Discovery of a Novel Bottlenose Dolphin Coronavirus Reveals a Distinct Species of Marine Mammal Coronavirus in *Gammacoronavirus*

Patrick C. Y. Woo,^{a,b,c,d} Susanna K. P. Lau,^{a,b,c,d} Carol S. F. Lam,^b Alan K. L. Tsang,^b Suk-Wai Hui,^e Rachel Y. Y. Fan,^b Paolo Martelli,^e Kwok-Yung Yuen^{a,b,c,d}

State Key Laboratory of Emerging Infectious Diseases,^a Department of Microbiology,^b Research Centre of Infection and Immunology,^c and Carol Yu Centre for Infection,^d The University of Hong Kong, and Ocean Park Corporation,^e Hong Kong

While gammacoronaviruses mainly comprise infectious bronchitis virus (IBV) and its closely related bird coronaviruses (CoVs), the only mammalian gammacoronavirus was discovered from a white beluga whale (beluga whale CoV [BWCoV] SW1) in 2008. In this study, we discovered a novel gammacoronavirus from fecal samples from three Indo-Pacific bottlenose dolphins (*Tursiops aduncus*), which we named bottlenose dolphin CoV (BdCoV) HKU22. All the three BdCoV HKU22-positive samples were collected on the same date, suggesting a cluster of infection, with viral loads of 1×10^3 to 1×10^5 copies per ml. Clearance of virus was associated with a specific antibody response against the nucleocapsid of BdCoV HKU22. Complete genome sequencing and comparative genome analysis showed that BdCoV HKU22 and BWCoV SW1 have similar genome characteristics and structures. Their genome size is about 32,000 nucleotides, the largest among all CoVs, as a result of multiple unique open reading frames (NS5a, NS5b, NS5c, NS6, NS7, NS8, NS9, and NS10) between their membrane (M) and nucleocapsid (N) protein genes. Although comparative genome analysis showed that BdCoV HKU22 and BWCoV SW1 should belong to the same species, a major difference was observed in the proteins encoded by their spike (S) genes, which showed only 74.3 to 74.7% amino acid identities. The high ratios of the number of synonymous substitutions per synonymous site (K_s) to the number of nonsynonymous substitutions per nonsynonymous site (K_a) in multiple regions of the genome, especially the S gene (K_a/K_s ratio, 2.5), indicated that BdCoV HKU22 may be evolving rapidly, supporting a recent transmission event to the bottlenose dolphins. We propose a distinct species, *Cetacean coronavirus*, in *Gammacoronavirus*, to include BdCoV HKU22 and BWCoV SW1, whereas IBV and its closely related bird CoVs represent another species, *Avian coronavirus*, in *Gammacoronavirus*.

Coronaviruses (CoVs) are found in a wide variety of animals, in which they can cause respiratory, enteric, hepatic, and neurological diseases of various severities. On the basis of genotypic and serological characterization, CoVs were traditionally divided into three distinct groups (1–3). Recently, the Coronavirus Study Group of the International Committee for the Taxonomy of Viruses has proposed three genera, *Alphacoronavirus*, *Betacoronavirus*, and *Gammacoronavirus*, to replace the traditional group 1, 2, and 3 CoVs, respectively. As a result of the unique mechanism of viral replication, CoVs have a high frequency of recombination (2). Their tendency for recombination and high mutation rates may allow them to adapt to new hosts and ecological niches (4, 5).

The severe acute respiratory syndrome (SARS) epidemic, the discovery of SARS coronavirus (SARS-CoV), and the identification of SARS-CoV-like viruses from Himalayan palm civets and a raccoon dog from wild live markets in China in 2003 have boosted interest in the discovery of novel CoVs in both humans and animals (6–11). A novel human CoV (HCoV) of the genus *Alphacoronavirus*, HCoV-NL63, was reported in 2004 (12, 13). In 2005, we also described the discovery, complete genome sequence, clinical features, and molecular epidemiology of another novel HCoV, HCoV-HKU1, in the genus *Betacoronavirus* (14–16). As for animal CoVs, we and others have described the discovery of SARS-CoV-like viruses in horseshoe bats in the Hong Kong Special Administrative Region (HKSAR) and other provinces of China (17, 18). In addition, we have also discovered 19 other animal CoVs, which include two novel lineages in *Betacoronavirus* and a novel genus, *Deltacoronavirus* (19–27). From our studies, it

was shown that bats are the gene source for *Alphacoronavirus* and *Betacoronavirus* and birds are the gene source for *Gammacoronavirus* and *Deltacoronavirus*, to fuel coronavirus evolution and dissemination (26). Recently, a novel CoV, named Middle East respiratory coronavirus (MERS-CoV), that is closely related to *Tylonycteris* bat coronavirus HKU4 and *Pipistrellus* bat coronavirus HKU5 has emerged as a cause of severe respiratory infections associated with high rates of mortality (28–30). Its isolation supported the suggestion that CoVs are important causes of major epidemics; and therefore, continuous discovery of novel CoVs and genomic and phylogenetic studies of these viruses are of crucial importance.

In the genus *Deltacoronavirus*, it was shown that, in addition to bird CoVs, it also comprises porcine coronavirus HKU15 from pigs (26). Similarly, in the genus *Gammacoronavirus*, in addition to *Avian coronavirus*, which consists of infectious bronchitis virus (IBV) and its closely related bird CoVs, a novel CoV was discovered from a white beluga whale in 2008 (31). As a result of this discovery, we hypothesized that there is a distinct group or species

Received 20 August 2013 Accepted 5 November 2013

Published ahead of print 13 November 2013

Address correspondence to Kwok-Yung Yuen, kyyuen@hkucc.hku.hk.

P.C.Y.W. and S.K.P.L. contributed equally to this article.

Copyright © 2014, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.02351-13

TABLE 1 Marine mammals screened in the present study

Marine mammal	Scientific name	Sample type	No. of samples tested	No. (%) of samples positive for CoV	CoV recovered
California sea lions	<i>Zalophus californianus</i>	Respiratory swab	6	0 (0)	
		Fecal swab	44	0 (0)	
Harbor seals	<i>Phoca vitulina</i>	Respiratory swab	3	0 (0)	
		Fecal swab	12	0 (0)	
Indo-Pacific bottlenose dolphins	<i>Tursiops aduncus</i>	Blood	31	0 (0)	BdCoV HKU22
		Respiratory swab	49	0 (0)	
		Fecal swab	48	3 (6.3)	

of marine mammal CoVs in *Gammacoronavirus*. To test this hypothesis, we carried out a molecular epidemiology study in marine mammals of the Ocean Park in HKSAR. Based on the results of comparative genome and phylogenetic analyses in the present study, we propose a novel bottlenose dolphin CoV in *Gammacoronavirus*. This distinct species of marine mammal CoV in *Gammacoronavirus* is also discussed.

MATERIALS AND METHODS

Marine mammal surveillance and sample collection. All respiratory, fecal, and blood samples were collected from the marine mammals by veterinary surgeons of the Ocean Park in HKSAR from August 2008 to July 2010 using standard procedures approved by the Animal Welfare Committee of Ocean Park (5, 25). A total of 18 Indo-Pacific bottlenose dolphins (*Tursiops aduncus*), 20 California sea lions (*Zalophus californianus*), and 7 harbor seals (*Phoca vitulina*) in the Ocean Park were tested in this study.

RNA extraction. Viral RNA was extracted from the respiratory, fecal, and blood samples using a QIAamp viral RNA minikit (Qiagen, Hilden, Germany). The RNA was eluted in 60 µl of AVE buffer (Qiagen, Hilden, Germany) and was used as the template for reverse transcription-PCR (RT-PCR).

RT-PCR of the RdRp gene of CoVs using conserved primers and DNA sequencing. Initial CoV screening was performed by amplifying a 440-bp fragment of the RNA-dependent RNA polymerase (RdRp) gene of CoVs using conserved primers (5'-GGTGGGACTATCCTAAGTGTG A-3' and 5'-ACCATCATCNGANARDATCATNA-3') designed by multiple alignments of the nucleotide sequences of available RdRp genes of known CoVs. After the detection of the novel CoV from the bottlenose dolphin sample, subsequent screening was performed by amplifying the same 440-bp fragment of the RdRp gene using specific primers (5'-GGT TGGGACTATCCTAAGTGTGA-3' and 5'-CCATCATCGCTCAATATC ATGAGA-3'). Reverse transcription was performed using a SuperScript III kit (Invitrogen, San Diego, CA). The PCR mixture (25 µl) contained cDNA, PCR buffer (10 mM Tris-HCl, pH 8.3, 50 mM KCl, 3 mM MgCl₂, 0.01% gelatin), 200 µM each deoxynucleoside triphosphate, and 1.0 U *Taq* polymerase (Applied Biosystems, Foster City, CA). The mixtures were amplified by 60 cycles of 94°C for 1 min, 48°C for 1 min, and 72°C for 1 min with a final extension at 72°C for 10 min in an automated thermal cycler (Applied Biosystems, Foster City, CA). Standard precautions were taken to avoid PCR contamination, and no false-positive results were observed for the negative controls.

