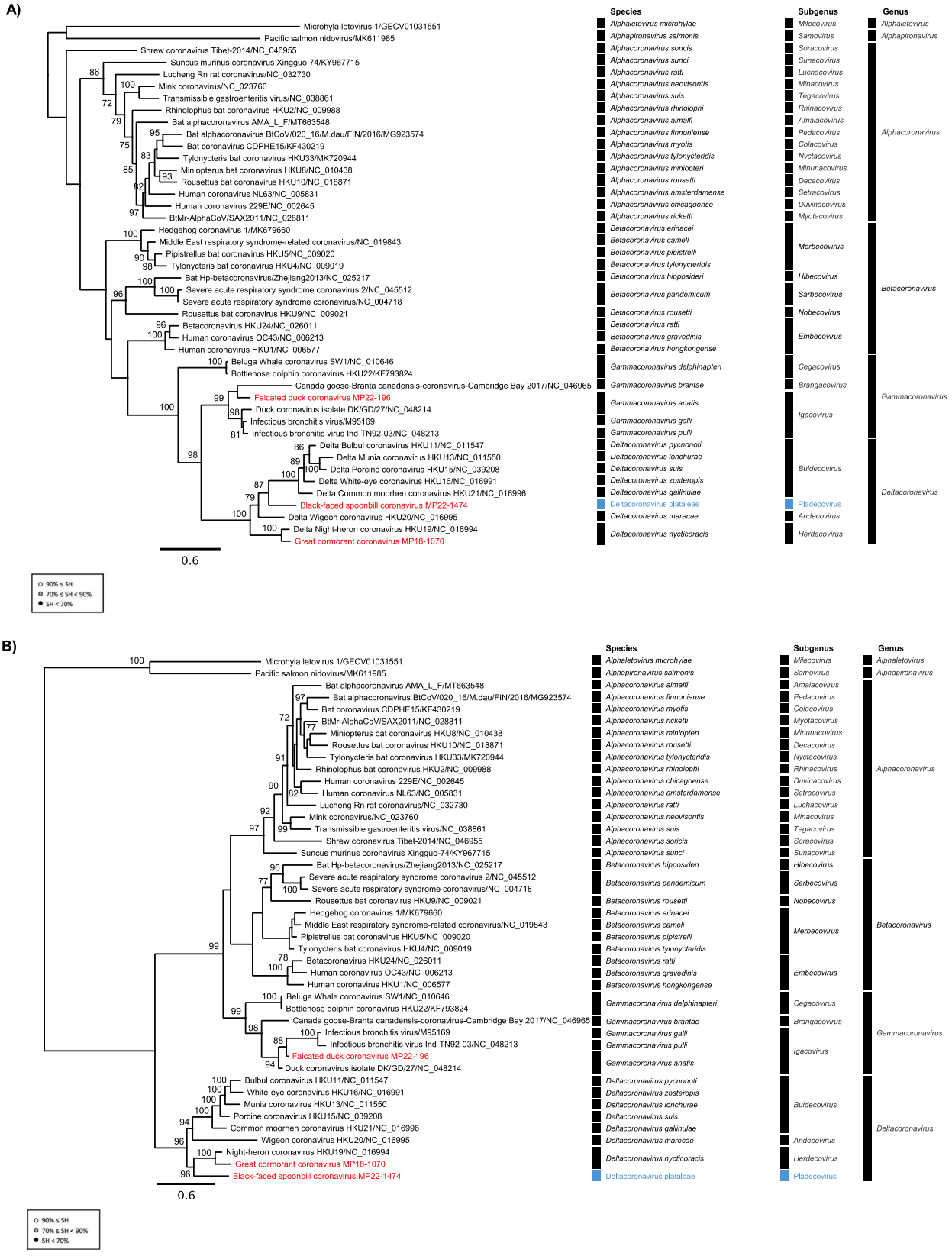

Figure 3. Phylogenetic trees of seven protein domains of coronaviruses. Depicted are trees for amino acid sequences of 3CLpro, NIRAN, RdRp, ZBD, HEL 1, S protein, and N protein of coronaviruses, including GCCoV MP18-1070, BSCoV MP22-1474, and FDCoV MP22-196. Each tree is rooted with *Microhyla letovirus 1* (GECV01031551) and *Pacific salmon nidovirus* (MK611985) as an outgroup. These trees are constructed by IQ-TREE v2.3.5 with the LG + F + G substitution model. Branch support is estimated using SH-aLRT with 1000 replicates. The support values of SH-aLRT are depicted with dots of varying shades at the internal nodes, according to details within the key box in the left corner. The branch values are bootstrap supports (%). Demarcations of species, subgenus, and genus are shown on the right side (see Fig. 4 for classification of the newly identified coronaviruses). For BSCoV MP22-1474, the subgenus is to be named Pladecovirus, and the species name is to be named Deltacoronavirus plataleae.

