JOURNAL OF VIROLOGY, Feb. 2007, p. 1574–1585 Vol. 81, No. 4
0022-538X/07/$08.00+0 doi:10.1128/JVI.02182-06

Comparative Analysis of Twelve Genomes of Three Novel Group 2c and Group 2d Coronaviruses Reveals Unique Group and Subgroup Features▽

Patrick C. Y. Woo,[1,2,3]† Ming Wang,[4]† Susanna K. P. Lau,[1,2,3]† Huifang Xu,[4] Rosana W. S. Poon,[1] Rongtong Guo,[4] Beatrice H. L. Wong,[1] Kai Gao,[4] Hoi-wah Tsoi,[1] Yi Huang,[1] Kenneth S. M. Li,[1] Carol S. F. Lam,[1] Kwok-hung Chan,[1] Bo-jian Zheng,[1,2,3] and Kwok-yung Yuen[1,2,3]*

Department of Microbiology,[1] Research Centre of Infection and Immunology,[2] and State Key Laboratory of Emerging Infectious Diseases,[3] The University of Hong Kong, Hong Kong, and Guangzhou Center for Disease Control and Prevention, Guangzhou, China[4]

Received 5 October 2006/Accepted 13 November 2006

Twelve complete genomes of three novel coronaviruses—bat coronavirus HKU4 (bat-CoV HKU4), bat-CoV HKU5 (putative group 2c), and bat-CoV HKU9 (putative group 2d)—were sequenced. Comparative genome analysis showed that the various open reading frames (ORFs) of the genomes of the three coronaviruses had significantly higher amino acid identities to those of other group 2 coronaviruses than group 1 and 3 coronaviruses. Phylogenetic trees constructed using chymotrypsin-like protease, RNA-dependent RNA polymerase, helicase, spike, and nucleocapsid all showed that the group 2a and 2b and putative group 2c and 2d coronaviruses are more closely related to each other than to group 1 and 3 coronaviruses. Unique genomic features distinguishing between these four subgroups, including the number of papain-like proteases, the presence or absence of hemagglutinin esterase, small ORFs between the membrane and nucleocapsid genes and ORFs (NS7a and NS7b), bulged stem-loop and pseudoknot structures downstream of the nucleocapsid gene, transcription regulatory sequence, and ribosomal recognition signal for the envelope gene, were also observed. This is the first time that NS7a and NS7b downstream of the nucleocapsid gene has been found in a group 2 coronavirus. The high Ka/Ks ratio of NS7a and NS7b in bat-CoV HKU9 implies that these two group 2d-specific genes are under high selective pressure and hence are rapidly evolving. The four subgroups of group 2 coronaviruses probably originated from a common ancestor. Further molecular epidemiological studies on coronaviruses in the bats of other countries, as well as in other animals, and complete genome sequencing will shed more light on coronavirus diversity and their evolutionary histories.

Coronaviruses are found in a wide variety of animals and can cause respiratory, enteric, hepatic, and neurological diseases of varying severity. Based on genotypic and serological characterization, coronaviruses were divided into three distinct groups (3, 12, 36). As a result of the unique mechanism of viral replication, coronaviruses have a high frequency of recombination (12). Their tendency for recombination and high mutation rates may allow them to adapt to new hosts and ecological niches (8, 33).

The recent severe acute respiratory syndrome (SARS) epidemic, the discovery of SARS coronavirus (SARS-CoV), and identification of SARS-CoV-like viruses from Himalayan palm civets and a raccoon dog from wild live markets in China have boosted interest in the discovery of novel coronaviruses in both humans and animals (6, 17, 19, 21, 31). In 2004, a novel group 1 human coronavirus, human coronavirus NL63 (HCoV-NL63), was reported independently by two groups (5, 27). In 2005, we described the discovery, complete genome sequence, clinical features, and molecular epidemiology of another novel group 2 human coronavirus, coronavirus HKU1 (CoV-HKU1) (14, 29, 32). Recently, we have also described the discovery of SARS-CoV-like virus in Chinese horseshoe bats and a novel group 1 coronavirus in large bent-winged bats, lesser bent-winged bats, and Japanese long-winged bats in Hong Kong (13, 20). SARS-CoV-like viruses have also been identified in horseshoe bats in other provinces of China (15). Based on these findings, a territory-wide molecular surveillance study was conducted to examine the diversity of coronaviruses in bats of our locality, and in this search six novel coronavirus species were discovered (30). From phylogenetic analysis of the RNA-dependent RNA polymerase (*pol*) and helicase genes, two of the viruses, bat coronavirus HKU4 (bat-CoV HKU4) and bat coronavirus HKU5 (bat-CoV HKU5), seemed to form a distinct subgroup in group 2 coronavirus.

In the present study, we extended our survey to include specimens of bats in the Guangdong province of Southern China where the SARS epidemic originated and wet-markets and game food restaurants serving bat dishes are commonly found (34). Five different coronaviruses were identified, including two previously undescribed coronavirus species: bat coronavirus HKU9 (bat-CoV HKU9) and bat coronavirus HKU10 (bat-CoV HKU10). In addition, we sequenced four complete genomes each of the two putative group 2c coronaviruses (bat-CoV HKU4 and bat-CoV HKU5) we discovered in Hong Kong (30) and the putative group 2d coronavirus

* Corresponding author. Mailing address: State Key Laboratory of Emerging Infectious Diseases, Department of Microbiology, The University of Hong Kong, University Pathology Building, Queen Mary Hospital, Hong Kong. Phone: (852) 28554892. Fax: (852) 28551241. E-mail: hkumicro@hkucc.hku.hk.

† P.C.Y.W., M.W., and S.K.P.L. contributed equally to this study.

▽ Published ahead of print on 22 November 2006.

TABLE 1. Bat species captured and associated coronaviruses in the present surveillance study

Bat				Coronavirus(es) (*n*)[a]
Scientific name	Common name	No. of bats tested	No. (%) of bats positive for coronaviruses	
Hipposideros larvatus	Intermediate roundleaf bat	2	0 (0)	
Hipposideros armiger	Great roundleaf bat	26	0 (0)	
Hipposideros pomona	Pomona roundleaf bat	1	0 (0)	
Miniopterus magnater	Greater bent-winged bat	14	0 (0)	
Miniopterus pusillus	Lesser bent-winged bat	13	2 (15)	Bat-CoV HKU8
Myotis ricketti	Rickett's big-footed bat	1	0 (0)	
Rhinolophus osgoodi	Osgood's horseshoe bat	1	0 (0)	
Rhinolophus pusillus	Least horseshoe bat	12	0 (0)	
Rhinolophus affinus	Intermediate horseshoe bat	25	0 (0)	
Rhinolophus sinicus	Chinese horseshoe bat	64	7 (11)	Bat-CoV HKU2 (6), Bat-SARS-CoV HKU3 (1)
Rousettus lechenaulti	Leschenault's rousette	350	43 (12%)	Bat-CoV HKU9 (42), Bat-CoV HKU10 (1)

[a] *n*, number of bats positive for indicated virus.

(bat-CoV HKU9) discovered in the present study and compared the 12 genomes with those of other coronaviruses. Based on the results of the present study, we propose two novel subgroups, group 2c and group 2d, among group 2 coronaviruses.

MATERIALS AND METHODS

Sample collection. A total of 509 bats (11 different species) were captured from various locations in the Guangdong province of Southern China over a 7-month period (October 2005 to April 2006). Respiratory and alimentary specimens were collected by procedures described previously (13, 35).

RNA extraction. Viral RNA was extracted from the respiratory and alimentary specimens by using QIAamp viral RNA minikit (QIAGEN, Hilden, Germany). The RNA was eluted in 50 μl of AVE buffer and was used as the template for reverse transcription-PCR (RT-PCR).

RT-PCR of *pol* gene of coronaviruses using conserved primers and DNA sequencing. Coronavirus screening was performed by amplifying a 440-bp fragment of the *pol* gene of coronaviruses using the conserved primers (5′-GGTTG GGACTATCCTAAGTGTGA-3′ and 5′-CCATCATCAGATAGAATCATCA TA-3′) designed by multiple alignments of the nucleotide sequences of available *pol* genes of known coronaviruses (29). RT was performed by 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 $MgCl_2$, and 0.01% gelatin), 200 μM concentrations of each deoxynucleoside triphosphate, and 1.0 U of *Taq* polymerase (Applied Biosystems, Foster City, CA). The mixtures were amplified in 60 cycles of 94°C for 1 min, 48°C for 1 min, and 72°C for 1 min and a final extension at 72°C for 10 min in an automated thermal cycler (Applied Biosystems). Standard precautions were taken to avoid PCR contamination, and no false-positive was observed in negative controls.

The PCR products were gel purified by using a QIAquick gel extraction kit (QIAGEN). Both strands of the PCR products were sequenced twice with an ABI Prism 3700 DNA analyzer (Applied Biosystems) using the two PCR primers. The sequences of the PCR products were compared to known sequences of the *pol* genes of coronaviruses in the GenBank database.

Viral culture. Two of the samples positive for bat-CoV HKU9 and the sample positive for bat-CoV HKU10 were cultured in LLC-Mk2 (rhesus monkey kidney), MRC-5 (human lung fibroblast), FRhK-4 (rhesus monkey kidney), Huh-7.5 (human hepatoma), Vero E6 (African green monkey kidney), and HRT-18 (colorectal adenocarcinoma) cells.