The PCR products were gel purified using a QIAquick gel extraction kit (Qiagen, Hilden, Germany). Both strands of the PCR products were sequenced twice with an ABI Prism 3700 DNA analyzer (Applied Biosystems, Foster City, CA), using the two PCR primers. The sequences of the PCR products were compared with known sequences of the RdRp genes of CoVs in the GenBank database.

Viral culture. Original fecal samples from the three bottlenose dolphins that tested positive for CoV were cultured in cells of the HRT-18G

(human colorectal adenocarcinoma; ATCC CRL-11663), Vero E6 (African green monkey kidney; ATCC CRL-1586), Caco-2 (human colorectal adenocarcinoma; ATCC HTB-37), and LLC-MK2 (rhesus monkey kidney; ATCC CCL-7) cell lines.

Complete genome sequencing. Three complete genomes of bottlenose dolphin coronavirus (BdCoV) HKU22 were amplified and sequenced using the RNA extracted from the fecal specimens as the templates. The RNA was converted to cDNA by a combined random priming and oligo(dT) priming strategy. The cDNA was amplified by degenerate primers designed by multiple alignments of the genomes of other CoVs with complete genomes available, using strategies described in our previous publications (15, 20) and the CoV database, CoVDB (32), for sequence retrieval. Additional primers were designed from the results of the first and subsequent rounds of sequencing. The 5' ends of the viral genomes were confirmed by rapid amplification of cDNA ends (RACE) using a 5'/3' RACE kit (Roche Diagnostics GmbH, Mannheim, Germany). Sequences were assembled and manually edited to produce final sequences of the viral genomes.

Genome analysis. The nucleotide sequences of the genomes and the deduced amino acid sequences of the open reading frames (ORFs) were compared to those of other CoVs using the EMBOSS needle program (<http://www.ebi.ac.uk>). Phylogenetic tree construction was performed using the neighbor-joining method with 1,000 bootstraps and the Jones-Taylor-Thornton (JTT) substitution model with a gamma distribution among sites conducted in the MEGA (v5) program (33). Protein family analysis was performed using the PFAM and InterProScan programs (34, 35). Prediction of transmembrane domains was performed using the TM-HMM program (36).

Genomes of other CoVs (and their GenBank accession numbers) included in the comparative analysis were as follows: porcine epidemic diarrhea virus (PEDV; NC_003436), *Scotophilus* bat coronavirus 512 (Sc-BatCoV-512; NC_009657), transmissible gastroenteritis virus (TGEV; DQ811789), feline infectious peritonitis virus (FIPV; AY994055), canine coronavirus (CCoV; GQ477367), porcine respiratory coronavirus (PRCV; DQ811787), mink coronavirus (MCoV; HM245925), *Rhinolophus* bat coronavirus HKU2 (RhBatCoV HKU2; EF203064), *Miniopterus* bat coronavirus 1A (Mi-BatCoV 1A; NC_010437), *Miniopterus* bat coronavirus 1B (Mi-BatCoV 1B; NC_010436), *Miniopterus* bat coronavirus HKU8 (Mi-BatCoV HKU8; NC_010438), human coronavirus 229E (HCoV-229E; NC_002645), human coronavirus NL63 (HCoV-NL63; NC_005831), *Rousettus* bat coronavirus HKU10 (Ro-BatCoV HKU10; JQ989270), *Hipposideros* bat coronavirus HKU10 (Hi-BatCoV HKU10; JQ989266), human coronavirus OC43 (HCoV OC43; NC_005147), bovine coronavirus (BCoV; NC_003045), sable antelope CoV (antelope CoV; EF424621), giraffe coronavirus (GiCoV; EF424622), equine coronavirus (ECoV; NC_010327), porcine hemagglutinating encephalomyelitis virus (PHEV; NC_007732), murine hepatitis virus (MHV; NC_001846), canine respiratory coronavirus (CRCoV; JX860640), rat coronavirus (RCoV; NC_012936), human coronavirus HKU1 (HCoV-HKU1; NC_006577), rabbit coronavirus HKU14 (RbCoV HKU14; JN874559), *Tylonycteris* bat coronavirus HKU4 (Ty-Bat-

TABLE 2 Comparison of genomic features and amino acid identities of BdCoV HKU22 and other CoVs with complete genome sequences available^a

CoV (sample no.)	Genome features		% pairwise amino acid identity with BdCoV HKU22 (CF090325)				
	Size (no. of bases)	G+C content	3CL ^{pro}	Pol	Hel	S	N
<i>Alphacoronavirus</i>							
PEDV	28,033	0.42	40.8	58.5	57.5	28.9	25.0
TGEV	28,586	0.38	40.3	58.7	58.0	28.0	27.0
FIPV	29,355	0.38	41.3	58.2	58.0	28.6	25.7
CCoV	29,363	0.38	40.7	58.5	58.0	28.2	27.0
PRCV	27,550	0.37	40.3	58.7	58.0	27.9	27.5
HCoV-229E	27,317	0.38	42.3	58.0	57.4	28.1	25.1
HCoV-NL63	27,553	0.34	41.8	58.7	57.0	28.4	28.5
RhBatCoV HKU2	27,165	0.39	41.6	58.5	55.4	23.1	28.4
Mi-BatCoV 1A	28,326	0.38	38.8	59.1	57.7	25.4	29.6
Mi-BatCoV 1B	28,476	0.39	37.5	58.3	57.4	26.7	28.3
Mi-BatCoV HKU8	28,773	0.42	40.6	59.4	58.0	27.4	27.1
Sc-BatCoV-512	28,179	0.40	39.9	57.8	57.4	27.1	30.6
Ro-BatCoV HKU10	28,494	0.38	41.2	57.6	58.4	26.9	26.5
Hi-BatCoV HKU10	28,492	0.38	41.2	57.3	58.4	25.9	25.6
MCoV	28,894	0.38	41.5	58.3	57.2	28.3	26.7
<i>Betacoronavirus</i>							
Lineage A							
HCoV OC43	30,738	0.37	45.6	60.3	58.2	26.1	29.0
BCoV	31,028	0.37	45.3	60.0	58.2	25.0	28.4
PHEV	30,480	0.37	45.3	60.2	58.2	26.5	28.8
Antelope CoV	30,995	0.37	45.3	60.0	58.2	24.6	28.4
GiCoV	30,979	0.37	45.3	60.0	58.2	25.2	28.4
ECoV	30,992	0.37	45.3	60.3	58.4	27.0	28.9
MHV	31,357	0.42	45.6	59.7	58.9	26.7	30.3
HCoV-HKU1	29,926	0.32	45.6	59.3	56.1	26.5	31.2
RCoV	31,250	0.41	46.3	59.5	58.6	26.7	30.1
RbCoV HKU14	31,100	0.38	45.0	60.2	58.3	25.7	29.5
CRCoV	31,028	0.37	45.3	59.8	58.2	26.8	28.9
Lineage B							
SARS-CoV	29,751	0.41	42.7	60.6	55.7	22.4	25.7
SARSr-CiCoV	29,728	0.41	42.7	60.6	55.7	22.5	25.7
SARSr-Rh-BatCoV HKU3	29,704	0.41	42.7	60.7	55.5	23.1	26.1
SARSr-CoV CFB	29,734	0.41	42.7	60.6	55.5	23.4	25.7
Lineage C							
Ty-BatCoV HKU4	30,286	0.38	43.9	58.7	58.6	26.7	26.1
Pi-BatCoV HKU5	30,488	0.43	42.3	59.0	59.1	24.8	26.7
MERS-CoV	30,119	0.41	43.1	59.4	58.8	25.2	26.3
Lineage D, Ro-BatCoV HKU9	29,114	0.41	39.5	60.3	58.2	25.4	28.9
<i>Gammacoronavirus</i>							
<i>Avian coronavirus</i>							
IBV	27,608	0.38	50.5	71.0	74.9	27.9	36.2
TCoV	27,657	0.38	50.5	70.2	74.0	28.2	35.4
<i>Cetacean coronavirus</i>							
BWCoV SW1	31,686	0.39	99.0	99.2	99.3	74.7	96.3
BdCoV HKU22 (CF090327)	31,750	0.39	100.0	99.8	99.8	99.2	98.4
BdCoV HKU22 (CF090331)	31,759	0.39	100.0	100.0	100.0	98.7	98.4
<i>Deltacoronavirus</i>							
BuCoV HKU11	26,476	0.39	39.9	53.0	51.8	27.2	31.2
ThCoV HKU12	26,396	0.38	39.2	53.3	51.8	26.8	30.2
MunCoV HKU13	26,552	0.43	37.2	53.1	53.0	27.5	30.4
PorCoV HKU15	25,421	0.43	39.5	53.4	53.0	27.3	31.4
WECov HKU16	26,027	0.40	39.5	53.3	51.8	27.3	30.1
SpCoV HKU17	26,067	0.45	39.5	53.6	53.0	28.3	30.8
MRCov HKU18	26,674	0.47	37.2	53.0	52.3	27.3	30.7
NHCoV HKU19	26,064	0.38	41.1	53.1	54.7	27.2	28.9
WiCoV HKU20	26,211	0.39	42.0	52.5	51.2	27.6	30.0
CMCoV HKU21	26,216	0.35	38.8	52.6	52.9	28.3	31.0