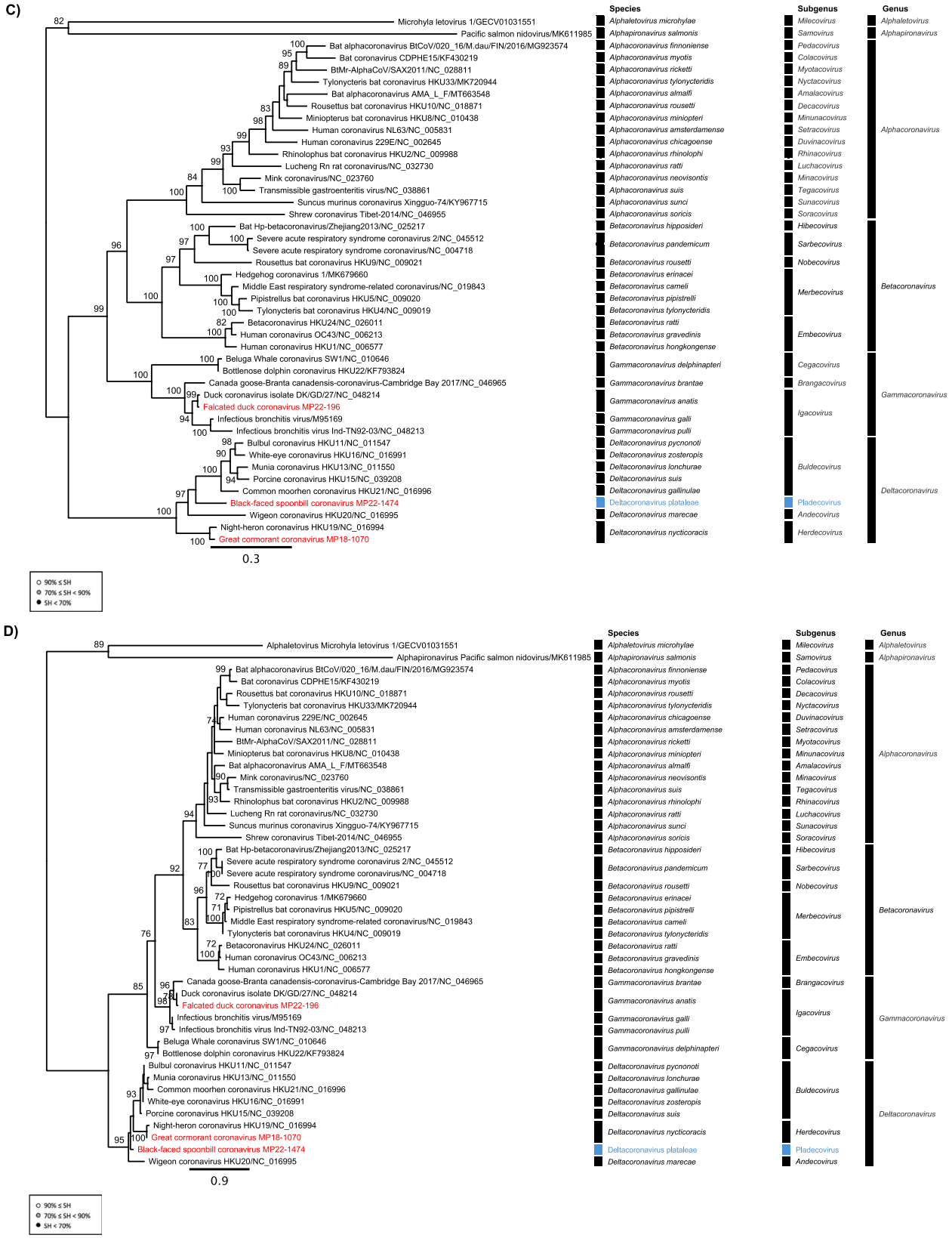

Figure 3. (Continued).

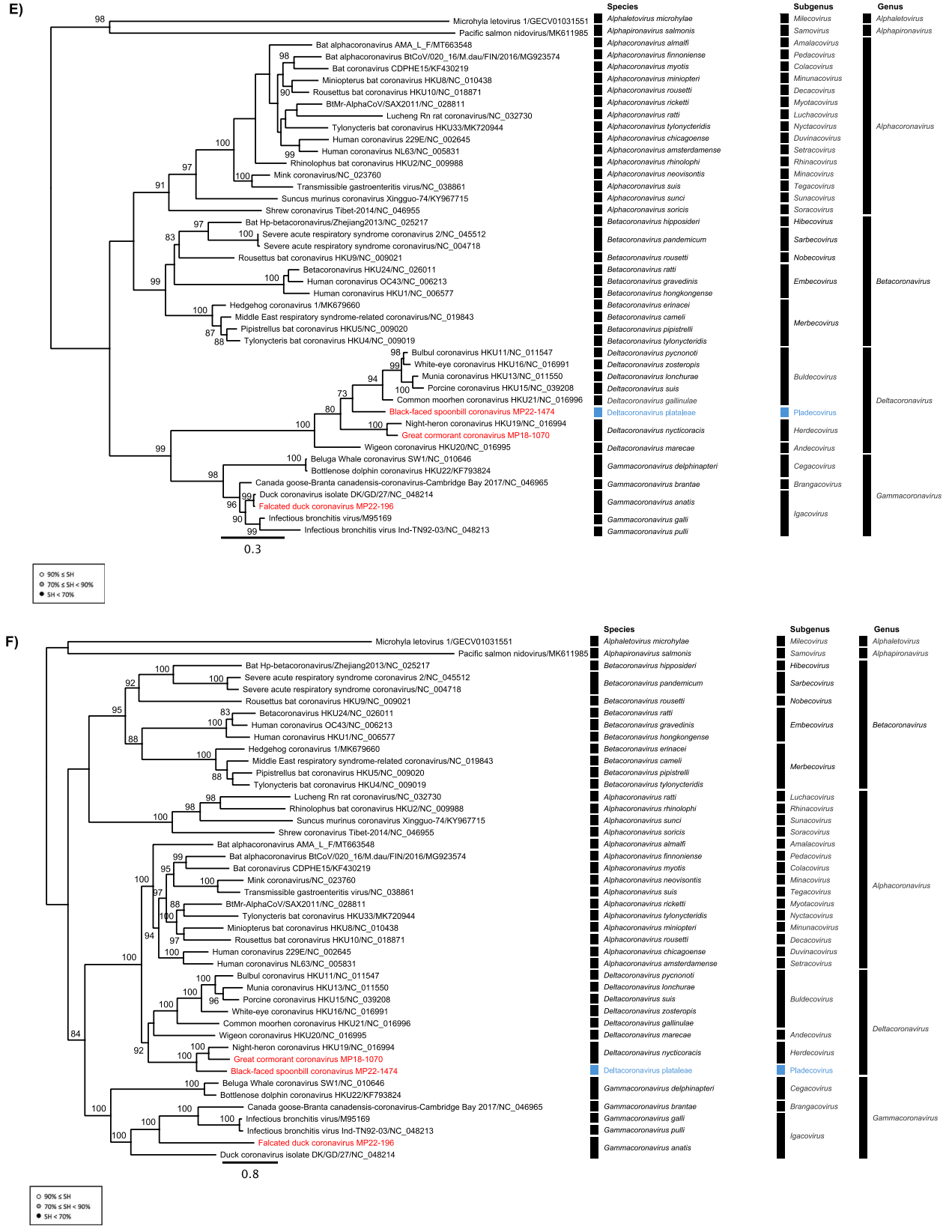

Figure 3. (Continued).