Complete genome sequencing. Twelve complete genomes of bat-CoV HKU4 (30), bat-CoV HKU5 (30), and the novel bat coronavirus discovered in the present study (bat-CoV HKU9) were amplified and sequenced using the RNA extracted from the alimentary specimens as templates. The RNA was converted to cDNA by a combined random-priming and oligo(dT) priming strategy. Since the initial results revealed that these coronaviruses were group 2 coronaviruses, the cDNA was amplified by degenerate primers designed by multiple alignment of the genomes of CoV-HKU1 (GenBank accession no. NC_006577), murine hepatitis virus (GenBank accession no. NC_006852), human coronavirus OC43 (GenBank accession no. NC_005147), bovine coronavirus (GenBank accession no. NC_003045), rat sialodacryoadenitis coronavirus (GenBank accession no. AF207551), equine coronavirus NC99 (GenBank accession no. AY316300), porcine hemagglutinating encephalomyelitis virus (GenBank accession no. NC_007732), SARS-CoV (GenBank accession no. NC_004718), and bat-SARS-CoV HKU3 (GenBank accession no. DQ022305) and additional primers designed from the results of the first and subsequent rounds of sequencing. These primer sequences are available on request. The 5′ ends of the viral genomes were confirmed by rapid amplification of cDNA ends using a 5′/3′ RACE kit (Roche, 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 coronaviruses. Phylogenetic tree construction was performed by using the neighbor-joining method with CLUSTAL X 1.83. Protein family analysis was performed by using PFAM and InterProScan (1, 2). Prediction of transmembrane domains was performed by using TMpred and TMHMM (9, 23).

Estimation of synonymous and nonsynonymous substitution rates. The number of synonymous substitutions per synonymous site (Ks) and the number of nonsynonymous substitutions per nonsynonymous site (Ka) for each coding region between each pair of strains were calculated by using the Nei-Gojobori method (Jukes-Cantor) in MEGA 3.1 (11). Since the sequences of three of the four genomes of bat-CoV HKU4 are almost identical and the sequences of three of the four genomes of bat-CoV HKU5 are almost identical, the Ka/Ks ratios for the coding regions in bat-CoV HKU4 and bat-CoV HKU5 were each calculated using one of these three genomes and the remaining genome that possessed more differences. For the four strains of bat-CoV HKU9, six pairwise comparisons were performed for each coding region.

Nucleotide sequence accession numbers. The nucleotide sequences of the 12 genomes of bat-CoV HKU4, bat-CoV HKU5, and bat-CoV HKU9 have been submitted to the GenBank sequence database under accession numbers EF065505 to EF065516.

RESULTS

Bat surveillance and identification of two novel coronaviruses. A total of 1,018 respiratory and alimentary specimens from 509 bats of 11 different species were obtained in the Guangdong province in Southern China (Table 1). RT-PCR analyses for a 440-bp fragment in the *pol* genes of coronaviruses were positive in alimentary specimens from 52 (10.2%) and in a respiratory specimen from 1 (0.2%) of 509 bats. Sequencing results suggested the presence of five different coronaviruses (Table 1 and Fig. 1). The sequences of two samples from lesser bent-winged bat (*Miniopterus pusillus*) possessed >97% nucleotide identities to a group 1 coronavirus (bat-CoV HKU8) that we described recently from lesser bent-

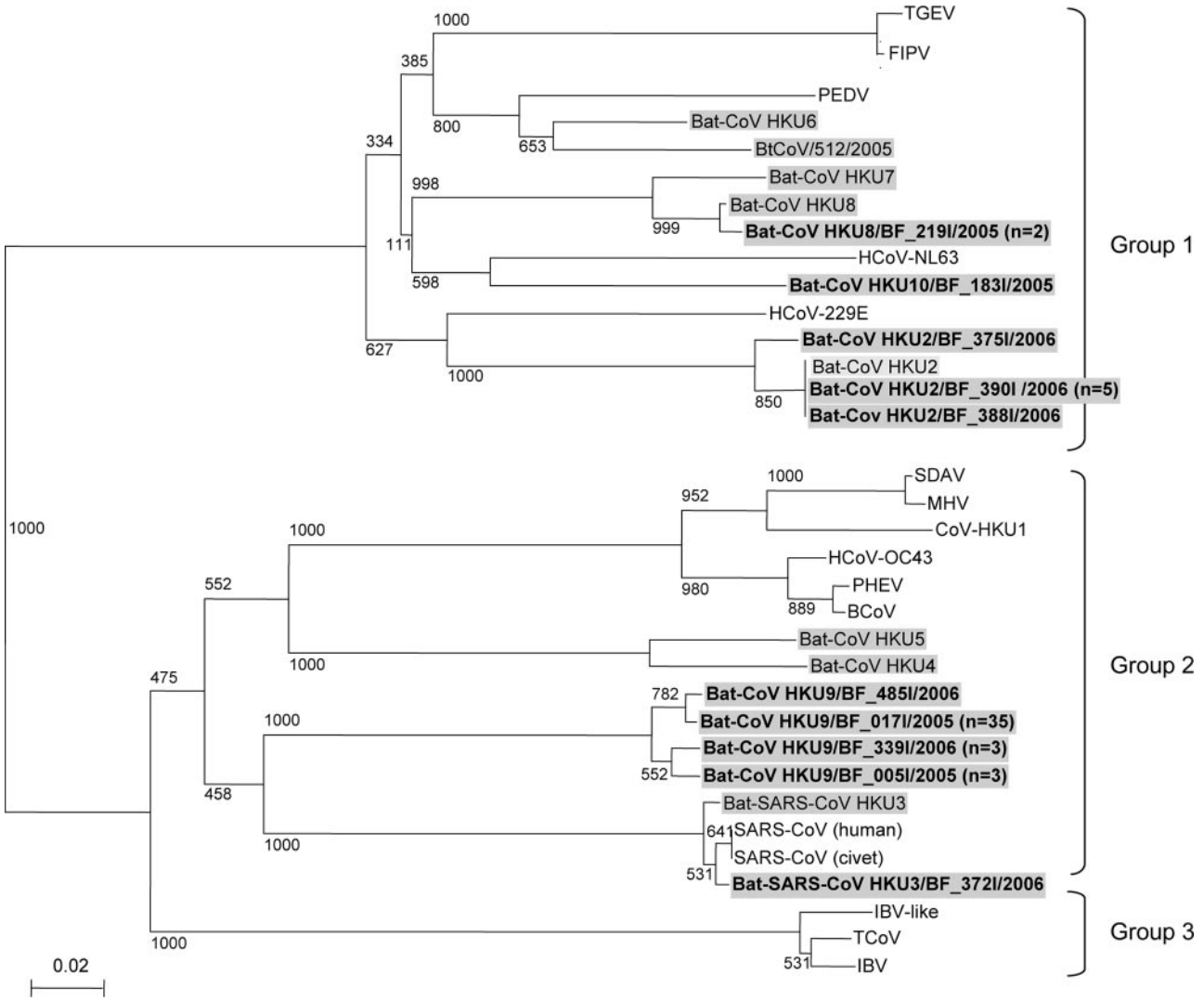

FIG. 1. Phylogenetic analysis of amino acid sequences of the 393-bp fragment of RNA-dependent RNA polymerase of coronaviruses identified from bats in the present study. The tree was constructed by the neighbor-joining method using the Jukes-Cantor correction and bootstrap values calculated from 1,000 trees. The scale bar indicates the estimated number of substitutions per 50 amino acids. Coronaviruses identified in the present study are shown in boldface. Coronaviruses from bats are shaded in gray. HCoV-229E (NC_002645); PEDV, porcine epidemic diarrhea virus (NC_003436); TGEV(NC_002306); FIPV (AY994055); HCoV-NL63 NL63 (NC_005831); bat-CoV HKU2 (DQ249235), HKU4 (DQ074652), HKU5 (DQ249219), HKU6 (DQ249224), HKU7 (DQ249226), and HKU8 (DQ249228); CoV-HKU1 (NC_006577); HCoV-OC43 (NC_005147); MHV, murine hepatitis virus (NC_006852); BCoV, bovine coronavirus (NC_003045); PHEV, porcine hemagglutinating encephalomyelitis virus (NC_007732); SDAV; SARS-CoV (human), human SARS coronavirus (NC_004718); SARS-CoV (Civet), civet SARS-like coronavirus (AY304488); bat-SARS-CoV HKU3, bat-SARS-like coronavirus HKU3 (DQ022305); IBV, infectious bronchitis virus (NC_001451); TCoV, turkey coronavirus (AF124991); IBV-like, IBV isolated from peafowl (AY641576). Other abbreviations are as defined in the text.

winged bats in Hong Kong (30), those of six alimentary specimens and one respiratory specimen (obtained from one of the six bats with positive alimentary specimens) from Chinese horseshoe bat (*Rhinolophus sinicus*) possessed >97% nucleotide identities to another group 1 coronavirus (bat-CoV HKU2) that we described recently from Chinese horseshoe bats in Hong Kong (30), and that of one sample from a Chinese horseshoe bat (*Rhinolophus sinicus*) possessed >98% nucleotide identities to bat-SARS-CoV HKU3 that we described recently from Chinese horseshoe bats in Hong Kong (13). The sequences of 42 samples from Leschenault's rousette bats (*Rousettus lechenaulti*) had <70% nucleotide identities to all known coronaviruses, suggesting a novel group 2 coronavirus (bat-CoV HKU9); that of one sample from a Leschenault's rousette bat (*Rousettus lechenaulti*) had <80% nucleotide identities to all known coronaviruses, suggesting a novel group 1 coronavirus (bat-CoV HKU10).