^a Comparison of genomic features of BdCoV HKU22 and other CoVs for which complete genome sequences are available and of amino acid identities between the predicted 3CL^{pro}, RdRp, Hel, S, and N proteins of BdCoV HKU22 and the corresponding proteins of other CoVs.

CoV HKU4; [NC_009019](#)), *Pipistrellus* bat coronavirus HKU5 (Pi-BatCoV HKU5; [NC_009020](#)), Middle East respiratory syndrome coronavirus (MERS-CoV; [JX869059](#)), SARS-related human coronavirus (SARS-CoV; [NC_004718](#)), SARS-related *Rhinolophus* bat coronavirus HKU3 (SARSr-Rh-BatCoV HKU3; [DQ022305](#)), SARS-related Chinese ferret badger coronavirus (SARSr-CoV CFB; [AY545919](#)), SARS-related palm civet coronavirus (SARSr-CiCoV; [AY304488](#)), *Rousettus* bat coronavirus HKU9

(Ro-BatCoV HKU9; [NC_009021](#)), infectious bronchitis virus (IBV; [NC_001451](#)), partridge coronavirus (IBV-partridge; [AY646283](#)), turkey coronavirus (TCoV; [NC_010800](#)), peafowl coronavirus (IBV-peafowl; [AY641576](#)), Beluga whale coronavirus SW1 (BWCoV SW1; [NC_010646](#)), bulbul coronavirus HKU11 (BuCoV HKU11; [FJ376619](#)), thrush coronavirus HKU12 (ThCoV HKU12; [FJ376621](#)), munia coronavirus HKU13 (MunCoV HKU13; [FJ376622](#)), porcine coronavirus HKU15 (PorCoV

TABLE 3 Coding potential and putative TRSs of the genomes of BdCoV HKU22

BdCoV HKU22 genome sample no.	ORF	Location (nt)	Length			Putative TRS	
			No. of nucleotides	No. of amino acids	Frame(s)	TRS location (nucleotide)	TRS (distance [no. of bases] to AUG) ^a
CF090325	1ab	520–20510	19991	6663	+1, +3	68	AAACA (447)AUG
	S	20446–24927	4482	1493	+1	20411	AAACA (30)AUG
	E	24933–25220	288	95	+3		
	M	25204–26028	825	274	+1	25137	AAACA (62)AUG
	5a	26031–26447	417	138	+3	26027	AAACA UG
	5b	26440–26958	519	172	+1		
	5c	26948–27475	528	175	+2		
	6	27479–28165	687	228	+2	27474	AAACAA UG
	7a	28168–28359	192	63	+1	28164	AAACA UG
	7b	28382–28486	105	34	+2		
	8	28804–28983	180	59	+1	28792	AAACAGAAAGGCA UG
	9	29040–29498	459	152	+3	29036	AAACA UG
	10	29501–30133	633	210	+2	29497	AAACA UG
	N	30245–31384	1140	379	+2	30132	AAACA (108)AUG
CF090327	1ab	520–20510	19991	6663	+1, +3	68	AAACA (447)AUG
	S	20446–24924	4479	1492	+1	20411	AAACA (30)AUG
	E	24930–25217	288	95	+3		
	M	25201–26025	825	274	+1	25134	AAACA (62)AUG
	5a	26028–26444	417	138	+3	26024	AAACA UG
	5b	26437–26955	519	172	+1		
	5c	26945–27472	528	175	+2		
	6	27476–28162	687	228	+2	27471	AAACAA UG
	7	28165–28650	486	161	+1	28161	AAACA UG
	8	28801–28980	180	59	+1	28789	AAACAGAAAGGCA UG
	9	29037–29495	459	152	+3	29033	AAACA UG
	10	29498–30130	633	210	+2	29494	AAACA UG
	N	30242–31381	1140	379	+2	30129	AAACA (108)AUG
CF090331	1ab	520–20510	19991	6663	+1, +3	68	AAACA (447)AUG
	S	20446–24933	4488	1495	+1	20411	AAACA (30)AUG
	E	24939–25226	288	95	+3		
	M	25210–26034	825	274	+1	25143	AAACA (62)AUG
	5a	26037–26453	417	138	+3	26033	AAACA UG
	5b	26446–26964	519	172	+1		
	5c	26954–27481	528	175	+2		
	6	27485–28171	687	228	+2	27480	AAACAA UG
	7a	28174–28365	192	63	+1	28170	AAACA UG
	7b	28388–28492	105	34	+2		
	8	28810–28989	180	59	+1	28798	AAACAGAAAGGCA UG
	9	29046–29504	459	152	+3	29042	AAACA UG
	10	29507–30139	633	210	+2	29503	AAACA UG
	N	30251–31390	1140	379	+2	30138	AAACA (108)AUG

^a Boldface indicates putative TRSs.

HKU15; [JQ065042](#)), white-eye coronavirus HKU16 (WECov HKU16; [JQ065044](#)), sparrow coronavirus HKU17 (SpCoV HKU17; [JQ065045](#)), magpie-robin coronavirus HKU18 (MRCov HKU18; [JQ065046](#)), night-heron coronavirus HKU19 (NHCoV HKU19; [JQ065047](#)), wigeon coronavirus HKU20 (WiCoV HKU20; [JQ065048](#)), and common moorhen coronavirus HKU21 (CMCoV HKU21; [JQ065049](#)).

Real-time quantitative RT-PCR. For real-time quantitative RT-PCR assays, cDNA was amplified with a FastStart DNA Master SYBR green I mix reagent kit (Roche Diagnostics GmbH, Mannheim, Germany). Briefly, 20 µl of reaction mixtures containing 2 µl cDNA, 3 mM MgCl₂, and 0.25 M forward and reverse primers (5'-CTGCTTATGCCAACAGT GCTT-3' and 5'-AAGTCCATATCGGGCTTAT-3') was thermal cycled at 95°C for 10 min, followed by 50 cycles of 95°C for 10 s, 55°C for 5 s, and 72°C for 7 s, using a LightCycler apparatus (Roche Diagnostics GmbH,

Mannheim, Germany). A plasmid with the target sequence was used for generating the standard curve. At the end of the assay, PCR products (a 165-bp fragment of the RdRp gene) were subjected to a melting curve analysis (65 to 95°C, 0.1°C/s) to confirm the specificity of the assay.