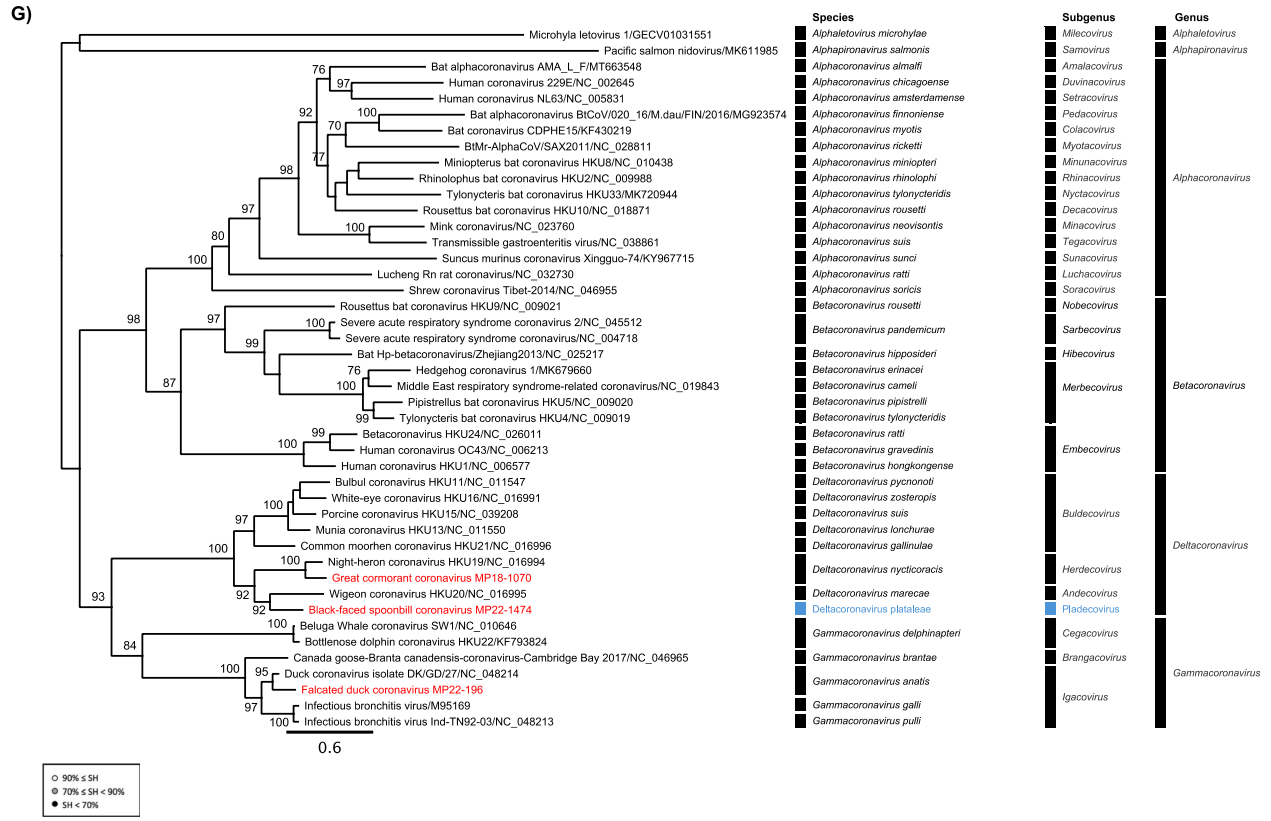

Figure 3. (Continued).

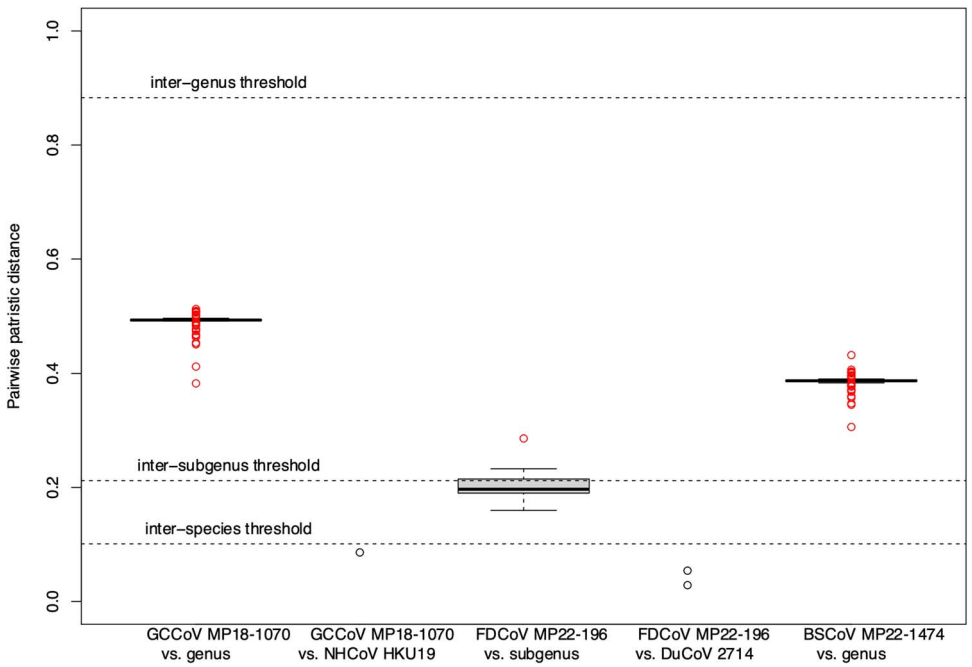

Figure 4. Hierarchical classification of three avian coronaviruses using DEmARC. Shown PPD of three avian coronaviruses with the most closely related coronaviruses of taxa up to the genus they belong to, using box-and-whisker plots. Outliers are indicated by hollow circular symbols, while solid symbols represent typical values. For GCCoV MP18-1070 and FDCoV MP22-196, their intraspecies PPDs are plotted separately from those with more distantly related viruses within either genus or subgenus, respectively. Intraspecies PPDs were derived from GCCoV MP18-1070 and night-heron coronavirus HKU19 (NHCoV HKU19, NC_016994), and also FDCoV MP22–196 and duck coronavirus isolate DK/GD/27/2014 (DuCoV 2714, NC_048214). Family-wide demarcation PPD thresholds for ranks of species, subgenus, and genus are shown as broken horizontal lines.

junction of the ORF1b and S genes (Fig. 5B), as well as with Avian coronavirus MK204411 at the junction between the M and N genes (Fig. 5D). These findings indicate potential recombination regions. The results confirmed that the region between the recombination breakpoints was the putative recombination region, consistent with the RDP5 results.