Viral culture. No cytopathic effect was observed in any of the cell lines inoculated with bat specimens positive for bat-CoV HKU9 and bat-CoV HKU10. Quantitative RT-PCR using the culture supernatants and cell lysates for monitoring the presence of viral replication also showed negative results.

Genome organization and coding potential of bat-CoV HKU4, bat-CoV HKU5, and bat-CoV HKU9. Since analysis of

TABLE 2. Comparison of genomic features of bat-CoV-HKU4, bat-CoV HKU5, bat-CoV HKU9, and other coronaviruses and amino acid identities between the predicted chymotrypsin-like protease ($3CL^{pro}$), RNA-dependent RNA polymerase (Pol), helicase (Hel), spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins of bat-CoV-HKU4 and bat-CoV HKU5 and the corresponding proteins of other coronaviruses[a]

Coronavirus	Genome features		Pairwise amino acid identity (%)																				
			Bat-CoV HKU4							Bat-CoV HKU5							Bat-CoV HKU9						
	Size (bases)	G+C content	$3CL^{pro}$	Pol	Hel	S	E	M	N	$3CL^{pro}$	Pol	Hel	S	E	M	N	$3CL^{pro}$	Pol	Hel	S	E	M	N
Group 1																							
HCoV-229E	27,317	0.38	48.2	58.8	62.1	24.4	26.4	32.8	24.5	49.2	58.0	62.6	25.2	28.2	32.9	27.0	42.7	56.7	60.0	26.2	16.5	31.2	19.7
PEDV	28,033	0.42	46.7	60.1	61.9	24.6	20.0	37.8	24.2	48.2	59.5	62.8	23.8	20.2	35.5	23.5	44.5	59.2	59.5	23.2	15.2	33.0	24.0
TGEV	28,586	0.38	48.4	59.7	61.3	25.7	22.5	32.7	28.9	47.4	59.6	61.9	24.6	20.2	31.7	25.6	44.2	57.9	61.3	23.0	14.9	27.2	25.8
FIPV	29,355	0.38	49.0	59.7	61.1	27.3	23.9	31.1	29.5	48.0	59.8	61.8	26.3	21.6	31.3	27.4	42.2	58.2	61.1	22.2	15.1	27.0	25.6
HCoV-NL63	27,553	0.34	49.4	58.3	61.8	25.1	18.8	32.2	27.7	48.7	58.1	62.6	25.7	18.6	33.8	25.4	44.0	57.7	60.5	25.8	13.6	29.1	22.1
Group 2a																							
CoV-HKU1	29,926	0.32	51.6	67.7	66.0	31.7	26.5	44.6	31.7	52.0	68.2	65.8	30.0	24.1	44.9	31.1	47.7	66.5	65.7	29.9	24.7	39.5	27.7
HCoV-OC43	30,738	0.37	51.6	68.7	67.8	32.1	27.3	42.4	32.0	51.6	68.8	67.7	30.8	30.7	42.0	33.3	48.4	66.6	67.5	29.1	27.4	40.7	31.2
MHV	31,357	0.42	53.3	67.9	66.2	30.9	22.0	43.2	33.3	53.9	68.1	65.8	30.2	27.2	42.2	34.2	51.0	65.2	67.1	28.5	22.5	41.6	29.6
BCoV	31,028	0.37	52.0	68.6	67.5	32.2	26.7	44.6	32.2	51.6	68.7	67.3	31.2	25.6	43.5	34.2	48.4	66.5	67.5	28.6	27.4	42.7	31.2
PHEV	30,480	0.37	52.3	68.7	67.7	32.2	27.8	44.6	31.4	52.0	68.8	67.5	30.5	26.7	42.6	32.0	48.0	66.7	67.7	29.2	28.6	41.5	29.7
Group 2b																							
SARS-CoV	29,751	0.41	50.0	71.7	70.7	32.4	39.0	43.2	44.0	51.1	71.8	71.7	31.9	34.9	43.1	43.2	52.0	72.1	73.4	31.8	29.3	43.3	39.2
Bat-SARS-CoV HKU3	29,728	0.41	50.3	71.8	70.5	32.7	39.0	43.2	44.4	51.4	71.7	71.5	31.7	34.9	42.5	43.6	52.0	71.9	73.6	32.2	29.3	43.3	39.9
Group 2c																							
Bat-CoV HKU4	30,286	0.38								83.7	92.2	93.5	66.9	79.3	82.7	74.4	51.1	69.4	72.9	29.7	19.8	42.3	37.2
Bat-CoV HKU5	30,488	0.43	83.7	92.2	93.5	66.9	79.3	82.7	74.4								50.6	69.0	73.0	30.5	23.3	43.7	35.1
Group 2d																							
Bat-CoV HKU9	29,114	0.41	51.1	69.4	72.9	29.8	19.6	42.6	37.2	50.6	69.0	73.0	30.7	23.1	44.0	35.1							
Group 3																							
IBV	27,608	0.38	40.6	61.0	57.8	25.6	19.3	28.9	26.4	38.5	59.6	58.5	23.0	16.8	27.5	28.6	36.6	61.9	61.0	26.8	20.0	30.8	25.7

[a] Abbreviations are as defined in the text and figure legends.

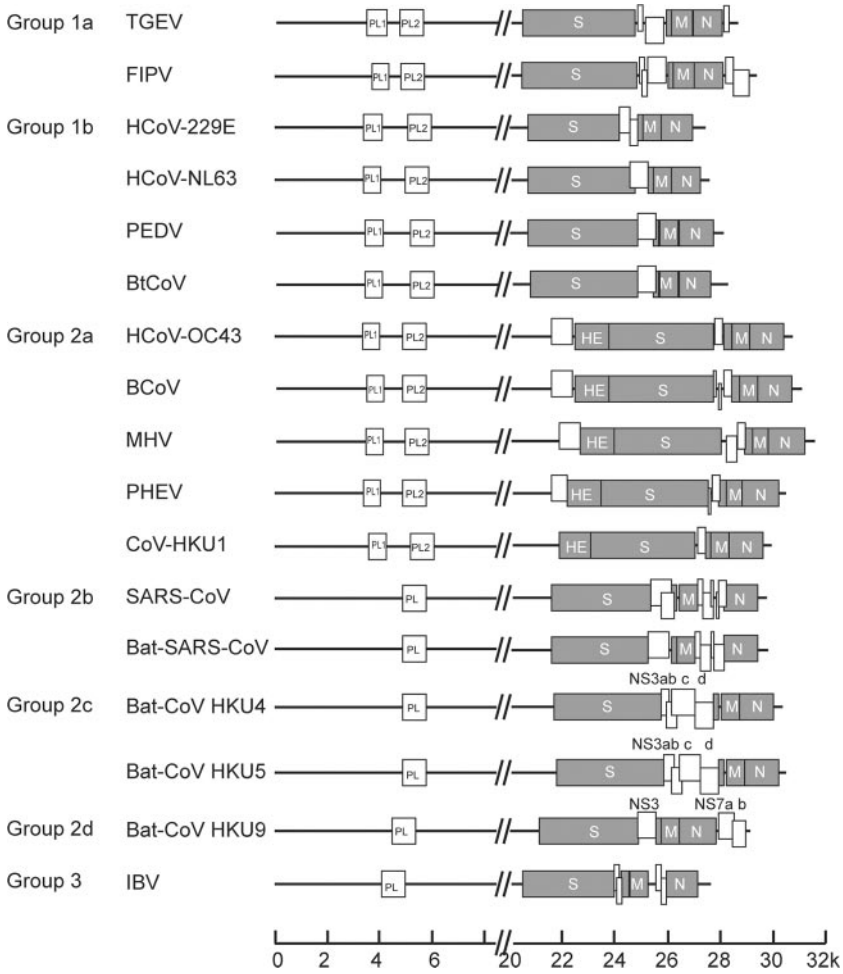

FIG. 2. Genome organizations of bat-CoV HKU4, bat-CoV HKU5, bat-CoV HKU9, and representative coronaviruses from each group. Papain-like proteases (PL1, PL2, and PL) and the nonstructural proteins are represented by white boxes. Hemagglutinin esterase (HE), spike (S), envelope (E), membrane (M), and nucleocapsid (N) are represented by gray boxes.

the 440-bp fragment of the *pol* gene of bat-CoV HKU9 suggests a distinct subgroup in group 2 coronavirus and our previous findings suggest that bat-CoV HKU4 and bat-CoV HKU5 represent another distinct subgroup of group 2 coronavirus, complete genome sequence data of four strains each of bat-CoV HKU4, bat-CoV HKU5, and bat-CoV HKU9 were obtained by assembly of the sequences of the RT-PCR products from the corresponding individual specimens.

The sizes of the genomes of bat-CoV HKU4, bat-CoV HKU5, and bat-CoV HKU9 are 30,286 to 30,316 bases, 30,482 to 30,488 bases, and 29,017 to 29,155 bases, respectively, and their G+C contents are 38, 43, and 41% (Table 2). Their genome organizations are similar to those of other coronaviruses, with the characteristic gene order: 5′-replicase ORF1ab, spike (S), envelope (E), membrane (M), and nucleocapsid (N)-3′ (Fig. 2 and Table 3). Both 5′ and 3′ ends contain short untranslated regions. The replicase ORF1ab occupies 20.8 to 21.5 kb of the genomes (Table 3). This ORF encodes a number of putative proteins, including nsp3 (which contains the putative papain-like protease [PL^{pro}]), nsp5 (putative chymotrypsin-like protease [$3CL^{pro}$]), nsp12 (putative RNA-dependent RNA polymerase [Pol]), nsp13 (putative helicase), and other proteins of unknown functions (Table 4). These proteins are produced by proteolytic cleavage of the large replicase polyprotein by PL^{pro} and $3CL^{pro}$ at specific sites (Table 4).

