JOURNAL OF VIROLOGY, Nov. 2011, p. 11325–11337 Vol. 85, No. 21
0022-538X/11/$12.00 doi:10.1128/JVI.05512-11

Molecular Epidemiology of Human Coronavirus OC43 Reveals Evolution of Different Genotypes over Time and Recent Emergence of a Novel Genotype due to Natural Recombination▽

Susanna K. P. Lau,[1,2,3,4]* Paul Lee,[1] Alan K. L. Tsang,[1] Cyril C. Y. Yip,[1] Herman Tse,[1,2,3,4] Rodney A. Lee,[5] Lok-Yee So,[6] Yu-Lung Lau,[7] Kwok-Hung Chan,[1] Patrick C. Y. Woo,[1,2,3,4]* and Kwok-Yung Yuen[1,2,3,4]

Department of Microbiology,[1] Research Centre of Infection and Immunology,[2] State Key Laboratory of Emerging Infectious Diseases,[3] Carol Yu Centre for Infection,[4] The University of Hong Kong, Hong Kong; Department of Pathology[5] and Department of Paediatrics and Adolescent Medicine,[6] Pamela Youde Nethersole Eastern Hospital, Hong Kong; and Department of Pediatrics and Adolescent Medicine, Queen Mary Hospital, The University of Hong Kong, Hong Kong[7]

Received 25 June 2011/Accepted 8 August 2011

Although human coronavirus OC43-OC43 (HCoV-OC43) is the coronavirus most commonly associated with human infections, little is known about its molecular epidemiology and evolution. We conducted a molecular epidemiology study to investigate different genotypes and potential recombination in HCoV-OC43. Twenty-nine HCoV-OC43 strains from nasopharyngeal aspirates, collected from 2004 to 2011, were subjected to RNA-dependent RNA polymerase (RdRp), spike, and nucleocapsid gene analysis. Phylogenetic analysis showed at least three distinct clusters of HCoV-OC43, although 10 unusual strains displayed incongruent phylogenetic positions between RdRp and spike genes. This suggested the presence of four HCoV-OC43 genotypes (A to D), with genotype D most likely arising from recombination. The complete genome sequencing of two genotype C and D strains and bootscan analysis showed recombination events between genotypes B and C in the generation of genotype D. Of the 29 strains, none belonged to the more ancient genotype A, 5 from 2004 belonged to genotype B, 15 from 2004 to 2006 belonged to genotype C, and 1 from 2004 and all 8 from 2008 to 2011 belonged to the recombinant genotype D. Molecular clock analysis using spike and nucleocapsid genes dated the most recent common ancestor of all genotypes to the 1950s, genotype B and C to the 1980s, genotype B to the 1990s, and genotype C to the late 1990s to early 2000s, while the recombinant genotype D strains were detected as early as 2004. This represents the first study to describe natural recombination in HCoV-OC43 and the evolution of different genotypes over time, leading to the emergence of novel genotype D, which is associated with pneumonia in our elderly population.

Coronaviruses cause infections in a wide variety of animals, resulting in respiratory, enteric, hepatic, and neurological diseases of various levels of severity. Based on genotypic and serological characterization, coronaviruses traditionally were classified into three distinct groups, groups 1, 2, and 3 (4). Recently, the Coronavirus Study Group of the International Committee for Taxonomy of Viruses has renamed the traditional group 1, 2, and 3 coronaviruses as *Alphacoronavirus*, *Betacoronavirus*, and *Gammacoronavirus*, respectively (http://talk.ictvonline.org/media/p/1230.aspx).

The recent severe acute respiratory syndrome (SARS) epidemic due to SARS coronavirus (SARS-CoV) and the identification of SARS-related coronaviruses (SARSr-CoVs) from Himalayan palm civets and horseshoe bats in mainland China have led to a boost in interest in the study of coronaviruses in both humans and animals (5, 13, 24, 26, 33, 37, 55). Before the SARS epidemic in 2003, there were only 19 known coronaviruses, including 2 human, 13 mammalian, and 4 avian coronaviruses. After the SARS epidemic, more than 20 additional novel coronaviruses have been described with complete genome sequences (9, 24–26, 31, 42, 45, 50, 53, 54, 57). These include 3 human coronaviruses, 15 mammalian coronaviruses, and 4 avian coronaviruses. For human coronaviruses, human coronavirus NL63 (HCoV-NL63) (an alphacoronavirus) and human coronavirus HKU1 (HCoV-HKU1) (a betacoronavirus) have been discovered in addition to the two previously known human coronaviruses, human coronavirus 229E (HCoV-229E) (an alphacoronavirus) and human coronavirus OC43 (HCoV-OC43) (a betacoronavirus), as well as SARS-CoV (a betacoronavirus) (9, 45, 53, 56). While HCoV-229E and HCoV-OC43 were thought to account for 5 to 30% of human respiratory tract infections, HCoV-NL63 and HCoV-HKU1 often were detected in <5% of respiratory tract samples (23, 29, 38). Outbreaks due to HCoV-OC43 also have been reported (3, 32, 44). Nevertheless, the different HCoVs often cocirculate, with one or two HCoVs being predominant depending on the geographical area and year (8, 11, 19, 23).

Coronaviruses are unique in having a high frequency of homologous RNA recombination, which is a result of random

* 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) 22554892. Fax: (852) 28551241. E-mail for S. K. P. Lau: skplau@hkucc.hku.hk. E-mail for P. C. Y. Woo: pcywoo@hkucc.hku.hk.

▽ Published ahead of print on 17 August 2011.