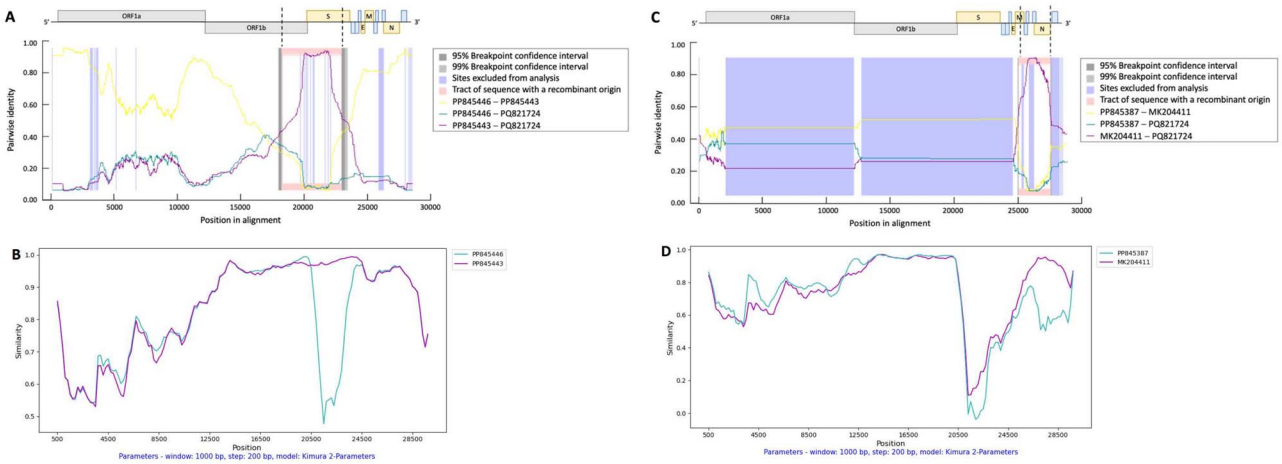

Figure 5. Putative recombination events of FDCoV MP22-196 are identified using the RDP method in RDP5 v5.61, and full-genome nucleotide sequence similarities are assessed using Simplot++ v1.3. FDCoV MP22-196 is used as the query sequence (PQ821724, recombinant) and analysed with the reference genome (Table S3) by RDP5 v5.61. Putative major and minor parent sequences are identified by RDP5 and used for the similarity plot analysis of FDCoV MP22-196 by Simplot++. The sites and lines are depicted in different colours according to the details within the key box on the right side. (A) RDP and (B) similarity plot derived from the major parent is bird gammacoronavirus AnasCN24 isolate AvAa-GammaCoV/SH21-SH145 (PP845446), and the minor parent is bird gammacoronavirus AnasCN24 isolate AvAz-GammaCoV/SH19-SH17B (PP845443). (C) RDP method and (D) similarity plot derived from the major parent is bird gammacoronavirus AnasCN24 isolate AvAz-GammaCoV/SH20-SH66 (PP845387), and the minor parent is avian coronavirus isolate MW18 (MK204411).

Coevolutionary analyses of avian coronaviruses and their hosts

We applied the global testing function of ParaFit and PACo on the RdRp gene of all positive samples for coronavirus and the COX1 gene (Table S5, $N = 24$) of the representative host. The obtained results strongly support the existence of substantial co-evolution between the coronavirus and the corresponding avian host (ParaFitGlobal = 165.7418, $P \leq .001$; m^2global value = 10.04692, $P \leq .0001$). Of the 207 host–virus pairs analysed, 101 were significantly linked, with P-values $\leq .05$ for ParaFit1 and ParaFit2 (Fig. 6). It is worth noting that the significant co-evolutionary relationships are predominantly limited to the subgenera *Igacovirus* and *Brangacovirus* in the genus *Gammacoronavirus* and derived from seven bird species (*S. clypeata, M. falcata, M. strepera, M. penelope, Anas acuta, A. crecca,* and *Branta canadensis*) which belong to the order of *Anseriformes* (Fig. 6). Also, we found gammacoronaviruses and deltacoronaviruses in many birds of *R. avosetta* in 2023–24, but global test results using ParaFit and PACo showed that the host–virus link did not have a significant coevolutionary fit.

Discussion

A high prevalence of coronaviruses in wild birds was discovered by our previous surveillance research in 2009–10 (Chu *et al.* 2011). In this study, we have applied a similar strategy to screen for coronaviruses in wild bird faecal samples collected from 2018 to 2024 at the same site, Mai Po Wetland. The positivity rates have remained under 10% from 2020 to 2024, which is relatively low compared to 15.0% in 2010 (Chu *et al.* 2011). This difference is statistically significant, with P-values of .0280 for 2020–21, .0001 for 2022–23, and <.0001 for 2023–24, as determined by a two-sample Z-test for proportions. The prevalence of gammacoronaviruses (64.4%) from 2020 to 2024 was significantly higher than that of deltacoronaviruses (35.6%), consistent with the findings from 2010 (P-value < .0001) (Chu *et al.* 2011). The majority of positive samples in this study came from the birds of the order *Anseriformes* and *Charadriiformes*. A literature review of the research on coronavirus-carrying wild birds worldwide found that 75.0% belong to the order *Anseriformes*, with the second largest group of 56.2% in the order *Charadriiformes* (Rahman *et al.* 2021). The results of our study fit this pattern. Interestingly, we identified one bird sample (*S. clypeata*) positive for both gammacoronavirus and a low-pathogenic influenza A virus H7N7. This observation highlights the potential for wild birds to be coinfected by multiple pathogens that may affect virus transmission and evolution. Further investigation into such co-infections is warranted, especially regarding their potential impact on viral ecology and epidemiology. This study examines coronaviruses in wild birds in Hong Kong over 12 years. While our results indicate a difference in the positivity rates between gammacoronaviruses and deltacoronaviruses, the exact reasons for this difference are unclear. Factors of technical origin, such as primer compatibility with new viral groups, or biological nature, including variations in mucosal immunity, population size, and migration patterns, as well as climate change, may influence the derived rates. We have excluded the possibility that such differences are due to biases created by our primer sets (data not shown). Nonetheless, considering the evolution of coronaviruses, we should continue to improve and refine the primer sets used in future coronavirus surveillance to enhance experimental accuracy. Additionally, further research that accounts for immune responses, group sizes, migratory routes, and seasonal climate effects can improve our understanding of coronavirus ecology in wild aquatic birds.

In this study, two deltacoronaviruses were detected in the faecal samples of the Great Cormorant and Black-faced Spoonbill, respectively, and complete genome sequences were successfully determined. A nearly complete genome sequence was also derived for a gammacoronavirus detected in the falcated duck. The genome sizes of these three coronaviruses are similar to other deltacoronaviruses (25400–26 689 nt) and gammacoronaviruses (27600–31 759 nt) (Woo *et al.* 2014, Papineau *et al.* 2019, Vlasova *et al.* 2020). Also, the genome architecture of these three viruses is similar to other coronaviruses, with the main part covered by two overlapping ORFs, which can be translated into polyproteins

Figure 6. Tanglegram of cophylogenetic relationships between wild bird hosts and coronaviruses. Two trees are constructed by IQ-TREE v2.3.5 with the GTR + I + G substitution model. The left is the host evolutionary tree (COX1 gene, 670 bp), and the branches in the same colour represent the same order. The host tree is rooted with *Apteryx mantelli* North Island brown kiwi (EU525317) as an outgroup. The right is the phylogenetic tree of the viral RdRp gene (440 bp), and the branches in the same colour represent the same subgenus. The virus tree is rooted with human coronavirus 229E (NC_002645). The cophylogenetic relationships are performed by the ParaFit and PACo functions in the R program. Solid lines denoted significant coevolution links between coronaviruses and their hosts, and dotted lines denoted nonsignificant links.