Bat-CoV HKU4 and bat-CoV HKU5 have the same genome structure (Fig. 2). They also possess the same putative transcription regulatory sequence (TRS) motif, 5′-ACGAAC-3′, at the 3′ end of the leader sequence and precede each ORF except NS3c and N (Table 3). This TRS has also been shown to be the TRS for SARS-CoV (10). No TRS was observed upstream of NS3c, whereas the TRS for N is ACGAAU in all eight strains of bat-CoV HKU4 and bat-CoV HKU5. Similar to other group 2b coronaviruses, the genomes of bat-CoV HKU4 and bat-CoV HKU5 have putative PL^{pro}, which are homologous to $PL2^{pro}$ of group 1 and group 2a and PL^{pro} of group 3 coronaviruses (Fig. 3). In the genomes of bat-CoV HKU4 and bat-CoV HKU5, between S and E, four ORFs that encode putative nonstructural proteins (NS3a, NS3b, NS3c, and NS3d) were observed. A BLAST search revealed no amino acid similarities between these four putative nonstructural proteins and other known proteins, and no functional domains were identified by PFAM and InterProScan. TMHMM and TMpred analyses showed three putative transmembrane do-

TABLE 3. Coding potential and putative transcription regulatory sequences of the genomes of bat-CoV HKU4, bat-CoV HKU5, and bat-CoV HKU9

Coronavirus	ORF	Start end (nucleotide position)	No. of nucleotides	No. of amino acids	Frame	Putative TRS	
						Nucleotide position in genome	TRS sequence
Bat-CoV HKU4	1a	267–13550	13,284	4,428	+3	63	ACGAAC(198)AUG
	1b	13550–21625	8,076	2,692	+2		
	S	21570–25628	4,059	1,352	+3	21519	ACGAAC(45)AUG
	NS3a	25655–25930	276	91	+2	25636	ACGAAC(13)AUG
	NS3b	25948–26307	360	119	+1	25940	ACGAACUUAUG
	NS3c	26111–26968	858	285	+2		
	NS3d	26984–27667	684	227	+2	26976	ACGAACUUAUG
	E	27737–27985	249	82	+2	27730	ACGAACUAUG
	M	28000–28659	660	219	+1	27985	ACGAAC(9)AUG
	N	28697–29968	1,272	423	+2	28674	ACGAAU(16)AUG
Bat-CoV HKU5	1a	260–13681	13,422	4,474	+2	61	ACGAAC(193)AUG
	1b	13681–21798	8,118	2,706	+1		
	S	21725–25798	4,074	1,357	+2	21674	ACGAAC(45)AUG
	NS3a	25761–26126	366	121	+3	25807	ACGAACUUAUG
	NS3b	26139–26498	360	119	+3	26130	ACGAACUUCAUG
	NS3c	26380–27150	771	256	+1		
	NS3d	27160–27831	672	223	+1	27152	ACGAACUUAUG
	E	27909–28157	249	82	+3	27902	ACGAACUAUG
	M	28172–28834	663	220	+2	28157	ACGAAC(9)AUG
	N	28884–30167	1,284	427	+3	28861	ACGAAU(16)AUG
Bat-CoV HKU9	1a	229–12951	12,723	4,241	+1	71	ACGAAC(152)AUG
	1b	12951–21020	8,070	2,690	+3		
	S	20974–24798	3,825	1,274	+1	20926	ACGAAC(42)AUG
	NS3	24795–25457	663	220	+3	24786	ACGAACAGUAUG
	E	25457–25696	240	79	+2	25448	UCGAACUAUAAUG
	M	25689–26357	669	222	+3	25662	ACGAAC(21)AUG
	N	26419–27825	1,407	468	+1	26408	ACGAACCUAUUAUG
	NS7a	27869–28426	558	185	+2	27863	ACGAACAUG
	NS7b	28433–28882	450	149	+2	28427	ACGAACAUG

TABLE 4. Characteristics of putative nonstructural proteins of replicase in bat-CoV HKU4, bat-CoV HKU5, and bat-CoV HKU9

nsp	Putative function or domain[a]	Amino acids (first residue^{position}-last residue^{position})		
		Bat-CoV HKU4	Bat-CoV HKU5	Bat-CoV HKU9
nsp1	Unknown	M^{1}-G^{195}	M^{1}-G^{195}	M^{1}-G^{175}
nsp2	Unknown	D^{196}-G^{847}	D^{196}-G^{851}	D^{176}-G^{772}
nsp3	Putative PL^{pro} domain	M^{848}-G^{2784}	A^{852}-G^{2829}	G^{773}-G^{2609}
nsp4	Hydrophobic domain	G^{2785}-Q^{3291}	G^{2830}-Q^{3337}	G^{2610}-Q^{3103}
nsp5	$3CL^{pro}$	S^{3292}-Q^{3597}	S^{3338}-Q^{3643}	A^{3104}-Q^{3409}
nsp6	Hydrophobic domain	S^{3598}-Q^{3889}	S^{3644}-Q^{3935}	G^{3410}-Q^{3699}
nsp7	Unknown	S^{3890}-Q^{3972}	S^{3936}-Q^{4018}	S^{3700}-Q^{3782}
nsp8	Unknown	A^{3973}-Q^{4171}	A^{4019}-Q^{4217}	A^{3783}-Q^{3982}
nsp9	Unknown	N^{4172}-Q^{4281}	N^{4218}-Q^{4327}	N^{3983}-H^{4094}
nsp10	Unknown	A^{4282}-Q^{4420}	A^{4328}-Q^{4466}	A^{4095}-Q^{4233}
nsp11	Unknown (short peptide at the end of ORF1a)	S^{4421}-V^{4434}	S^{4467}-L^{4480}	A^{4234}-E^{4248}
nsp12	Pol	S^{4421}-Q^{5354}	S^{4467}-Q^{5400}	A^{4234}-Q^{5165}
nsp13	Hel	A^{5355}-Q^{5952}	A^{5401}-Q^{5998}	S^{5166}-Q^{5766}
nsp14	ExoN	S^{5953}-Q^{6475}	S^{5999}-Q^{6522}	S^{5767}-Q^{6296}
nsp15	XendoU	G^{6476}-Q^{6817}	G^{6523}-Q^{6871}	S^{6297}-Q^{6633}
nsp16	2′-O-MT	A^{6818}-L^{7119}	A^{6872}-R^{7179}	A^{6634}-V^{6930}

[a] PL^{pro}, papain-like protease; $3CL^{pro}$, chymotrypsin-like protease; Pol, RNA-dependent RNA polymerase; Hel, helicase; ExoN, 3′-to-5′ exonuclease; XendoU, poly(U)-specific endoribonuclease and 2′-O-MT, *S*-adenosylmethionine-dependent 2′-*O*-ribose methyltransferase.

mains in NS3d of bat-CoV HKU4 (residues 37 to 59, 71 to 90, and 94 to 111) and bat-CoV HKU5 (residues 32 to 54, 67 to 84, and 89 to 108). Similar to group 2a and 2b coronaviruses, 18 to 81 and 19 to 82 nucleotides downstream of the N genes (nucleotide positions 29986 to 30049 in bat-CoV HKU4 and nucleotide positions 30186 to 30249 in bat-CoV HKU5), the 3′ untranslated regions of the two genomes contain predicted bulged stem-loop structures (Fig. 4). Downstream of the bulged stem-loop structures, 77 to 126 and 78 to 129 nucleotides downstream of the N genes (nucleotide positions 30045 to 30094 in bat-CoV HKU4 and nucleotide positions 30245 to 30296 in bat-CoV HKU5), pseudoknot structures are present (Fig. 4).

For the genome of bat-CoV HKU9, similar to bat-CoV HKU4, bat-CoV HKU5, and the group 2b coronaviruses, the putative TRS motif, 5′-ACGAAC-3′, is also observed. This putative TRS is present at the 3′ end of the leader sequence and precedes each ORF except E, of which the putative TRS is UCGAAC (Table 3). Interestingly, the P1 position of the putative cleavage site by $3CL^{pro}$ at the junction between nsp9 and nsp10 is occupied by histidine instead of glutamine. This exception was also previously observed at the junction between the helicase and nsp14 in CoV-HKU1 and HCoV-NL63, where the P1 positions are also occupied by histidine instead of glutamine (26, 28). One ORF, which encodes a putative nonstructural protein (NS3), is observed between the S and E genes.