Cloning and purification of His₆-tagged recombinant N protein of BdCoV HKU22. To produce a plasmid for protein purification, primers 5'-TACAGCTAGCATGGCCTCTACATCGGGAAAAG-3' and 5'-GCAT GCTAGCTTAAGCTTCAGACCATTCAAG-3' were used to amplify the gene encoding the nucleocapsid (N) protein of BdCoV HKU22 by RT-PCR. The sequence coding for amino acid residues 1 to 380 of the N protein was amplified and cloned into the NheI sites of expression vector pET-28b(+) (Novagen, Madison, WI) in frame and downstream of the series of six histidine residues. The recombinant N protein was expressed and purified using an Ni²⁺-loaded HiTrap chelating system (GE Health-

HCoV-229E

HCoV-NL63

TGEV

PEDV

PRCV

FIPV

Sc-BatCoV-512

RhBatCoV-HKU2

Mi-BatCoV 1

Mi-BatCoV-HKU8

Ro-BatCoV HKU10

CCoV

MCoV

Betacoronavirus

A

HCoV-OC43

HCoV-HKU1

BCoV

MHV

RCoV

PHEV

ECoV

AntelopeCoV

GiCoV

RbCoV HKU14

CRCoV

B

SARS CoV

SARSr-CiCoV

SARSr-Rh-BatCoV HKU3

SARSr CoV CFB

C

Ty-BatCoV HKU4

Pi-BatCoV HKU5

MERS-CoV

D

Ro-BatCoV HKU9

Gammacoronavirus

Avian coronavirus

IBV

TCoV

Cetacean coronavirus

BWCoV-SW1

BdCoV HKU22

Deltacoronavirus

BuCoV HKU11

ThCoV HKU12

MunCoV HKU13

PorCoV HKU15

WECov HKU16

SpCoV HKU17

MRCov HKU18

NHCoV HKU19

WiCoV HKU20

CMCoV HKU21

FIG 1 Genome organizations of BdCoV HKU22 and representative CoVs from each genus. Orange boxes, papain-like proteases (PL1^{PRO}, PL2^{PRO}, and PL^{PRO}), chymotrypsin-like protease (3CL^{PRO}), and RNA-dependent RNA polymerase (RdRp); green boxes, hemagglutinin esterase (HE), spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins; blue boxes, putative accessory proteins.

TABLE 4 Characteristics of putative nonstructural proteins of ORF1ab in BdCoV HKU22, BWCoV SW1, and IBV

nsp	Putative function or domain ^a	Protein and cleavage site characteristic					
		BdCoV HKU22 (CF090325)		BWCoV SW1		IBV	
		ORF1a/ORF1b positions (amino acids)	C-end cleavage ^b	ORF1a/ORF1b positions (amino acids)	C-end cleavage	ORF1a/ORF1b positions (amino acids)	C-end cleavage
nsp2	Unknown	M ¹ -G ⁶³⁷	DG^DI	M ¹ -G ⁶³⁷	DG^DI	M ¹ -G ⁶⁷³	AG^GK
nsp3	ADRP, PL ^{pro}	D ⁶³⁸ -G ²²²²	LG^GV	D ⁶³⁸ -G ²²²²	LG^GV	G ⁶⁷⁴ -G ²²⁶⁵	AG^GI
nsp4	Hydrophobic domain	G ²²²³ -Q ²⁷⁵⁹	LQ^AG	G ²²²³ -Q ²⁷⁵⁹	LQ^AG	G ²²⁶⁶ -Q ²⁷⁷⁹	LQ^SG
nsp5	3CL ^{pro}	A ²⁷⁶⁰ -Q ³⁰⁶²	LQ^SN	A ²⁷⁶⁰ -Q ³⁰⁶²	LQ^SN	S ²⁷⁸⁰ -Q ³⁰⁸⁶	LQ^SS
nsp6	Hydrophobic domain	S ³⁰⁶³ -Q ³³⁶⁵	VQ^SK	S ³⁰⁶³ -Q ³³⁶⁵	VQ^SK	S ³⁰⁸⁷ -Q ³³⁷⁹	VQ^AK
nsp7	Unknown	S ³³⁶⁶ -Q ³⁴⁴⁸	LQ^AV	S ³³⁶⁶ -Q ³⁴⁴⁸	LQ^AV	A ³³⁸⁰ -Q ³⁴⁶²	LQ^SV
nsp8	Unknown	A ³⁴⁴⁹ -Q ³⁶⁴⁶	LQ^NN	A ³⁴⁴⁹ -Q ³⁶⁴⁶	LQ^NN	S ³⁴⁶³ -Q ³⁶⁷²	LQ^NN
nsp9	Unknown	N ³⁶⁴⁷ -Q ³⁷⁵⁸	LQ^HG	N ³⁶⁴⁷ -Q ³⁷⁵⁸	LQ^HG	N ³⁶⁷³ -Q ³⁷⁸³	LQ^SK
nsp10	Unknown	H ³⁷⁵⁹ -Q ³⁹⁴⁷	LQ^SV	H ³⁷⁵⁹ -Q ³⁹⁴⁷	LQ^SV	S ³⁷⁸⁴ -Q ³⁹²⁸	LQ^SV
nsp11	Unknown (short peptide at the end of ORF1a)	S ³⁹⁴⁸ -D ³⁹⁶⁴		S ³⁹⁴⁸ -D ³⁹⁶⁴		S ³⁹²⁹ -G ³⁹⁵¹	
nsp12	RdRp	S ³⁹⁴⁸ -Q ⁴⁸⁷³	LQ^AS	S ³⁹⁴⁸ -Q ⁴⁸⁷³	LQ^AS	S ³⁹²⁹ -Q ⁴⁸⁶⁸	LQ^SC
nsp13	Hel	A ⁴⁸⁷⁴ -Q ⁵⁴⁷⁴	LQ^SQ	A ⁴⁸⁷⁴ -Q ⁵⁴⁷⁴	LQ^SQ	S ⁴⁸⁶⁹ -Q ⁵⁴⁶⁸	LQ^GT
nsp14	ExoN	S ⁵⁴⁷⁵ -Q ⁶⁰⁰²	LQ^SL	S ⁵⁴⁷⁵ -Q ⁶⁰⁰²	LQ^SL	G ⁵⁴⁶⁹ -Q ⁵⁹⁸⁹	LQ^SI
nsp15	NendoU	S ⁶⁰⁰³ -Q ⁶³⁵¹	LQ^AD	S ⁶⁰⁰³ -Q ⁶³⁵¹	LQ^AD	S ⁵⁹⁹⁰ -Q ⁶³²⁷	LQ^SA
nsp16	2'-O-MT	A ⁶³⁵² -I ⁶⁶⁶³		A ⁶³⁵² -N ⁶⁶⁶³		S ⁶³²⁸ -M ⁶⁶²⁹	

^a ADRP, ADP-ribose-1"-phosphatase; PL^{pro}, papain-like protease; 3CL^{pro}, 3C-like protease; RdRp, RNA-dependent RNA polymerase; Hel, helicase; ExoN, 3'-to-5' exonuclease; NendoU, nidoviral uridylyate-specific endoribonuclease; 2'-O-MT, 2'-O-ribose methyltransferase.

^b ^, cleavage site.

care, Buckinghamshire, United Kingdom) according to the manufacturer's instructions.

Western blot analysis. Western blot analysis was performed according to our published protocol (15). Briefly, 600 ng of purified His₆-tagged recombinant N protein of BdCoV HKU22 was loaded into each well of a sodium dodecyl sulfate (SDS)-10% polyacrylamide gel and subsequently electroblotted onto a nitrocellulose membrane (Bio-Rad, Hercules, CA). The blot was cut into strips, and the strips were incubated separately with 1:1,000 dilutions of serial serum samples obtained from the three bottlenose dolphins that were RT-PCR positive for BdCoV HKU22 as well as nine bottlenose dolphins that were RT-PCR negative for BdCoV HKU22. Antigen-antibody interaction was detected with 1:8,000-diluted horseradish peroxidase-conjugated anti-bottlenose dolphin IgG (Abcam) and an ECL fluorescence system (GE Healthcare, Buckinghamshire, United Kingdom).

Estimation of synonymous and nonsynonymous substitution rates.

The number of synonymous substitutions per synonymous site (K_s) and the number of nonsynonymous substitutions per nonsynonymous site (K_a) for each coding region between each pair of strains were calculated using the Nei-Gojobori method (Jukes-Cantor) in MEGA (v5) (33).

Estimation of divergence dates. A maximum likelihood tree under the GTR+ gamma model of substitution was constructed using the PhyML program (37). A root-to-tip genetic distance determined from a maximum likelihood tree was plotted against the year of sampling using the Path-O-Gen (v1.4) tool (<http://tree.bio.ed.ac.uk/software/pathogen/>). The complete genome sequences of three strains of BdCoV HKU22 and one strain of BWCoV SW1 were used. The crossing point was taken as the time of the most recent common ancestor (tMRCA) for the four viruses under analysis.

Nucleotide sequence accession numbers. The nucleotide sequences of the three genomes of BdCoV HKU22 have been lodged within the

FIG 2 Amino acid differences and selection pressure analysis of S proteins of BdCoV HKU22 and BWCoV SW1. (A) Distribution of amino acid changes in the S proteins of BdCoV HKU22 and BWCoV SW1. The positions of the amino acid changes are depicted by vertical lines, and deletions are marked by asterisks. (B) Distribution of positively selected sites in S protein identified using random effects likelihood (REL) (39). $dN > dS$, more nonsynonymous than synonymous substitutions.