1a and 1ab through the ribosomal frameshift mechanism (Agranovsky 2021, Kelly *et al.* 2021). We also tentatively mapped nsp2 to nsp16 and their corresponding cleavage sites in the ORF1ab-encoded polyproteins of these three viruses (Table 4). Among them, the amino acids upstream of the putative cleavage site of BSCoV MP22-1474 (nsp3/nsp4: KG, nsp8/nsp9: CQ, and nsp13/nsp14: TQ) differ from those observed in other deltacoronaviruses (Lau *et al.* 2018). Amino acid replacements in these P2-P1 sites can affect viral protease recognition and cleavage efficiency (Kanjanahaluethai *et al.* 2003). The putative TRS elements are separated from the nearest start codon of the genome ORFs (S gene, partial NS6 and 7) by a distance in the range of (110–201 nt) in the described three coronaviruses (Table 3). According to the literature, this relatively large distance was only described for the S gene of deltacoronaviruses (Woo *et al.* 2012, Lau *et al.* 2018). The position of TRS can influence the efficiency and accuracy of transcription. If the TRS is too far from the start codon, it might affect the binding of transcription factors and the overall transcriptional machinery, potentially leading to less efficient transcription initiation (Zhang *et al.* 2022). The potential impact of the TRS location in these three viruses needs further research.

We used seven conserved domains (3CLpro, NiRAN, RdRp, ZBD, HEL1, S gene, and N gene) for distance analysis using different metrics and programs. The phylogenetic tree and amino acid identity indicate that GCCoV MP18-1070 and BSCoV MP22-1474 belong to the genus *Deltacoronavirus*, while FDCoV MP22-196 belongs to the genus *Gammacoronavirus*. According to our DEmARC-based results, BSCoV MP22-1474 most likely constitutes a new putative subgenus. Although the results show that GCCoV MP18-1070 is closely related to NHCoV HKU19 (NC_016994) and FDCoV MP22-196 is closely associated with DuCoV 2 714 (NC_048214), the nucleotide identity of the whole genome is only 73.5% and 71.8%. Also, the amino acid identity of the S gene is only 66.3% and 29.4% (Table 5). This suggests that these viruses are considerably different from previously identified viruses of these species, warranting further research.

According to the ICTV practice, subgenus names usually consist of a unique part derived from the host or virus name, followed by the first one or two letters of the genus name (Woo *et al.* 2023). Therefore, the new subgenus is provisionally named '*Pladecovirus*', 'Pla' is taken from the host *Platalea minor*, and 'de' is the genus Delta and ends with 'covirus'. In contrast, GCCoV MP18-1070 and FDCoV MP22-196 were found to belong to two known species *Deltacoronavirus nycticoracis*, and *Gammacoronavirus anatis*, respectively, but the PPD results showed that they were very different from the known representatives of these species (Fig. 4). In particular, the distance between GCCoV MP18-1070 and its closest sibling is very close to the interspecies demarcation PPD threshold. These results reveal considerable intraspecies variability and contribute to the appreciation among and within avian coronavirus species.

Recombination phenomena occur frequently in coronaviruses, which may facilitate interspecies transmission, adaptation to new hosts, and affect virulence (Lau *et al.* 2018, Wang *et al.* 2021, Li *et al.* 2022, Guo *et al.* 2024). We used RDP5 software to screen potential recombination events in these viral genomes and further confirmed them using SimPlot++. Two putative recombination events were observed in FDCoV MP22-196, located from the downstream of ORF1b to the S gene and from the downstream of the M gene to the N gene (Fig. 5, Table S4). Genetic recombination events frequently occur in the S gene of coronaviruses (Flores-Alanis *et al.* 2020, Antas *et al.* 2021, Francois *et al.* 2023). The literature mentioned that IBV involving the S gene recombination may produce new strains or serotypes that possibly enhance pathogenicity in host species (Feng *et al.* 2018). In contrast, the proteins encoded by the M and N genes are highly conserved and have a lower likelihood of recombination events (Fehr and Perlman 2015). However, previous literature also suggests that diversity in M and N genes may result in lower efficacy of IBV cross-protection (Feng *et al.* 2018). Although no recombination events were identified for the two deltacoronaviruses in this study, many studies have provided evidence for genetic recombination among avian deltacoronaviruses. In a Dubai study, four novel deltacoronaviruses involving multiple recombination events were detected in different bird species, demonstrating the potential for interspecies transmission in birds from the Middle East (Lau *et al.* 2018). Evidence of at least two recombination events was observed in the S gene of common magpies coronavirus HNU1 and HNU2 (Wang *et al.* 2022). In 2024, a study found that the deltacoronavirus that infects black-headed gulls has recombination phenomena, indicating a greater structural similarity between the receptor-binding domain of BHG-QH-2021 and alphacoronavirus human coronavirus 229E, which may be a key factor in the ability to infect humans (Tian *et al.* 2024). The potential role of recombination in the evolution of viruses in the genus *Deltacoronavirus* deserves exploration in future studies.

We conducted a global fitting analysis of phylogenies to understand the co-evolution between avian hosts and coronaviruses. The results demonstrated that gammacoronaviruses are host-specific, supporting our 2011 study results (Chu *et al.* 2011). This observation highlights the virus–host linkage between the genus *Gammacoronavirus* and birds of the order *Anseriformes* (Fig. 6). These waterfowl can fly long distances and are common worldwide due to their high survival rates. Prior research indicates that *Anseriformes* avian species are the main carriers of coronaviruses (accounting for 75%), and it was also found that transmission between these waterfowl populations in different countries or regions is very common through connected migration routes (Hepojoki *et al.* 2017, Rahman *et al.* 2021). In addition, we observed that the prevalence of gammacoronaviruses and deltacoronaviruses in the birds of the order *Charadriiformes* has increased remarkably from 7.6% in 2022–23 to 40.2% in 2023–24, indicating that the circulation of these viruses may have changed (Table 2). Although deltacoronaviruses are not known to have a specific host species association, our research and others have discovered these viruses mainly in *Charadriiformes* (Fig. 6) (Wille and Holmes 2020, Domanska-Blicharz *et al.* 2023). Different studies have also documented a considerable diversity of deltacoronaviruses and evidence of transmission in multiple host species (Vlasova *et al.* 2020, Wille and Holmes 2020). The widespread spread of deltacoronaviruses and their potential for interspecific transmission are notable. Regular monitoring is necessary to understand the evolution of deltacoronaviruses in aquatic wild birds.