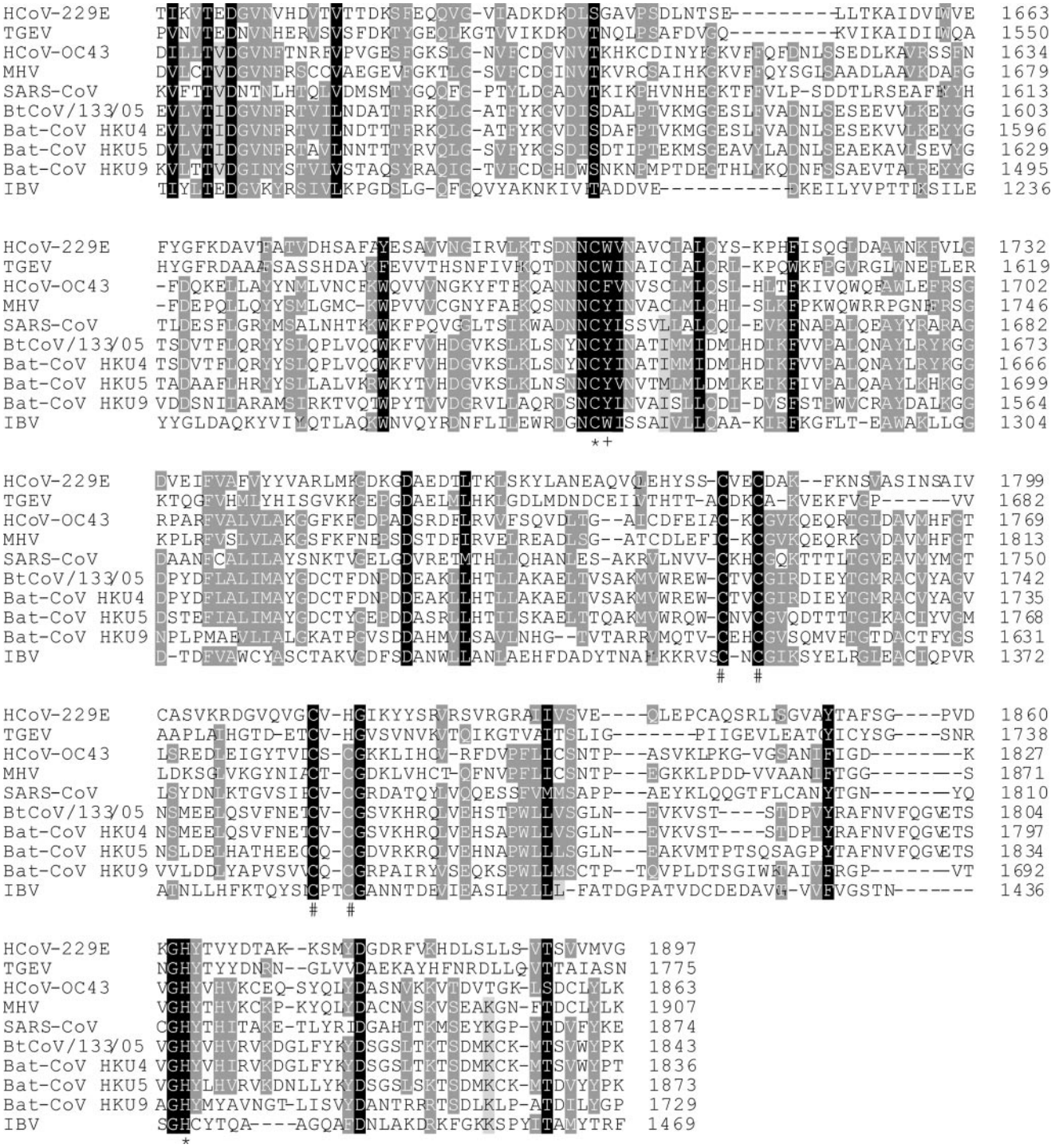

FIG. 3. Multiple alignments of PL^{pro} of SARS-CoV, btCoV/133/05 (NC_008315), bat-CoV HKU4, bat-CoV HKU5, bat-CoV HKU9, and IBV and $PL2^{pro}$ of HCoV-229E, TGEV, HCoV-OC43, and MHV. Amino acids conserved across all coronaviruses are highlighted in black. Amino acids conserved in 60 to 90% of the coronaviruses are highlighted in gray. The conserved Cys and His amino acid residues of the catalytic dyad are marked with an asterisk, the conserved postulated metal-chelating Cys and His residues are marked with a "#" symbol, and the conserved aromatic amino acid immediately downstream of the catalytic Cys is marked with a "+" symbol.

Notably, at the 3′ end of the genome, it contains the longest stretch of nucleotides (1,289 bases) after the N gene among all known coronaviruses with complete genomes available, where two ORFs that encode putative nonstructural proteins (NS7a and NS7b) are observed. A BLAST search revealed no amino acid similarities between these three putative nonstructural proteins and other known proteins,, and no functional domain was identified by PFAM and InterProScan. TMHMM and TMpred analysis showed three putative transmembrane domains in NS3 (residues 30 to 47, 54 to 76, and 80 to 99). No

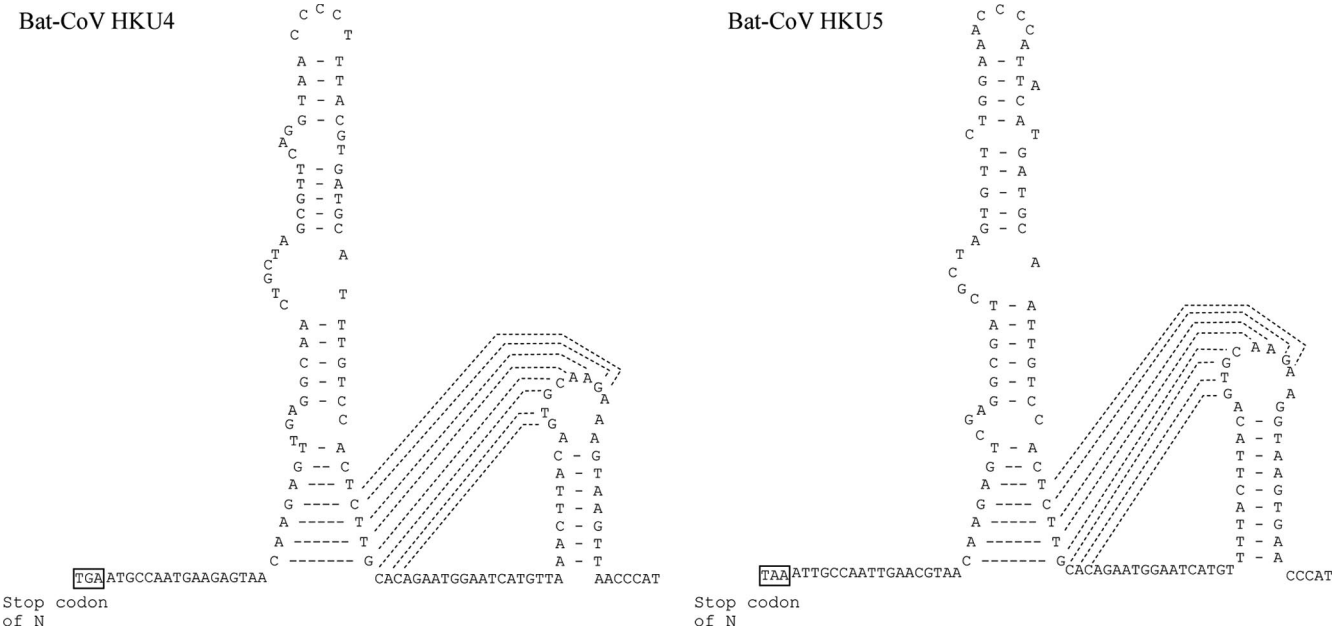

FIG. 4. Predicted bulged stem-loop and pseudoknot structures downstream of N in genomes of bat-CoV HKU4 and bat-CoV HKU5. Stop codons for the N genes are boxed. Broken lines indicate alternative base pairing.

bulged stem-loop and pseudoknot structures, similar to those in other group 2 coronaviruses, are observed downstream to N, NS7a, or NS7b in the bat-CoV HKU9 genomes.

Phylogenetic analyses. The phylogenetic trees constructed using the amino acid sequences of the $3CL^{pro}$, Pol, helicase, S, and N of bat-CoV HKU4, bat-CoV HKU5, bat-CoV HKU9, and other coronaviruses are shown in Fig. 5, and the corresponding pairwise amino acid identities are shown in Table 2. For all of the five genes, bat-CoV HKU4, bat-CoV HKU5, and bat-CoV HKU9 possess higher amino acid identities to the homologous genes in other group 2 coronaviruses than to those of group 1 and group 3 coronaviruses (Table 2). In all five trees, all strains of bat-CoV HKU4, bat-CoV HKU5, and another strain of coronavirus recently described (24) were clustered together, with bootstrap values of 1,000 in all cases, forming a distinct subgroup (Fig. 5). Within this subgroup, all four strains of bat-CoV HKU4 were clustered with the strain of coronavirus recently described (BtCoV/133/05) (24), and all four strains of bat-CoV HKU5 were clustered separately, forming two distinct sublineages. Furthermore, in all five trees, all strains of bat-CoV HKU9 were clustered together, with bootstrap values of 1,000 in all cases, forming another distinct subgroup (Fig. 5). From both phylogenetic tree analysis and amino acid differences, the strains of bat-CoV HKU9 subgroup were more closely related to the group 2b coronaviruses than the others (Fig. 5 and Table 2). We propose two novel subgroups, group 2c and group 2d, of coronavirus to describe these two distinct subgroups, respectively.

Estimation of synonymous and nonsynonymous substitution rates. The Ka/Ks ratio for the various coding regions in bat-CoV HKU4, bat-CoV HKU5, and bat-CoV HKU9 is shown in Table 5. For bat-CoV HKU4, the numbers of synonymous and nonsynonymous mutations were small. Therefore, the Ka/Ks ratios of the various coding regions, as, for example, the exceptional high Ka/Ks ratios of nsp6, NS3c and N, were not conclusive. For bat-CoV HKU5, the Ka/Ks ratios of the various coding regions were small, implying that the genes were stably evolving. Notably, the Ka/Ks ratio for NS3c of bat-CoV HKU5 is 0.027, which suggested that this gene is expressed and stably evolving. However, NS3c possesses neither TRS nor internal ribosomal entry site (IRES). Further experiments are necessary to elucidate whether NS3c is expressed and, if it is expressed, what signal sequence is involved for ribosomal recognition. For bat-CoV HKU9, the mean Ka/Ks ratio of NS7a and 7b (0.961 and 0.529) was significantly higher than those of other coding regions, implying that these two genes are rapidly evolving.