FIG 3 Phylogenetic analyses of the ORF1b polyprotein and the S and N proteins of BdCoV HKU22. The trees were constructed by using the neighbor-joining method in the JTT substitution model with a gamma-distributed rate variation and bootstrap values calculated from 1,000 trees. Bootstrap values below 70% are not shown. A total of 2,716, 1,495, and 379 amino acid positions in ORF1b polyprotein, S protein, and N protein, respectively, were included in the analyses. The tree was rooted to Breda virus (AY_427798). For the ORF1b polyprotein, the scale bars indicate the estimated number of substitutions per 10 amino acids. For the S and N proteins, the scale bars indicate the estimated number of substitutions per 5 amino acids. The three strains of BdCoV HKU22 characterized in this study are in boldface.

FIG 3 continued

GenBank sequence database under accession numbers [KF793824](#), [KF793825](#), and [KF793826](#).

RESULTS

Marine mammal surveillance and identification of CoV in Indo-Pacific bottlenose dolphin. A total of 193 samples, including 58 respiratory, 31 blood, and 104 fecal samples, were obtained from Indo-Pacific bottlenose dolphins, California sea lions, and harbor seals ([Table 1](#)). RT-PCR was positive for a 440-bp fragment in the RdRp genes of CoVs in specimens from the fecal samples of three Indo-Pacific bottlenose dolphins (samples [CF090325](#), [CF090327](#), and [CF090331](#)). Sequencing results suggested the presence of a CoV with 98% nucleotide sequence identity to BWCoV SW1 ([Table 1](#)). No positive results were obtained from any of the California sea lions and harbor seals tested ([Table 1](#)).

Viral culture. Attempts to stably passage BdCoV HKU22 in cell cultures were unsuccessful, with no cytopathic effect or viral replication being detected.

Viral load. Quantitative real-time RT-PCR showed that the amounts of BdCoV HKU22 RNA ranged from 1×10^3 to 1×10^5

copies per ml in the three fecal samples positive for BdCoV HKU22.

Genome organization and coding potential of BdCoV HKU22.

Complete genome sequence data for the three strains of BdCoV HKU22 from the three bottlenose dolphins were obtained by assembly of the sequences of the RT-PCR products from the RNA extracted from the corresponding individual specimens. The sizes of the three genomes of BdCoV HKU22 were 31,750 to 31,759 bases, and their G+C contents were 39% ([Table 2](#)).

BdCoV HKU22 and BWCoV SW1 possess the same putative transcription regulatory sequence (TRS) motif, 5'-AAACA-3', at the 3' end of the leader sequence and preceding most ORFs ([Table 3](#)). This TRS has not been found to be the TRS for any CoVs for which complete genome sequences are available other than BdCoV HKU22 and BWCoV SW1. Interestingly, this putative TRS overlapped with the initiation codon AUG by 1 base in ORFs NS5a, NS7, NS9, and NS10. This has never been observed in any other CoVs for which complete genome sequences are available. Notably, the TRS of the N protein in the genomes of BdCoV HKU22 was separated from the corresponding AUG by 108 bases

FIG 4 Western blot analysis of purified recombinant BdCoV HKU22 N-protein antigen. Lanes 1 to 4, 5 to 8, and 9 to 12, serial serum samples collected from the three bottlenose dolphins RT-PCR positive for BdCoV HKU22, respectively. Prominent immunoreactive protein bands of about 42 kDa (arrowheads) were visible 4 to 8 weeks after the fecal samples of the three bottlenose dolphins tested positive for BdCoV HKU22 (lanes 3 and 4, 7 and 8, and 11 and 12). Only very faint bands were observed for serum samples obtained before and within a few days after their fecal samples tested positive for BdCoV HKU22 (lanes 1 and 2, 5 and 6, and 9 and 10). The dates of serum collection are indicated below each lane. The dates that the dolphins were RT-PCR positive and RT-PCR negative for BdCoV HKU22 are also shown.

(Table 3). This long stretch of nucleotides between the TRS of N and the corresponding AUG was also observed in the genomes of other gammacoronaviruses (90 bases for IBV, 93 bases for TCoV, and 105 bases for BWCoV SW1), which is in contrast to the relatively small number of bases (≤ 16 bases) between the TRS for the N protein and the corresponding AUG in other genera of CoVs. Upstream of the initiation codons of NS5a, a short stretch of 13 nucleotides, CUUUAUUCUGUUU in BdCoV HKU22 and CUUUAUUCGUUU in BWCoV SW1, was observed. These short stretches of 13 nucleotides were homologous to the known internal ribosome entry site (IRES) element UUUUAUUCUUUUU of the envelope (E) gene in murine hepatitis virus and, hence, could be potential IRES elements for NS5b and/or NS5c of BdCoV HKU22 and BWCoV SW1, as these two ORFs do not possess the putative TRS upstream of their initiation codons.

In addition to the same putative TRS, BdCoV HKU22 also possesses the same genome structure as BWCoV SW1 (Fig. 1). The replicase ORF1ab occupies 19.991 kb of the BdCoV HKU22 genome (Table 3). This ORF encodes a number of putative non-structural proteins (nsp's), including nsp3 (which contains the putative papain-like protease [PL^{pro}]), nsp5 (a putative chymotrypsin [3C]-like protease [3CL^{pro}]), nsp12 (a putative RdRp), nsp13 (a putative helicase [Hel]), and other proteins of unknown functions. The proteolytic cleavage sites between the nonstructural proteins of ORF1ab in the genomes of BdCoV HKU22 and BWCoV SW1 were identical (Table 4). Similar to BWCoV SW1, the genome of BdCoV HKU22 contained one putative PL^{pro},

TABLE 5 Estimation of nonsynonymous and synonymous substitution rates in the genomes of BdCoV HKU22

Gene	K_a	K_s	K_a/K_s
nsp2	0.002	0.002	1
nsp3	0	0.001	0
nsp4	0.001	0	
nsp5	0	0	
nsp6	0.003	0.003	1
nsp7	0	0	
nsp8	0.005	0.002	2.5
nsp9	0.005	0	
nsp10	0.002	0.005	0.4
nsp11	0	0	
nsp12	0.001	0.007	0.143
nsp13	0	0.005	0
nsp14	0.002	0.006	0.333
nsp15	0.002	0.006	0.333
nsp16	0.002	0.006	0.333
S	0.005	0.002	2.5
E	0	0	
M	0.003	0	
NS5a	0.004	0	
NS5b	0.003	0.006	0.5
NS5c	0.002	0	
NS6	0	0.009	0
NS7 ^a			
NS8	0.005	0	
NS9	0	0.006	0
NS10	0.006	0.005	1.2
N	0.005	0.005	1

^a The sequences contain a premature stop codon.

FIG 5 Estimation of date of divergence of BdCoV HKU22 and BWCov SW1. (A) Regression of root-to-tip distances against the date of sampling of four genomic sequences to estimate the rate of evolution and the time of the most recent common ancestor. (B) Maximum likelihood tree with time scale using the estimated rate of evolution.

which is homologous to the PL2^{pro} enzymes of *Alphacoronavirus* and *Betacoronavirus* lineage A and the PL^{pro} enzymes of *Betacoronavirus* lineages B, C, and D, *Avian coronavirus* of *Gammacoronavirus*, and *Deltacoronavirus* (Fig. 1).