Aquatic birds are considered a natural reservoir of a wide range of coronaviruses. However, our understanding of avian coronavirus circulation is limited and challenging to study due to diverse geographical locations and numerous bird species. In this study of coronavirus surveillance among wild birds in Hong Kong, we recovered a full-length genome sequence of a coronavirus that prototypes a new subgenus of the *Deltacoronavirus* (GCCoV MP18-1070) and two other highly divergent coronaviruses (BSCoV MP22-1474 and FDCoV MP22-196). Further analysis revealed the genome architecture and protein domain organization, phylogenetic relationships, and putative recombination events of these three

coronaviruses, as well as the cophylogenetic linkage between hosts and viruses, demonstrating the genetic diversity of coronaviruses in aquatic birds and their potential for global spread. Extensive and long-term surveillance of gammacoronaviruses and deltacoronaviruses is prudent. In the future, it is necessary to conduct in-depth research on the evolution and host adaptability of coronaviruses in different geographical locations to prevent the future impact of these viruses on ecosystems or human society.

Acknowledgements

We gratefully acknowledge the Agriculture, Fisheries and Conservation Department and World Wildlife Fund Hong Kong for preserving this precious wetland and providing the opportunity to obtain the specimens. We acknowledge the technical support provided by colleagues in the surveillance team from the School of Public Health of the University of Hong Kong.

Supplementary data

Supplementary data are available at *VEVOLU Journal* online.

Conflict of interest

None declared.

Funding

The work described in this paper was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (U01AI151810), Theme-based Research Scheme (T11-705/21-N) of the Research Grants Council of the Hong Kong Special Administrative Region, China.

Data availability

Two deltacoronaviruses and one gammacoronavirus genome sequence have been deposited in the NCBI GenBank with the accession numbers PQ821722-PQ821724.

Ethics declaration

Not applicable.

References

Agranovsky AA. Structure and expression of large (+)Rna genomes of viruses of higher eukaryotes. *Biochemistry (Mosc)* 2021;**86**:248–56. https://doi.org/10.1134/S0006297921030020

Akiva E, Brown S, Almonacid DE *et al.* The structure-function linkage database. *Nucleic Acids Res* 2014;**42**:D521–30. https://doi.org/10.1093/nar/gkt1130

Alluwaimi AM, Alshubaith IH, Al-Ali AM *et al.* The coronaviruses of animals and birds: their zoonosis, vaccines, and models for Sars-Cov and Sars-Cov2. *Front Vet Sci* 2020;**7**:582287. https://doi.org/10.3389/fvets.2020.582287

Antas M, Olech M, Szczotka-Bochniarz A. Molecular characterization of porcine epidemic diarrhoea virus (Pedv) in Poland reveals the presence of swine enteric coronavirus (Secov) sequence in S gene. *PLoS One* 2021;**16**:E0258318. https://doi.org/10.1371/journal.pone.0258318

Attwood TK, Coletta A, Muirhead G *et al.* The Prints database: a fine-grained protein sequence annotation and analysis resource–its status In 2012. *Database (Oxford)* 2012;**2012**:Bas019.

Balbuena JA, Miguez-Lozano R, Blasco-Costa I. Paco: a novel Procrustes application to Cophylogenetic analysis. *PLoS One* 2013;**8**:E61048. https://doi.org/10.1371/journal.pone.0061048

Blum M, Andreeva A, Florentino LC *et al.* Interpro: the protein sequence classification resource In 2025. *Nucleic Acids Res* 2025;**53**:D444–56. https://doi.org/10.1093/nar/gkae1082

Boni MF, Posada D, Feldman MW. An exact nonparametric method for inferring mosaic structure In sequence triplets. *Genetics* 2007;**176**:1035–47. https://doi.org/10.1534/genetics.106.068874

Brinton MA, Gulyaeva AA, Balasuriya UBR *et al.* Ictv virus taxonomy profile: Arteriviridae 2021. *J Gen Virol* 2021;**102**:001632. https://doi.org/10.1099/jgv.0.001632

Camacho C, Coulouris G, Avagyan V *et al.* Blast+: architecture and applications. *BMC Bioinformatics* 2009;**10**:421. https://doi.org/10.1186/1471-2105-10-421

Cavanagh D. Coronaviruses In poultry and other birds. *Avian Pathol* 2005;**34**:439–48. https://doi.org/10.1080/03079450500367682

Chen S, Zhou Y, Chen Y *et al.* Fastp: an ultra-fast all-In-one Fastq Preprocessor. *Bioinformatics* 2018;**34**:I884–90. https://doi.org/10.1093/bioinformatics/bty560

Chen H, Zhu Z, Qiu Y *et al.* Prediction of coronavirus 3c-like protease cleavage sites using machine-learning algorithms. *Virol Sin* 2022;**37**:437–44. https://doi.org/10.1016/j.virs.2022.04.006

Cheung PP, Leung YH, Chow CK *et al.* Identifying the species-origin of faecal droppings used for avian influenza virus surveillance In wild-birds. *J Clin Virol* 2009;**46**:90–3. https://doi.org/10.1016/j.jcv.2009.06.016

Chu DK, Leung CY, Gilbert M *et al.* Avian coronavirus In wild aquatic birds. *J Virol* 2011;**85**:12815–20. https://doi.org/10.1128/JVI.05838-11

Coronaviridae Study Group Of The International Committee On Taxonomy Of, V. The species severe acute respiratory syndrome-related coronavirus: classifying 2019-Ncov and naming it Sars-Cov-2. *Nat Microbiol* 2020;**5**:536–44. https://doi.org/10.1038/s41564-020-0695-z

Domanska-Blicharz K, Kuczkowski M, Sajewicz-Krukowska J. Whole genome characterisation of quail Deltacoronavirus detected in Poland. *Virus Genes* 2019;**55**:243–7. https://doi.org/10.1007/s11262-019-01639-1

Domanska-Blicharz K, Milek-Krupa J, Pikula A. Gulls As a host for both gamma and Deltacoronaviruses. *Sci Rep* 2023;**13**:15104. https://doi.org/10.1038/s41598-023-42241-8

Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol* 2015;**1282**:1–23. https://doi.org/10.1007/978-1-4939-2438-7_1

Feng KY, Chen T, Zhang X *et al.* Molecular characteristic and pathogenicity analysis of a virulent recombinant Avain infectious bronchitis virus isolated In China. *Poult Sci* 2018;**97**:3519–31. https://doi.org/10.3382/ps/pey237

Flores-Alanis A, Sandner-Miranda L, Delgado G *et al.* The receptor binding domain of Sars-Cov-2 spike protein is the result of an ancestral recombination between the bat-Cov Ratg13 and the pangolin-Cov Mp789. *BMC Res Notes* 2020;**13**:398. https://doi.org/10.1186/s13104-020-05242-8