DISCUSSION

Two putative new subgroups, 2c and 2d, of coronaviruses, are described. The four strains of bat-CoV HKU4 and the four strains of bat-CoV HKU5 formed two distinct branches in the putative subgroup 2c lineage in all five phylogenetic trees analyzed (Fig. 5). Moreover, all strains of bat-CoV HKU4 were found in lesser bamboo bats, whereas all strains of bat-CoV HKU5 were found in Japanese pipistrelle (30). These findings support the view that bat-CoV HKU4 and bat-CoV HKU5 are two separate coronavirus species. Since bat-CoV HKU4 and bat-CoV HKU5 have the same genome organization and share the same TRS, we speculate that these two coronaviruses originated from the same ancestor, and their subsequent divergence into two separate species was due to the adaptation to different hosts and ecological niches. As for bat-CoV HKU9, the S and N genes showed quite marked nucleotide polymorphism and amino acid sequence changes, but the amino acid sequences of $3CL^{pro}$, Pol, and helicase are relatively conserved

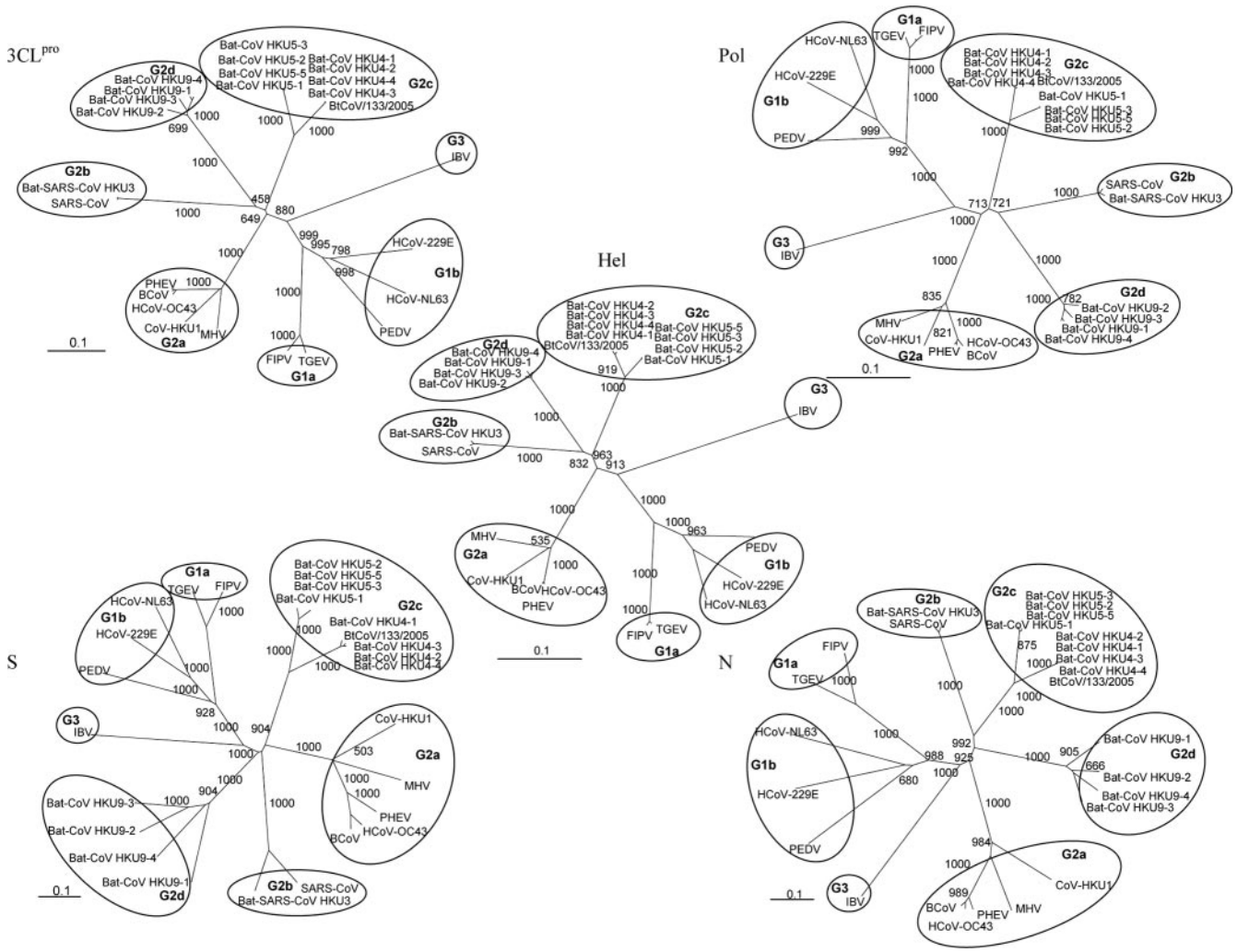

FIG. 5. Phylogenetic analysis of chymotrypsin-like protease (3CL^{pro}), RNA-dependent RNA polymerase (Pol), helicase (Hel), spike (S), and nucleocapsid (N) of bat-CoV HKU4, bat-CoV HKU5, and bat-CoV HKU9. The trees were constructed by the neighbor-joining method using the Jukes-Cantor correction and bootstrap values calculated from 1,000 trees. We included 327, 949, 609, 1,661, and 582 amino acid positions in 3CL^{pro}, Pol, helicase, S and N, respectively, in the analysis. The scale bar indicates the estimated number of substitutions per 10 amino acids. Abbreviations are as defined in the text or in the legend to Fig. 1.

(Fig. 5). Furthermore, all 42 strains of bat-CoV HKU9 were found in the same bat species, Leschenault's rousette. These findings support the view that all of the 42 strains of bat-CoV HKU9 belong to one coronavirus species. Complete genome sequencing of more bat-CoV HKU9 strains may show genotypes and even recombination events as in the case of CoV-HKU1 (33). Based on phylogenetic tree analysis, although coronaviruses of groups 2c (bat-CoV HKU4 and bat-CoV HKU5) and group 2d (bat-CoV HKU9) are more closely related to the other group 2 coronaviruses, they formed branches distinct from the group 2a and 2b coronaviruses. Furthermore, bat-CoV HKU4, bat-CoV HKU5, and bat-CoV HKU9 of these two new proposed subgroups possessed additional genomic features different from those of other group 2 coronaviruses (Table 6). For the coding potentials of the genomes, group 2a coronaviruses possess PL1^{pro} and PL2^{pro}, but group 2b, 2c, and 2d coronaviruses only possess one PL^{pro} that is homologous to PL2^{pro}. It is noteworthy that in an article recently published, the authors mentioned that no PL^{pro} was identified in nsp3 of the genome of BtCoV/133/05 (NC_008315, >95% overall nucleotide identities with bat-CoV HKU4) (24). However, after careful analysis of their nsp3 by multiple alignment and a search of the conserved domains and amino acid residues (37), it was found that PL^{pro} is present in the genome of BtCoV/133/05, with the conserved Cys and His residues of the catalytic dyad, conserved aromatic amino acid residue (Trp, Phe, or Tyr) immediately downstream to the catalytic Cys, and the postulated metal-chelating Cys and His residues of the zinc fingers (Fig. 3). The genomes of group 2a coronavirus, but not those of group 2b, 2c, and 2d coronaviruses, encode hemagglutinin esterase. The genomes of group 2b coronavirus, but not those of group 2a, 2c, and 2d coronaviruses, contain several small ORFs between the M and N genes. The genomes of group 2d coronavirus, but not those of group 2a, 2b, and 2c coronaviruses, contain two ORFs downstream of the N gene. As for the TRS, the sequence for the

TABLE 5. Estimation of nonsynonymous substitution and synonymous rates in the genomes of bat-CoV HKU4, bat-CoV HKU5, and bat-CoV HKU9

Coding region	Ka/Ks ratio		
	Bat-CoV HKU4	Bat-CoV HKU5	Bat-CoV HKU9[a]
nsp1	0.031	Ka = 0, Ks = 0.03711	0.247
nsp2	0.133	0.061	0.131
nsp3	0.154	0.070	0.091
nsp4	0.155	0.045	0.066
nsp5	Ka = 0, Ks = 0.00239	0.016	0.035
nsp6	0.317	0.076	0.067
nsp7	Ka = 0, Ks = 0.00904	0.066	0.020
nsp8	Ka = 0, Ks = 0	0.011	0.025
nsp9	Ka = 0, Ks = 0.00691	0.021	0.019
nsp10	Ka = 0, Ks = 0	0.050	0.021
nsp11	Ka = 0, Ks = 0	Ka = 0, Ks = 0	0.283
nsp12	Ka = 0, Ks = 0.00163	0.003	0.027
nsp13	Ka = 0, Ks = 0	0.009	0.011
nsp14	Ka = 0, Ks = 0	0.007	0.028
nsp15	Ka = 0, Ks = 0.00665	0.091	0.044
nsp16	Ka = 0, Ks = 0	0.018	0.081
S	0.010	0.127	0.170
NS3			0.234
NS3a	0.187	Ka = 0.00181, Ks = 0	
NS3b	0.308	0.201	
NS3c	1.205	0.027	
NS3d	Ka = 0.00096, Ks = 0	0.166	
E	Ka = 0, Ks = 0.00865	Ka = 0, Ks = 0.03392	0.108
M	Ka = 0, Ks = 0.00325	0.014	0.097
N	0.473	0.060	0.096
NS7a			0.961
NS7b			0.529

[a] Mean of six comparisons.