The major difference between the genomes of BWCov SW1 and BdCoV HKU22 was observed in their spike (S) genes, with only 74.3 to 74.7% amino acid identities between the protein encoded by the S gene of BdCoV HKU22 and that encoded by the S gene of BWCov SW1. Most of the difference occurred in the N-terminal half of the S protein (Fig. 2). In addition to nonsynonymous substitutions, a number of deletions/insertions were also observed in the BdCoV HKU22 genomes compared to the sequence of the BWCov SW1 genome (Fig. 2A). Moreover, most of the positive selection sites were also observed in the N-terminal half of the S protein (Fig. 2B). In addition to the S protein, the amino acid sequence of the protein encoded by the membrane (M) gene in BdCoV HKU22 also showed a significant difference (14.5%) compared to that of the protein encoded by the M gene in

BWCov SW1, which was greater than the differences detected in most other parts of the genome. Uniquely, BdCoV HKU22 and BWCov SW1 possessed a number of ORFs (NS5a, NS5b, NS5c, NS6, NS7, NS8, NS9, and NS10) between their M and N genes. These ORFs lead BdCoV HKU22 and BWCov SW1 to have the largest genome sizes among all CoVs. Among these ORFs, the putative amino acid sequence of NS5a in BdCoV HKU22 showed the largest number of differences from the amino acid sequence of NS5a in BWCov SW1 (9.4% amino acid differences in NS5a and 0 to 8.5% amino acid differences in the proteins encoded by other ORFs). Similar to BWCov SW1, a putative transmembrane domain (residues 45 to 65) was detected in NS5b of BdCoV HKU22. The putative protein encoded by NS6 was homologous to the capsid proteins of human astroviruses encoded by ORF2. Similar to BWCov SW1, putative signal peptide sequences were detected in NS7 of BdCoV HKU22. NS7 of two of the three strains of BdCoV HKU22 (CF090325 and CF090331) contained a premature stop codon, resulting in NS7a and NS7b. Interestingly, a pu-

TABLE 6 Comparison of genome characteristics among alphacoronaviruses, betacoronaviruses, gammacoronaviruses, and deltacoronaviruses^a

Virus	Genome size (no. of nucleotides) ^b	Optimal TRS	No. of nsp's in ORF1ab	Papain-like protease	Presence of HE gene	Presence of conserved S1/S2 cleavage site	Presence (no. of) ORFs for putative accessory proteins			Presence of s2m
							Between S and E	Between M and N	Downstream of N	
<i>Alphacoronavirus</i>	27,000–29,000	CUAAAC	16	PL1 ^{pro} and PL2 ^{pro}	–	–	+ (1–4)	–	+/- (0–3)	–
<i>Betacoronavirus</i>										
A	30,000–31,000	CUAAAC	16	PL1 ^{pro} and PL2 ^{pro}	+	+	+ (1–3)	–	–	–
B	~30,000	ACGAAC	16	PL ^{pro}	–	–	+ (1–2)	+ (3–5)	–	+
C	~30,000	ACGAAC	16	PL ^{pro}	–	–	+ (4)	–	–	–
D	~29,000	ACGAAC	16	PL ^{pro}	–	–	+ (1)	–	+ (2)	–
<i>Gammacoronavirus</i>										
<i>Avian coronavirus</i>	~28,000	CUUAAAC	15	PL ^{pro}	–	+	+ (2–3)	+ (2–3)	–	+
<i>Cetacean coronavirus</i>	~32,000	AAACA	15	PL ^{pro}	–	–	–	+ (8–9)	–	–
<i>Deltacoronavirus</i>	25,000–27,000	ACACCA	15	PL ^{pro}	–	–	–	+ (1)	+/- (0–4)	+/-

^a HE, hemagglutinin esterase; s2m, stem-loop II motif; TRS, transcription regulatory sequence.
^b Genome size rounded to the nearest thousand.

tative signal peptide sequence was detected in NS8 of BWCoV SW1 but not BdCoV HKU22 by the Signal IP (v4) program, but transmembrane domains were detected in the corresponding regions in NS8 of BdCoV HKU22 but not BWCoV SW1 by TMHMM, although there were just four amino acid differences between this signal peptide and the transmembrane region in BWCoV SW1 and BdCoV HKU22. This may reflect inherent limitations of these bioinformatics analysis programs for signal peptide and transmembrane domain detection. Similar to BWCoV SW1, a putative transmembrane domain (residues 7 to 27) was detected in NS9 of BdCoV HKU22. Unlike *Avian coronavirus* of *Gammacoronavirus*, most deltacoronaviruses, SARS-related *Rhinolophus* bat coronavirus (SARSr-Rh-BatCoV), and SARS-CoV, no stem-loop II motif (s2m), a conserved RNA element downstream of the gene for the N protein and upstream of the poly(A) tail, was observed in BdCoV HKU22 or BWCoV SW1.

Phylogenetic analyses. The phylogenetic trees constructed us-

ing the amino acid sequences of the ORF1b polyprotein and the S and N proteins of BdCoV HKU22 and other CoVs are shown in Fig. 3, and the pairwise amino acid identities in 3CL^{pro}, RdRp, Hel, S, and N are shown in Table 2. In all three phylogenetic trees, BdCoV HKU22 was clustered with BWCoV SW1 (Fig. 3). On the basis of both phylogenetic tree analyses and amino acid sequence differences, BdCoV HKU22 and BWCoV SW1 should belong to the same clade in the genus *Gammacoronavirus*.

Western blot analysis. Prominent immunoreactive bands were visible for serum samples collected from the three bottlenose dolphins 4 to 8 weeks after their fecal samples tested positive for BdCoV HKU22 (Fig. 4, lanes 3 and 4, 7 and 8, and 11 and 12). The sizes of the bands were about 42 kDa, consistent with the expected size of 41.7 kDa for the full-length His₆-tagged recombinant N protein. Only very faint bands were observed for serum samples obtained before and within a few days after their fecal samples tested positive for BdCoV HKU22 (Fig. 4, lanes 1 and 2, 5 and 6,

FIG 6 Model of CoV evolution with a distinct species, *Cetacean coronavirus*, in *Gammacoronavirus*.

and 9 and 10), as well as for serum samples obtained from the nine bottlenose dolphins that were RT-PCR negative for BdCoV HKU22 (data not shown).

Estimation of synonymous and nonsynonymous substitution rates. The K_a , K_s , and K_a/K_s ratio for the various coding regions in BdCoV HKU22 are shown in Table 5. Although the numbers of synonymous and nonsynonymous mutations were small, it was notable that the K_a/K_s ratios for the nsp8, S, and NS10 genes were >1 , with the nsp8 and S genes having the highest K_a/K_s ratio of 2.5.

Estimation of divergence dates. The rate of nucleotide substitution for BWCoV SW1 and BdCoV HKU22 was estimated to be 6.19×10^{-4} from the root-to-tip regression. The estimated tMRCA for BWCoV SW1 and BdCoV HKU22 was approximately in 1959 (correlation coefficient = 0.96) (Fig. 5).

DISCUSSION

In this study, we discovered a novel CoV in Indo-Pacific bottlenose dolphins. A unique advantage of the present study was the invaluable opportunity that we had to collect serial fecal and serum samples from marine animals in an oceanarium. Among the serial fecal samples collected from the three bottlenose dolphins from which BdCoV HKU22 was detected, only one fecal sample from each bottlenose dolphin was positive for BdCoV HKU22, while the samples collected 7 months before and 13 months after the date of positive detection were negative for BdCoV HKU22 (data not shown). No persistent shedding of BdCoV HKU22 was observed in the same bottlenose dolphin. For the serial serum samples, only very faint bands were observed by Western blotting of the serum samples collected from the three bottlenose dolphins before and during the time that the fecal samples tested positive for BdCoV HKU22, but very high BdCoV HKU22 N-protein-specific antibody levels were detected in their serum samples at 4 to 8 weeks postinfection (Fig. 4). Moreover, an antibody response was not detected in bottlenose dolphins negative for BdCoV HKU22. These findings indicate that BdCoV HKU22 is associated with acute infections in bottlenose dolphins, where viral clearance was associated with a specific adaptive antibody response when the bottlenose dolphins recovered from the infections. In contrast to the previous report that described the detection of BWCoV SW1 in a 13-year-old beluga whale that died after generalized pulmonary disease and terminal acute liver failure (31), BdCoV HKU22 was probably associated with an asymptomatic or mild infection in these three bottlenose dolphins, as none of the three bottlenose dolphins positive for BdCoV HKU22 developed any notable symptoms.