Francois S, Nazki S, Vickers SH *et al.* Genetic diversity, recombination and cross-species transmission of a Waterbird Gammacoronavirus In the wild. *J Gen Virol* 2023;**104**:001883. https://doi.org/10.1099/jgv.0.001883

Gibbs MJ, Armstrong JS, Gibbs AJ. Sister-scanning: a Monte Carlo procedure for assessing signals In recombinant

sequences. *Bioinformatics* 2000;**16**:573–82. https://doi.org/10.1093/bioinformatics/16.7.573

Gorbalenya AE, Lieutaud P, Harris MR *et al.* Practical application of bioinformatics by the multidisciplinary vizier consortium. *Antivir Res* 2010;**87**:95–110. https://doi.org/10.1016/j.antiviral.2010.02.005

Gu X, Fu YX, Li WH. Maximum likelihood estimation of the heterogeneity of substitution rate among nucleotide sites. *Mol Biol Evol* 1995;**12**:546–57.

Guindon S, Dufayard JF, Lefort V *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of Phyml 3.0. *Syst Biol* 2010;**59**:307–21. https://doi.org/10.1093/sysbio/syq010

Guo J, He J, Liang Z *et al.* Birds As reservoirs: Unraveling the global spread of gamma- and Deltacoronaviruses. *MBio* 2024;**15**:E0232424. https://doi.org/10.1128/mbio.02324-24

Hafez HM, Attia YA. Challenges to the poultry industry: current perspectives and strategic future after the Covid-19 outbreak. *Front Vet Sci* 2020;**7**:516. https://doi.org/10.3389/fvets.2020.00516

Haft DH, Badretdin A, Coulouris G *et al.* Refseq and the prokaryotic genome annotation pipeline In the age of metagenomes. *Nucleic Acids Res* 2024;**52**:D762–9. https://doi.org/10.1093/nar/gkad988

Hepojoki S, Lindh E, Vapalahti O *et al.* Prevalence and genetic diversity of coronaviruses In wild birds, Finland. *Infect Ecol Epidemiol* 2017;**7**:1408360.

Hoang DT, Chernomor O, Von Haeseler A *et al.* UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol* 2018;**35**:518–22. https://doi.org/10.1093/molbev/msx281

Kalyaanamoorthy S, Minh BQ, Wong TKF *et al.* Modelfinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* 2017;**14**:587–9. https://doi.org/10.1038/nmeth.4285

Kanjanahaluethai A, Jukneliene D, Baker SC. Identification of the murine coronavirus Mp1 cleavage site recognized by papain-like proteinase 2. *J Virol* 2003;**77**:7376–82. https://doi.org/10.1128/JVI.77.13.7376-7382.2003

Katoh K, Standley DM. Mafft multiple sequence alignment software version 7: improvements In performance and usability. *Mol Biol Evol* 2013;**30**:772–80. https://doi.org/10.1093/molbev/mst010

Kelly JA, Woodside MT, Dinman JD. Programmed −1 ribosomal frameshifting In coronaviruses: a therapeutic target. *Virology* 2021;**554**:75–82. https://doi.org/10.1016/j.virol.2020.12.010

Lau SKP, Wong EYM, Tsang CC *et al.* Discovery and sequence analysis of four Deltacoronaviruses from birds In the Middle East reveal interspecies jumping with recombination As a potential mechanism for avian-to-avian and avian-to-mammalian transmission. *J Virol* 2018;**92**:e00265-18. https://doi.org/10.1128/JVI.00265-18

Lauber C, Gorbalenya AE. Partitioning the genetic diversity of a virus family: approach and evaluation through a case study of picornaviruses. *J Virol* 2012a;**86**:3890–904. https://doi.org/10.1128/JVI.07173-11

Lauber C, Gorbalenya AE. Toward genetics-based virus taxonomy: comparative analysis of a genetics-based classification and the taxonomy of picornaviruses. *J Virol* 2012b;**86**:3905–15. https://doi.org/10.1128/JVI.07174-11

Lauber C, Ziebuhr J, Junglen S *et al.* Mesoniviridae: a proposed new family In the order Nidovirales formed by a single species of mosquito-borne viruses. *Arch Virol* 2012;**157**:1623–8. https://doi.org/10.1007/s00705-012-1295-x

Lauber C, Zhang X, Vaas J *et al.* Deep mining of the sequence read archive reveals major genetic innovations In coronaviruses and other Nidoviruses of aquatic vertebrates. *PLoS Pathog* 2024;**20**:E1012163. https://doi.org/10.1371/journal.ppat.1012163

Lednicky JA, Tagliamonte MS, White SK *et al.* Independent infections of porcine Deltacoronavirus among Haitian children. *Nature* 2021;**600**:133–7. https://doi.org/10.1038/s41586-021-04111-z

Legendre P, Desdevises Y, Bazin E. A statistical test for host-parasite coevolution. *Syst Biol* 2002;**51**:217–34. https://doi.org/10.1080/10635150252899734

Letunic I, Khedkar S, Bork P. Smart: recent updates, new developments and status In 2020. *Nucleic Acids Res* 2021;**49**:D458–60. https://doi.org/10.1093/nar/gkaa937

Leung YH, Zhang LJ, Chow CK *et al.* Poultry drinking water used for avian influenza surveillance. *Emerg Infect Dis* 2007;**13**:1380–2. https://doi.org/10.3201/eid1309.070517

Li Q, Shah T, Wang B *et al.* Cross-species transmission, evolution and zoonotic potential of coronaviruses. *Front Cell Infect Microbiol* 2022;**12**:1081370.

Martin DP, Posada D, Crandall KA *et al.* A modified Bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. *AIDS Res Hum Retrovir* 2005;**21**:98–102. https://doi.org/10.1089/aid.2005.21.98

Martin DP, Varsani A, Roumagnac P *et al.* Rdp5: a computer program for Analyzing recombination In, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol* 2021;**7**:veaa087. https://doi.org/10.1093/ve/veaa087

Minh BQ, Schmidt HA, Chernomor O *et al.* Iq-tree 2: new models and efficient methods for phylogenetic inference In the genomic era. *Mol Biol Evol* 2020;**37**:1530–4. https://doi.org/10.1093/molbev/msaa015

Mistry J, Chuguransky S, Williams L *et al.* Pfam: the protein families database In 2021. *Nucleic Acids Res* 2021;**49**:D412–9. https://doi.org/10.1093/nar/gkaa913

Nguyen LT, Schmidt HA, Von Haeseler A *et al.* Iq-tree: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015;**32**:268–74. https://doi.org/10.1093/molbev/msu300

Nikolskaya AN, Arighi CN, Huang H *et al.* Pirsf family classification system for protein functional and evolutionary analysis. *Evol Bioinformatics Online* 2007;**2**:197–209.