TABLE 6. Comparison of characteristics in the genomes of group 2a, 2b, 2c, and 2d coronaviruses

Characteristics[a]	Group 2 coronavirus			
	2a	2b	2c	2d
Coding potential				
Papain-like protease	PL1^{pro} and PL2^{pro}	PL^{pro}	PL^{pro}	PL^{pro}
Hemagglutinin esterase	+	−	−	−
Small ORFs between M and N	−	+	−	−
NS7a and 7b downstream to N	−	−	−	+
TRS				
TRS sequence	CUAAAC	ACGAAC	ACGAAC	ACGAAC
TRS/IRES for E	IRES	TRS	TRS	TRS
Stem-loop and pseudoknot structures downstream to N	+	+	+	−

[a] TRS, transcription regulatory sequence; IRES, internal ribosome entry site.

TRS of group 2a coronaviruses is CUAAAC and that of the group 2b, 2c, and 2d coronaviruses is ACGAAC (10, 12, 16). For the E gene, TRS is present in group 2b, 2c, and 2d, but not 2a, coronaviruses, which use IRES for their translation. The genomes of group 2a, 2b, and 2c coronaviruses, but not of group 2d coronaviruses, contain bulged stem-loop and pseudoknot structures downstream of the N gene.

Coronaviruses are probably better classified into group 1 (subgroups 1a and 1b), group 2 (subgroups 2a, 2b, 2c, and 2d), and group 3 than into seven groups. Traditionally, coronaviruses have been classified into groups 1, 2, and 3. When SARS-CoV was first identified and its genome was sequenced, it was proposed that it constituted a fourth group of coronavirus (17, 21). However, after more extensive phylogenetic analyses, it was suggested that SARS-CoV probably represents a distant relative of group 2 coronaviruses, and it was subsequently classified as group 2b coronaviruses (4, 22). In 2005, we and another group in mainland China independently described additional members of group 2b coronaviruses (13, 15). Recently, we described the discovery of six novel coronaviruses from bats in Hong Kong (30). Phylogenetic analysis of the *pol* and helicase genes showed that two of them, bat-CoV HKU4 and bat-CoV HKU5, probably represent a novel subgroup in group 2 coronaviruses. Subsequently, another group reported similar diversity in coronaviruses found from bats in mainland China, and they proposed that coronaviruses should be classified into five groups, instead of groups 1, 2a, 2b, 2c, and 3 (24). In the present study, we discovered another distinct subgroup of coronaviruses (bat-CoV HKU9). We also performed complete genome sequencing of four strains each of bat-CoV HKU4, bat-CoV HKU5, and bat-CoV HKU9. This large amount of genome sequence data enabled us to perform a thorough comparative analysis of the genomes of the various groups of coronaviruses. The results showed that the amino acid identities in the various ORFs among the group 2 coronaviruses were significantly higher than those between group 2 coronaviruses and the group 1 and 3 coronaviruses. Phylogenetic trees constructed using 3CL^{pro}, Pol, helicase, S, and N all showed that the group 2a, 2b, 2c, and 2d coronaviruses are more closely related to each other than the group 1 and 3 coronaviruses (Fig. 5). These showed that the group 2 coronaviruses probably originated from one common ancestor before they diverge into the four subgroups, and therefore it would be more logical and informative if they are classified as subgroups of group 2 coronaviruses.

This is the first time that NS7a and 7b downstream of the N gene has been observed in group 2 coronaviruses. Previously, feline infectious peritonitis virus (FIPV), a group 1 coronavirus, is the only coronavirus known to possess two genes downstream of the N gene (18). FIPV infects macrophages in a variety of tissues systemically, whereas feline enteric coronavirus (FECV), a coronavirus closely related to FIPV, is restricted to replication in enterocytes. It has been found that the FECV genome lacks the 300 nucleotides at the 3′ end of FIPV, suggesting that this region may be important for virulence. Recently, it has been shown that an isogenic deletion mutant of FIPV missing the 7ab cluster protected cats against lethal challenge by FIPV, which makes the mutant a potential live attenuated vaccine candidate (7). In addition to FIPV, the genome of porcine transmissible gastroenteritis virus (TGEV) also possesses one gene downstream of N (25). This gene encodes a hydrophobic protein that associates with endoplasmic reticulum and cell surface membranes in TGEV-infected cells, suggesting that it may have a role in the membrane association of replication complexes or assembly of the virus (25). In the present comparative genomic analysis, ORFs downstream of the N gene were not found in any other coronaviruses other than group 1a coronaviruses and bat-CoV HKU9 (Fig. 2). While the presence of TRS supports that NS7a and 7b of bat-CoV HKU9 are probably expressed, the high

Ka/Ks ratio implies that these two genes are under high selective pressure and thus are rapidly evolving, which may be due to recent acquisition by recombination. Further experiments will delineate the function and essentiality of NS7a and NS7b in bat-CoV HKU9.

The huge diversity of coronaviruses is probably a result of both a higher mutation rate of RNA viruses due to the infidelity of their polymerases and a higher chance of recombination as a result of their unique replication mechanism. Before the SARS epidemic in 2003, a total of 19 (2 human, 13 mammalian, and 4 avian) coronaviruses were known. Since the SARS epidemic, two novel human coronaviruses have been discovered (5, 27, 29). In the past two years, at least 10 previously unrecognized coronaviruses from bats have been described in Hong Kong and mainland China (13, 15, 20, 24, 30). In addition to the generation of a large number of coronavirus species, recombination has also resulted in the generation of different genotypes in a particular coronavirus species. This is exemplified by the presence of at least three genotypes in CoV-HKU1 as a result of recombination (33). The astonishing diversity of coronaviruses in bats implies that there are probably a lot of other unknown coronaviruses in other animal species. Further molecular epidemiological studies in bats of other countries, as well as in other animals, and complete genome sequencing will shed more light on coronavirus diversity and the evolutionary histories of these viruses.

ACKNOWLEDGMENTS

We thank Stella Hung, Chik-Chuen Lay, and Ping-Man So (HKSAR Department of Agriculture, Fisheries, and Conservation [AFCD]) and the Hong Kong Police Force for facilitation and support; Chung-Tong Shek and Cynthia S. M. Chan from the AFCD for excellent technical assistance; and King-Shun Lo (Laboratory Animal Unit) and Cassius Chan for the collection of animal specimens. We are grateful for the generous support of Hui Hoy and Hui Ming in the genomic sequencing platform.

This study is partly supported by a Research Grant Council grant; a University Development Fund and Outstanding Young Researcher Award (The University of Hong Kong); The Tung Wah Group of Hospitals Fund for Research in Infectious Diseases; the HKSAR Research Fund for the Control of Infectious Diseases of the Health, Welfare, and Food Bureau; and the Providence Foundation Limited in memory of the late Lui Hac Minh.