BdCoV HKU22 and BWCoV SW1 represent a distinct species in *Gammacoronavirus*. Comparison of the amino acid identities of the seven conserved replicase domains for species demarcation (ADP-ribose-1"-phosphatase, nsp5 [3CL^{Pro}], nsp12 [RdRp], nsp13 [Hel], nsp14 [ExoN], nsp15 [NendoU], and nsp16 [2'-O-ribose methyltransferase]) (38) between BdCoV HKU22 and BWCoV SW1 showed that the sequences of all seven domains of the two CoVs had more than 90% identity to each other, indicating that these two CoVs should belong to the same CoV species. In addition, the genomes of BdCoV HKU22 and BWCoV SW1 possess unique characteristics distinct from those of other CoVs (Table 6). The size of their genomes is about 32,000 nucleotides, making their genomes the largest among those of all CoVs for which complete genome sequences are available. The putative

TRS for most of their ORFs is AAACA. A large number of ORFs which span more than 4 kb and encode putative nonstructural proteins were observed between their M and N genes. In addition to these unique characteristics, the genomes of BdCoV HKU22 and BWCoV SW1 also differ from those of members of *Avian coronavirus* by the absence of conserved S1/S2 cleavage sites, ORFs between the S and E genes, and s2m downstream of the N gene (Table 6). Phylogenetically, BdCoV HKU22 is closely related to BWCoV SW1, with high bootstrap support in all phylogenetic trees constructed, while these two CoVs are, in turn, closely related to members of *Avian coronavirus* (Fig. 3). The close relatedness of BdCoV HKU22 and BWCoV SW1 is in line with the similar close relatedness of their corresponding hosts, as dolphins and whales both belong to the cetaceans, one of the four groups of marine mammals, namely, cetaceans (whales, dolphins, and porpoises), pinnipeds (seals, sea lions, and walruses), sirenians (manatees and dugongs), and fissipeds (polar bear and otters). Based on all these pieces of evidence, we propose a distinct species, *Cetacean coronavirus*, in *Gammacoronavirus*, which includes BdCoV HKU22 and BWCoV SW1 in the model of CoV evolution (Fig. 6). In this model, BdCoV HKU22 and BWCoV SW1 originated from a common ancestral CoV approximately 54 years ago (Fig. 5). *Gammacoronavirus* also contains the species *Avian coronavirus*, which includes IBV and its closely related bird CoVs.

Despite their close relationships, BdCoV HKU22 and BWCoV SW1 represent two CoVs that probably use different receptors in Indo-Pacific bottlenose dolphins and white beluga whales, respectively. Although they belong to the same CoV species, there is a more than 25% difference between the amino acid sequences of the S protein in BdCoV HKU22 and that in BWCoV SW1. Most of the difference is observed in the N-terminal half of their S proteins, where the receptor binding domains should be located (Fig. 2). No recombination between the S gene of BdCoV HKU22 and that of BWCoV SW1 was observed by Simplot analysis (data not shown). Since all the three bottlenose dolphins were RT-PCR positive for BdCoV HKU22 on the same date and they were housed in the same oceanarium, their infections likely originated from a single source. However, the high K_a/K_s ratios in multiple regions of the genome, with the S gene having a very high K_a/K_s ratio of 2.5, indicated that BdCoV HKU22 may be evolving rapidly, supporting a recent transmission event to a new host. Epidemiological studies in bottlenose dolphins in other parts of the world as well as other groups of marine mammals will not only ascertain whether this CoV recently jumped from birds to marine mammals but also further delineate the phylogenetic map of CoVs in these animals.

ACKNOWLEDGMENTS

We thank Wing-Man Ko, secretary for food and health, and Constance Chan, director of the Department of Health, HKSAR, The Peoples' Republic of China. We also thank the Ocean Park Corporation for access to the specimen collection.

This work was partly supported by a Research Grant Council grant, University Grant Council; the Consultancy Service for Enhancing Laboratory Surveillance of Emerging Infectious Disease for the HKSAR Department of Health; the HKSAR Health and Medical Research Fund; and the Strategic Research Theme Fund, The University of Hong Kong.