Nurk S, Meleshko D, Korobeynikov A *et al.* Metaspades: a new versatile metagenomic assembler. *Genome Res* 2017;**27**:824–34. https://doi.org/10.1101/gr.213959.116

Oates ME, Stahlhacke J, Vavoulis DV *et al.* The superfamily 1.75 database In 2014: a doubling of data. *Nucleic Acids Res* 2015;**43**:D227–33. https://doi.org/10.1093/nar/gku1041

Padidam M, Sawyer S, Fauquet CM. Possible emergence of new Geminiviruses by frequent recombination. *Virology* 1999;**265**:218–25. https://doi.org/10.1006/viro.1999.0056

Papineau A, Berhane Y, Wylie TN *et al.* Genome organization of Canada goose coronavirus, a novel species identified In a mass die-off of Canada geese. *Sci Rep* 2019;**9**:5954. https://doi.org/10.1038/s41598-019-42355-y

Paradis E, Claude J, Strimmer K. Ape: analyses of Phylogenetics and evolution In R language. *Bioinformatics* 2004;**20**:289–90. https://doi.org/10.1093/bioinformatics/btg412

Pedruzzi I, Rivoire C, Auchincloss AH *et al.* Hamap In 2015: updates to the protein family classification and annotation system. *Nucleic Acids Res* 2015;**43**:D1064–70. https://doi.org/10.1093/nar/gku1002

Posada D, Crandall KA. Evaluation of methods for detecting recombination from Dna sequences: computer simulations. *Proc Natl Acad Sci USA* 2001;**98**:13757–62. https://doi.org/10.1073/pnas.241370698

Price MN, Dehal PS, Arkin AP. FastTree 2 – approximately maximum-likelihood trees for large alignments. *PLoS One* 2010;**5**:E9490. https://doi.org/10.1371/journal.pone.0009490

Quinlan AR, Hall IM. Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;**26**:841–2. https://doi.org/10.1093/bioinformatics/btq033

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation For Statistical Computing, 2024.

Rahman MM, Talukder A, Chowdhury MMH *et al.* Coronaviruses In wild birds - a potential and suitable vector for global distribution. *Vet Med Sci* 2021;**7**:264–72. https://doi.org/10.1002/vms3.360

Revell LJ. Phytools 2.0: an updated R ecosystem for phylogenetic comparative methods (and other things). *Peerj* 2024;**12**:E16505. https://doi.org/10.7717/peerj.16505

Samson S, Lord E, Makarenkov V. Simplot++: a python application for representing sequence similarity and detecting recombination. *Bioinformatics* 2022;**38**:3118–20. https://doi.org/10.1093/bioinformatics/btac287

Sigrist CJA, De Castro E, Cerutti L *et al.* New and continuing developments at Prosite. *Nucleic Acids Res* 2013;**41**:E344–7.

Sillitoe I, Bordin N, Dawson N *et al.* Cath: increased structural coverage of functional space. *Nucleic Acids Res* 2021;**49**:D266–73. https://doi.org/10.1093/nar/gkaa1079

Smith JS. Reusable software concepts applied to the development of Fms control software. *Int J Comput Integr Manuf* 1992;**5**:182–96. https://doi.org/10.1080/09511929208944527

Thomas PD, Ebert D, Muruganujan A *et al.* Panther: making genome-scale phylogenetics accessible to all. *Protein Sci* 2022;**31**:8–22. https://doi.org/10.1002/pro.4218

Tian Y, Yu T, Wang J *et al.* Genetic characterization of the first Deltacoronavirus from wild birds around Qinghai Lake. *Front Microbiol* 2024;**15**:1423367. https://doi.org/10.3389/fmicb.2024.1423367

Vlasova AN, Kenney SP, Jung K *et al.* Deltacoronavirus evolution and transmission: current scenario and evolutionary perspectives. *Front Vet Sci* 2020;**7**:626785.

Walker BJ, Abeel T, Shea T *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 2014;**9**:E112963. https://doi.org/10.1371/journal.pone.0112963

Wang Y, Zeng J, Zhang C *et al.* New framework for recombination and adaptive evolution analysis with application to the novel coronavirus Sars-Cov-2. *Brief Bioinform* 2021;**22**:bbab107. https://doi.org/10.1093/bib/bbab107

Wang Q, Zhou ZJ, You Z *et al.* Epidemiology and evolution of novel Deltacoronaviruses in birds in Central China. *Transbound Emerg Dis* 2022;**69**:632–44. https://doi.org/10.1111/tbed.14029

Wang J, Chitsaz F, Derbyshire MK *et al.* The conserved domain database In 2023. *Nucleic Acids Res* 2023;**51**:D384–8. https://doi.org/10.1093/nar/gkac1096

Weiss SR, Navas-Martin S. Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol Mol Biol Rev* 2005;**69**:635–64. https://doi.org/10.1128/MMBR.69.4.635-664.2005

Wille M, Holmes EC. Wild birds As reservoirs for diverse and abundant gamma- and Deltacoronaviruses. *FEMS Microbiol Rev* 2020;**44**:631–44. https://doi.org/10.1093/femsre/fuaa026

Woo PC, Lau SK, Huang Y *et al.* Coronavirus diversity, phylogeny and interspecies jumping. *Exp Biol Med (Maywood)* 2009;**234**:1117–27. https://doi.org/10.3181/0903-MR-94

Woo PC, Lau SK, Lam CS *et al.* 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* 2012;**86**:3995–4008. https://doi.org/10.1128/JVI.06540-11

Woo PC, Lau SK, Lam CS *et al.* Discovery of a novel bottlenose dolphin coronavirus reveals a distinct species of marine mammal coronavirus In Gammacoronavirus. *J Virol* 2014;**88**:1318–31. https://doi.org/10.1128/JVI.02351-13

Woo PCY, De Groot RJ, Haagmans B *et al.* Ictv virus taxonomy profile: Coronaviridae 2023. *J Gen Virol* 2023;**104**:001843. https://doi.org/10.1099/jgv.0.001843

Zhang C, Sashittal P, Xiang M *et al.* Accurate identification of transcription regulatory sequences and genes in coronaviruses. *Mol Biol Evol* 2022;**39**:msac133. https://doi.org/10.1093/molbev/msac133

Zhou Z, Qiu Y, Ge X. The taxonomy, host range and pathogenicity of coronaviruses and other viruses in the Nidovirales order. *Anim Dis* 2021;**1**:5. https://doi.org/10.1186/s44149-021-00005-9

Virus Evolution, 2025, 1–20
https://doi.org/10.1093/ve/veaf049
Research Article