REFERENCES

1. **Apweiler, R., T. K. Attwood, A. Bairoch, A. Bateman, E. Birney, M. Biswas, P. Bucher, L. Cerutti, F. Corpet, M. D. Croning, R. Durbin, L. Falquet, W. Fleischmann, J. Gouzy, H. Hermjakob, N. Hulo, I. Jonassen, D. Kahn, A. Kanapin, Y. Karavidopoulou, R. Lopez, B. Marx, N. J. Mulder, T. M. Oinn, M. Pagni, F. Servant, C. J. Sigrist, and E. M. Zdobnov.** 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res. **29:**37–40.
2. **Bateman, A., E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S. R. Eddy, S. Griffiths-Jones, K. L. Howe, M. Marshall, and E. L. Sonnhammer.** 2002. The Pfam protein families database. Nucleic Acids Res. **30:**276–280.
3. **Brian, D. A., and R. S. Baric.** 2005. Coronavirus genome structure and replication. Curr. Top. Microbiol. Immunol. **287:**1–30.
4. **Eickmann, M., S. Becker, H. D. Klenk, H. W. Doerr, K. Stadler, S. Censini, S. Guidotti, V. Masignani, M. Scarselli, M. Mora, C. Donati, J. H. Han, H. C. Song, S. Abrignani, A. Covacci, and R. Rappuoli.** 2003. Phylogeny of the SARS coronavirus. Science **302:**1504–1505.
5. **Fouchier, R. A., N. G. Hartwig, T. M. Bestebroer, B. Niemeyer, J. C. de Jong, J. H. Simon, and A. D. Osterhaus.** 2004. A previously undescribed coronavirus associated with respiratory disease in humans. Proc. Natl. Acad. Sci. USA **101:**6212–6216.
6. **Guan, Y., B. J. Zheng, Y. Q. He, X. L. Liu, Z. X. Zhuang, C. L. Cheung, S. W. Luo, P. H. Li, L. J. Zhang, Y. J. Guan, K. M. Butt, K. L. Wong, K. W. Chan, W. Lim, K. F. Shortridge, K. Y. Yuen, J. S. Peiris, and L. L. Poon.** 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science **302:**276–278.
7. **Haijema, B. J., H. Volders, and P. J. Rottier.** 2004. Live, attenuated coronavirus vaccines through the directed deletion of group-specific genes provide protection against feline infectious peritonitis. J. Virol. **78:**3863–3871.
8. **Herrewegh, A. A., I. Smeenk, M. C. Horzinek, P. J. Rottier, and R. J. de Groot.** 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.
9. **Hofmann, K. S. W.** 1993. TMbase: a database of membrane spanning proteins segments. Biol. Chem. Hoppe-Seyler's **374:**166.
10. **Hussain, S., J. Pan, Y. Chen, Y. Yang, J. Xu, Y. Peng, Y. Wu, Z. Li, Y. Zhu, P. Tien, and D. Guo.** 2005. Identification of novel subgenomic RNAs and noncanonical transcription initiation signals of severe acute respiratory syndrome coronavirus. J. Virol. **79:**5288–5295.
11. **Kumar, S., K. Tamura, and M. Nei.** 2004. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform. **5:**150–163.
12. **Lai, M. M., and D. Cavanagh.** 1997. The molecular biology of coronaviruses. Adv. Virus Res. **48:**1–100.
13. **Lau, S. K., P. C. Woo, K. S. Li, Y. Huang, H. W. Tsoi, B. H. Wong, S. S. Wong, S. Y. Leung, K. H. Chan, and K. Y. Yuen.** 2005. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc. Natl. Acad. Sci. USA **102:**14040–14045.
14. **Lau, S. K., P. C. Woo, C. C. Yip, H. Tse, H. W. Tsoi, V. C. Cheng, P. Lee, B. S. Tang, C. H. Cheung, R. A. Lee, L. Y. So, Y. L. Lau, K. H. Chan, and K. Y. Yuen.** 2006. Coronavirus HKU1 and other coronavirus infections in Hong Kong. J. Clin. Microbiol. **44:**2063–2071.
15. **Li, W., Z. Shi, M. Yu, W. Ren, C. Smith, J. H. Epstein, H. Wang, G. Crameri, Z. Hu, H. Zhang, J. Zhang, J. McEachern, H. Field, P. Daszak, B. T. Eaton, S. Zhang, and L. F. Wang.** 2005. Bats are natural reservoirs of SARS-like coronaviruses. Science **310:**676–679.
16. **Makino, S., and M. Joo.** 1993. Effect of intergenic consensus sequence flanking sequences on coronavirus transcription. J. Virol. **67:**3304–3311.
17. **Marra, M. A., S. J. Jones, C. R. Astell, R. A. Holt, A. Brooks-Wilson, Y. S. Butterfield, J. Khattra, J. K. Asano, S. A. Barber, S. Y. Chan, A. Cloutier, S. M. Coughlin, D. Freeman, N. Girn, O. L. Griffith, S. R. Leach, M. Mayo, H. McDonald, S. B. Montgomery, P. K. Pandoh, A. S. Petrescu, A. G. Robertson, J. E. Schein, A. Siddiqui, D. E. Smailus, J. M. Stott, G. S. Yang, F. Plummer, A. Andonov, H. Artsob, N. Bastien, K. Bernard, T. F. Booth, D. Bowness, M. Czub, M. Drebot, L. Fernando, R. Flick, M. Garbutt, M. Gray, A. Grolla, S. Jones, H. Feldmann, A. Meyers, A. Kabani, Y. Li, S. Normand, U. Stroher, G. A. Tipples, S. Tyler, R. Vogrig, D. Ward, B. Watson, R. C. Brunham, M. Krajden, M. Petric, D. M. Skowronski, C. Upton, and R. L. Roper.** 2003. The genome sequence of the SARS-associated coronavirus. Science **300:**1399–1404.
18. **Olsen, C. W.** 1993. A review of feline infectious peritonitis virus: molecular biology, immunopathogenesis, clinical aspects, and vaccination. Vet. Microbiol. **36:**1–37.
19. **Peiris, J. S., S. T. Lai, L. L. Poon, Y. Guan, L. Y. Yam, W. Lim, J. Nicholls, W. K. Yee, W. W. Yan, M. T. Cheung, V. C. Cheng, K. H. Chan, D. N. Tsang, R. W. Yung, T. K. Ng, and K. Y. Yuen.** 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet **361:**1319–1325.
20. **Poon, L. L., D. K. Chu, K. H. Chan, O. K. Wong, T. M. Ellis, Y. H. Leung, S. K. Lau, P. C. Woo, K. Y. Suen, K. Y. Yuen, Y. Guan, and J. S. Peiris.** 2005. Identification of a novel coronavirus in bats. J. Virol. **79:**2001–2009.
21. **Rota, P. A., M. S. Oberste, S. S. Monroe, W. A. Nix, R. Campagnoli, J. P. Icenogle, S. Penaranda, B. Bankamp, K. Maher, M. H. Chen, S. Tong, A. Tamin, L. Lowe, M. Frace, J. L. DeRisi, Q. Chen, D. Wang, D. D. Erdman, T. C. Peret, C. Burns, T. G. Ksiazek, P. E. Rollin, A. Sanchez, S. Liffick, B. Holloway, J. Limor, K. McCaustland, M. Olsen-Rasmussen, R. Fouchier, S. Gunther, A. D. Osterhaus, C. Drosten, M. A. Pallansch, L. J. Anderson, and W. J. Bellini.** 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science **300:**1394–1399.
22. **Snijder, E. J., P. J. Bredenbeek, J. C. Dobbe, V. Thiel, J. Ziebuhr, L. L. Poon, Y. Guan, M. Rozanov, W. J. Spaan, and A. E. Gorbalenya.** 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.
23. **Sonnhammer, E. L., G. von Heijne, and A. Krogh.** 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol. **6:**175–182.
24. **Tang, X. C., J. X. Zhang, S. Y. Zhang, P. Wang, X. H. Fan, L. F. Li, G. Li, B. Q. Dong, W. Liu, C. L. Cheung, K. M. Xu, W. J. Song, D. Vijaykrishna, L. L. Poon, J. S. Peiris, G. J. Smith, H. Chen, and Y. Guan.** 2006. Prevalence and genetic diversity of coronaviruses in bats from China. J. Virol. **80:**7481–7490.
25. **Tung, F. Y., S. Abraham, M. Sethna, S. L. Hung, P. Sethna, B. G. Hogue, and D. A. Brian.** 1992. The 9-kDa hydrophobic protein encoded at the 3′ end of the porcine transmissible gastroenteritis coronavirus genome is membrane-associated. Virology **186:**676–683.
26. **van der Hoek, L., K. Pyrc, and B. Berkhout.** 2006. Human coronavirus NL63, a new respiratory virus. FEMS Microbiol. Rev. **30:**760–773.
27. **van der Hoek, L., K. Pyrc, M. F. Jebbink, W. Vermeulen-Oost, R. J. Berkhout, K. C. Wolthers, P. M. Wertheim-van Dillen, J. Kaandorp, J. Spaar-**

garen, and B. Berkhout. 2004. Identification of a new human coronavirus. Nat. Med. **10:**368–373.
28. **Woo, P. C., Y. Huang, S. K. Lau, H. W. Tsoi, and K. Y. Yuen.** 2005. In silico analysis of ORF1ab in coronavirus HKU1 genome reveals a unique putative cleavage site of coronavirus HKU1 3C-like protease. Microbiol. Immunol. **49:**899–908.
29. **Woo, P. C., S. K. Lau, C. M. Chu, K. H. Chan, H. W. Tsoi, Y. Huang, B. H. Wong, R. W. Poon, J. J. Cai, W. K. Luk, L. L. Poon, S. S. Wong, Y. Guan, J. S. Peiris, and K. Y. Yuen.** 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J. Virol. **79:**884–895.
30. **Woo, P. C., S. K. Lau, K. S. Li, R. W. Poon, B. H. Wong, H. W. Tsoi, B. C. Yip, Y. Huang, K. H. Chan, and K. Y. Yuen.** 2006. Molecular diversity of coronaviruses in bats. Virology **351:**180–187.
31. **Woo, P. C., S. K. Lau, H. W. Tsoi, K. H. Chan, B. H. Wong, X. Y. Che, V. K. Tam, S. C. Tam, V. C. Cheng, I. F. Hung, S. S. Wong, B. J. Zheng, Y. Guan, and K. Y. Yuen.** 2004. Relative rates of non-pneumonic SARS coronavirus infection and SARS coronavirus pneumonia. Lancet **363:**841–845.
32. **Woo, P. C., S. K. Lau, H. W. Tsoi, Y. Huang, R. W. Poon, C. M. Chu, R. A. Lee, W. K. Luk, G. K. Wong, B. H. Wong, V. C. Cheng, B. S. Tang, A. K. Wu, R. W. Yung, H. Chen, Y. Guan, K. H. Chan, and K. Y. Yuen.** 2005. Clinical and molecular epidemiological features of coronavirus HKU1-associated community-acquired pneumonia. J. Infect. Dis. **192:**1898–1907.
33. **Woo, P. C., S. K. Lau, C. C. Yip, Y. Huang, H. W. Tsoi, K. H. Chan, and K. Y. Yuen.** 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.
34. **Woo, P. C., S. K. Lau, and K. Y. Yuen.** 2006. Infectious diseases emerging from Chinese wet-markets: zoonotic origins of severe respiratory viral infections. Curr. Opin. Infect. Dis. **19:**401–407.
35. **Yob, J. M., H. Field, A. M. Rashdi, C. Morrissy, B. van der Heide, P. Rota, A. bin Adzhar, J. White, P. Daniels, A. Jamaluddin, and T. Ksiazek.** 2001. Nipah virus infection in bats (order Chiroptera) in peninsular Malaysia. Emerg. Infect. Dis. **7:**439–441.
36. **Ziebuhr, J.** 2004. Molecular biology of severe acute respiratory syndrome coronavirus. Curr. Opin. Microbiol. **7:**412–419.
37. **Ziebuhr, J., V. Thiel, and A. E. Gorbalenya.** 2001. The autocatalytic release of a putative RNA virus transcription factor from its polyprotein precursor involves two paralogous papain-like proteases that cleave the same peptide bond. J. Biol. Chem. **276:**33220–33232.