REFERENCES

- Brian DA, Baric RS. 2005. Coronavirus genome structure and replication. *Curr. Top. Microbiol. Immunol.* 287:1–30. http://dx.doi.org/10.1007/3-540-26765-4_1.
- Lai MM, Cavanagh D. 1997. The molecular biology of coronaviruses. *Adv. Virus Res.* 48:1–100. [http://dx.doi.org/10.1016/S0168-1702\(96\)01421-9](http://dx.doi.org/10.1016/S0168-1702(96)01421-9).
- Ziebuhr J. 2004. Molecular biology of severe acute respiratory syndrome coronavirus. *Curr. Opin. Microbiol.* 7:412–419. <http://dx.doi.org/10.1016/j.mib.2004.06.007>.
- Herrewegh AA, Smeenk I, Horzinek MC, Rottier PJ, de Groot RJ. 1998. Feline coronavirus type II strains 79-1683 and 79-1146 originate from a double recombination between feline coronavirus type I and canine coronavirus. *J. Virol.* 72:4508–4514.
- Woo PC, Lau SK, Yip CC, Huang Y, Tsoi HW, Chan KH, Yuen KY. 2006. Comparative analysis of 22 coronavirus HKU1 genomes reveals a novel genotype and evidence of natural recombination in coronavirus HKU1. *J. Virol.* 80:7136–7145. <http://dx.doi.org/10.1128/JVI.00509-06>.
- Cheng VC, Lau SK, Woo PC, Yuen KY. 2007. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. *Clin. Microbiol. Rev.* 20:660–694. <http://dx.doi.org/10.1128/CMR.00023-07>.
- Guan Y, Zheng BJ, He YQ, Liu XL, Zhuang ZX, Cheung CL, Luo SW, Li PH, Zhang LJ, Guan YJ, Butt KM, Wong KL, Chan KW, Lim W, Shortridge KF, Yuen KY, Peiris JS, Poon LL. 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* 302:276–278. <http://dx.doi.org/10.1126/science.1087139>.
- Marra MA, Jones SJ, Astell CR, Holt RA, Brooks-Wilson A, Butterfield YS, Khattar J, Asano JK, Barber SA, Chan SY, Cloutier A, Coughlin SM, Freeman D, Girn N, Griffith OL, Leach SR, Mayo M, McDonald H, Montgomery SB, Pandoh PK, Petrescu AS, Robertson AG, Schein JE, Siddiqui A, Smailus DE, Stott JM, Yang GS, Plummer F, Andonov A, Artsob H, Bastien N, Bernard K, Booth TF, Bowness D, Czub M, Drebot M, Fernando L, Flick R, Garbutt M, Gray M, Grolla A, Jones S, Feldmann H, Meyers A, Kabani A, Li Y, Normand S, Stroher U, Tipples GA, Tyler S, et al. 2003. The genome sequence of the SARS-associated coronavirus. *Science* 300:1399–1404. <http://dx.doi.org/10.1126/science.1085953>.
- Rota PA, Oberste MS, Monroe SS, Nix WA, Campagnoli R, Icenogle JP, Penaranda S, Bankamp B, Maher K, Chen MH, Tong S, Tamin A, Lowe L, Frace M, DeRisi JL, Chen Q, Wang D, Erdman DD, Peret TC, Burns C, Ksiazek TG, Rollin PE, Sanchez A, Liffick S, Holloway B, Limor J, McCaustland K, Olsen-Rasmussen M, Fouchier R, Gunther S, Osterhaus AD, Drosten C, Pallansch MA, Anderson LJ, Bellini WJ. 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. *Science* 300:1394–1399. <http://dx.doi.org/10.1126/science.1085952>.
- Snijder EJ, Bredenbeek PJ, Dobbe JC, Thiel V, Ziebuhr J, Poon LL, Guan Y, Rozanov M, Spaan WJ, Gorbalenya AE. 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* 331:991–1004. [http://dx.doi.org/10.1016/S0022-2836\(03\)00865-9](http://dx.doi.org/10.1016/S0022-2836(03)00865-9).
- Woo PC, Lau SK, Tsoi HW, Chan KH, Wong BH, Che XY, Tam VK, Tam SC, Cheng VC, Hung IF, Wong SS, Zheng BJ, Guan Y, Yuen KY. 2004. Relative rates of non-pneumonic SARS coronavirus infection and SARS coronavirus pneumonia. *Lancet* 363:841–845. [http://dx.doi.org/10.1016/S0140-6736\(04\)15729-2](http://dx.doi.org/10.1016/S0140-6736(04)15729-2).
- Fouchier RA, Hartwig NG, Bestebroer TM, Niemeyer B, de Jong JC, Simon JH, Osterhaus AD. 2004. A previously undescribed coronavirus associated with respiratory disease in humans. *Proc. Natl. Acad. Sci. U. S. A.* 101:6212–6216. <http://dx.doi.org/10.1073/pnas.0400762101>.
- van der Hoek L, Pyrc K, Jebbink ME, Vermeulen-Oost W, Berkhout RJ, Wolthers KC, Wertheim-van Dillen PM, Kaandorp J, Spaargaren J, Berkhout B. 2004. Identification of a new human coronavirus. *Nat. Med.* 10:368–373. <http://dx.doi.org/10.1038/nm1024>.
- Lau SK, Woo PC, Yip CC, Tse H, Tsoi HW, Cheng VC, Lee P, Tang BS, Cheung CH, Lee RA, So LY, Lau YL, Chan KH, Yuen KY. 2006. Coronavirus HKU1 and other coronavirus infections in Hong Kong. *J. Clin. Microbiol.* 44:2063–2071. <http://dx.doi.org/10.1128/JCM.02614-05>.
- Woo PC, Lau SK, Chu CM, Chan KH, Tsoi HW, Huang Y, Wong BH, Poon RW, Cai JJ, Luk WK, Poon LL, Wong SS, Guan Y, Peiris JS, Yuen KY. 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J. Virol.* 79:884–895. <http://dx.doi.org/10.1128/JVI.79.2.884-895.2005>.
- Woo PC, Lau SK, Tsoi HW, Huang Y, Poon RW, Chu CM, Lee RA, Luk WK, Wong GK, Wong BH, Cheng VC, Tang BS, Wu AK, Yung RW, Chen H, Guan Y, Chan KH, Yuen KY. 2005. Clinical and molecular epidemiological features of coronavirus HKU1-associated community-acquired pneumonia. *J. Infect. Dis.* 192:1898–1907. <http://dx.doi.org/10.1086/497151>.
- Lau SK, Woo PC, Li KS, Huang Y, Tsoi HW, Wong BH, Wong SS, Leung SY, Chan KH, Yuen KY. 2005. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. *Proc. Natl. Acad. Sci. U. S. A.* 102:14040–14045. <http://dx.doi.org/10.1073/pnas.0506735102>.
- Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, Wang H, Cramer G, Hu Z, Zhang H, Zhang J, McEachern J, Field H, Daszak P, Eaton BT, Zhang S, Wang LF. 2005. Bats are natural reservoirs of SARS-like coronaviruses. *Science* 310:676–679. <http://dx.doi.org/10.1126/science.1118391>.
- Woo PC, Lau SK, Li KS, Poon RW, Wong BH, Tsoi HW, Yip BC, Huang Y, Chan KH, Yuen KY. 2006. Molecular diversity of coronaviruses in bats. *Virology* 351:180–187. <http://dx.doi.org/10.1016/j.virol.2006.02.041>.
- Woo PC, Wang M, Lau SK, Xu H, Poon RW, Guo R, Wong BH, Gao K, Tsoi HW, Huang Y, Li KS, Lam CS, Chan KH, Zheng BJ, Yuen KY. 2007. Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. *J. Virol.* 81:1574–1585. <http://dx.doi.org/10.1128/JVI.02182-06>.
- Lau SK, Woo PC, Li KS, Huang Y, Wang M, Lam CS, Xu H, Guo R, Chan KH, Zheng BJ, Yuen KY. 2007. Complete genome sequence of bat coronavirus HKU2 from Chinese horseshoe bats revealed a much smaller spike gene with a different evolutionary lineage from the rest of the genome. *Virology* 367:428–439. <http://dx.doi.org/10.1016/j.virol.2007.06.009>.
- Lau SK, Li KS, Huang Y, Shek CT, Tse H, Wang M, Choi GK, Xu H, Lam CS, Guo R, Chan KH, Zheng BJ, Woo PC, Yuen KY. 2010. Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related *Rhinolophus* bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. *J. Virol.* 84:2808–2819. <http://dx.doi.org/10.1128/JVI.02219-09>.
- Lau SK, Poon RW, Wong BH, Wang M, Huang Y, Xu H, Guo R, Li KS, Gao K, Chan KH, Zheng BJ, Woo PC, Yuen KY. 2010. Coexistence of different genotypes in the same bat and serological characterization of *Rousettus* bat coronavirus HKU9 belonging to a novel Betacoronavirus subgroup. *J. Virol.* 84:11385–11394. <http://dx.doi.org/10.1128/JVI.01121-10>.
- Lau SK, Li KS, Tsang AK, Shek CT, Wang M, Choi GK, Guo R, Wong BH, Poon RW, Lam CS, Wang SY, Fan RY, Chan KH, Zheng BJ, Woo PC, Yuen KY. 2012. Recent transmission of a novel alphacoronavirus, bat coronavirus HKU10, from Leschenault's rousettes to pomona leaf-nosed bats: first evidence of interspecies transmission of coronavirus between bats of different suborders. *J. Virol.* 86:11906–11918. <http://dx.doi.org/10.1128/JVI.01305-12>.
- Woo PC, Lau SK, Lam CS, Lai KK, Huang Y, Lee P, Luk GS, Dyrting KC, Chan KH, Yuen KY. 2009. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. *J. Virol.* 83:908–917. <http://dx.doi.org/10.1128/JVI.01977-08>.
- 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. <http://dx.doi.org/10.1128/JVI.06540-11>.
- Lau SK, Woo PC, Yip CC, Fan RY, Huang Y, Wang M, Guo R, Lam CS, Tsang AK, Lai KK, Chan KH, Che XY, Zheng BJ, Yuen KY. 2012. Isolation and characterization of a novel Betacoronavirus subgroup A coronavirus, rabbit coronavirus HKU14, from domestic rabbits. *J. Virol.* 86:5481–5496. <http://dx.doi.org/10.1128/JVI.06927-11>.
- Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA. 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* 367:1814–1820. <http://dx.doi.org/10.1056/NEJMoa1211721>.
- de Groot RJ, Baker SC, Baric RS, Brown CS, Drosten C, Enjuanes L, Fouchier RA, Galiano M, Gorbalenya AE, Memish ZA, Perlman S, Poon LL, Snijder EJ, Stephens GM, Woo PC, Zaki AM, Zambon M, Ziebuhr

- J. 2013. Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *J. Virol.* 87:7790–7792. <http://dx.doi.org/10.1128/JVI.01244-13>.
30. Lau SK, Li KS, Tsang AK, Lam CS, Ahmed S, Chen H, Chan KH, Woo PC, Yuen KY. 2013. Genetic characterization of *Betacoronavirus* lineage C viruses in bats reveals marked sequence divergence in the spike protein of *Pipistrellus* bat coronavirus HKU5 in Japanese pipistrelle: implications for the origin of the novel Middle East respiratory syndrome coronavirus. *J. Virol.* 87:8638–8650. <http://dx.doi.org/10.1128/JVI.01055-13>.
31. Mihindukulasuriya KA, Wu G, St Leger J, Nordhausen RW, Wang D. 2008. Identification of a novel coronavirus from a beluga whale by using a panviral microarray. *J. Virol.* 82:5084–5088. <http://dx.doi.org/10.1128/JVI.02722-07>.
32. Huang Y, Lau SK, Woo PC, Yuen KY. 2008. CoVDB: a comprehensive database for comparative analysis of coronavirus genes and genomes. *Nucleic Acids Res.* 36:D504–D511. <http://dx.doi.org/10.1093/nar/gkm754>.
33. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28:2731–2739. <http://dx.doi.org/10.1093/molbev/msr121>.
34. Apweiler R, Attwood TK, Bairoch A, Bateman A, Birney E, Biswas M, Bucher P, Cerutti L, Corpet F, Croning MD, Durbin R, Falquet L, Fleischmann W, Gouzy J, Hermjakob H, Hulo N, Jonassen I, Kahn D, Kanapin A, Karavidopoulou Y, Lopez R, Marx B, Mulder NJ, Oinn TM, Pagni M, Servant F, Sigrist CJ, Zdobnov EM. 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* 29:37–40. <http://dx.doi.org/10.1093/nar/29.1.37>.
35. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. 2012. The Pfam protein families database. *Nucleic Acids Res.* 40:D290–D301. <http://dx.doi.org/10.1093/nar/gkr1065>.
36. Sonnhammer EL, von Heijne G, Krogh A. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* 6:175–182.
37. Guindon S, Lethiec F, Duroux P, Gascuel O. 2005. PHYML online—a web server for fast maximum likelihood-based phylogenetic inference. *Nucleic Acids Res.* 33:W557–W559. <http://dx.doi.org/10.1093/nar/gki352>.
38. de Groot RJ, Baker SC, Baric R, Enjuanes L, Gorbalenya A, Holmes KV, Perlman S, Poon L, Rottier PJ, Talbot PJ, Woo PC, Ziebuhr J. 2011. *Coronaviridae*, p 806–828. In King AMQ, Adams MJ, Carstens EB, Lefkowitz EJ (ed), *Virus taxonomy, classification and nomenclature of viruses*. Ninth report of the International Committee on Taxonomy of Viruses. Elsevier Academic Press, San Diego, CA.
39. Pond SL, Frost SD. 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* 21:2531–2533. <http://dx.doi.org/10.1093/bioinformatics/bti320>.