TABLE 1. Primers used for PCR and sequencing of the complete RdRp, S, and N genes of the 29 HCoV-OC43 strains

Primer name	Primer direction, sequence (5′–3′)	Gene target
LPW 1114	Forward, CYGTTTGTATATATTGCCGC	RdRp
LPW 1118	Reverse, TGATTATCAAGTGTTAAAAC	RdRp
LPW 5188	Forward, GCCCCTAGTTAGCGCTACTGAGTT	RdRp
LPW 3709	Reverse, ACTTAGGATAATCCCAACCCAT	RdRp
LPW 3064	Forward, CTGGGATGATATGTTACGCCG	RdRp
LPW 2579	Reverse, GTGTGTTGTGAACARAAYTCRTG	RdRp
LPW 3064	Forward, CTGGGATGATATGTTACGCCG	RdRp
LPW 2579	Reverse, GTGTGTTGTGAACARAAYTCRTG	RdRp
LPW 1223	Forward, TAAGTGCCTTTCAACAGGT	RdRp
LPW 1127	Reverse, KGCCTTTTGCGTTTCTGC	RdRp
LPW 1162	Forward, CCYRTTTGTRTGTATGATCC	S
LPW 1166	Reverse, YGCATAAAAAGTACCACC	S
LPW 1261	Forward, CTRCTATARYTATAGGTAGT	S
LPW 2094	Reverse, GCCCAAATTACCCAATTGTAGG	S
LPW 2095	Forward, TGATGCTGCTAAGATATATGG	S
LPW 2098	Reverse, ATTCCGARATAGCAATGCTGG	S
LPW 1839	Forward, ATCTTTTGTATGATTCTAATGG	S
LPW 1178	Reverse, GACACCAAGMCCATTAAT	S
LPW 1177	Forward, CWGCAGGTGTRCCATTTT	S
LPW 1183	Reverse, CCACAYTTCTTRAAACAAC	S
LPW 1275	Forward, TRAAATGGCCTTGGTATGT	S
LPW 1189	Reverse, TKWMWAGGAACTCTACAATA	S
LPW 6547	Forward, CTTCAAAGAACTATGGCATT	S
LPW 6548	Reverse, GACTGCAAATAGCCCAAATT	S
LPW 1192	Forward, AACCCMGAAACAAACAAC	N
LPW 1045	Reverse, GCAAGAATGGGGAACTGTGG	N
LPW 1195	Forward, GAGAGGCCCTAATCAGAA	N
LPW 1198	Reverse, TYAACTTCATTCATTTACTA	N

template switching during RNA replication that is thought to be mediated by a copy-choice mechanism (28, 46). Their tendency for recombination and high mutation rates may allow them to adapt to new hosts and ecological niches. During our previous investigations on the molecular epidemiology of HCoV-HKU1, we documented the first evidence for natural recombination in coronavirus associated with human infection, resulting in the generation of different HCoV-HKU1 genotypes (23, 52, 56). Since some strains of HCoV-HKU1 were found to display incongruent phylogenetic relationships upon the analysis of the RNA-dependent RNA polymerase (RdRp), spike (S), and nucleocapsid (N) genes, recombination events were suspected and later confirmed with the complete genome sequencing of 22 strains of HCoV-HKU1 and recombination analysis (52). Although HCoV-OC43 is thought to be the most commonly encountered human coronavirus, no similar molecular epidemiology studies have been performed, and little is known about its evolution among humans. Only five complete genome sequences of HCoV-OC43, two from the same American Type Culture Collection (ATCC) strain, VR759, that was isolated in 1967, one Paris strain that was isolated in 2001, and two Belgium strains detected in 2003 and 2004, were available in GenBank (39, 47, 48). In this study, we investigate the presence of different genotypes among HCoV-OC43 strains and identify potential recombination events that lead to the generation of novel genotypes, a situation analogous to that observed for HCoV-HKU1. HCoV-OC43 detected from the nasopharyngeal aspirates (NPAs) from 29 patients with respiratory tract infections from 2004 to 2011 were subjected to complete RdRp, S, and N gene sequencing and analysis. The clinical characteristics of patients also were analyzed in relation to molecular epidemiology results. As initial analyses showed the presence of potential recombination events, two complete genomes of HCoV-OC43 were selected for sequencing and further analysis. The emergence of a novel genotype of HCoV-OC43 through recombination and the evolution of different HCoV-OC43 genotypes also was described.

MATERIALS AND METHODS

HCoV-OC43 strains. Twenty-nine HCoV-OC43 strains were detected from nasopharyngeal aspirates of patients with acute respiratory tract infections who had been admitted to two Hong Kong public hospitals, Queen Mary Hospital and Pamela Youde Nethersole Eastern Hospital. They were randomly selected during a 7-year period (November 2004 to February 2011) and were included in this study (23). All NPAs were negative for influenza A and B viruses, parainfluenza virus types 1, 2, and 3, respiratory syncytial virus, and adenovirus by direct immunofluorescence, and they also were negative for metapneumovirus, human coronavirus HKU1, human coronavirus 229E, and human coronavirus NL63 by reverse transcription-PCR (RT-PCR) (23, 56).

RNA extraction. Viral RNA was extracted from the NPAs of the corresponding patients using a QIAamp viral RNA Minikit (QIAgen, Hilden, Germany). The RNA pellet was eluted in 50 μl of DNase-free, RNase-free double-distilled water and was used as the template for RT-PCR.

RT-PCR and sequencing of the complete RdRp, S, and N genes of HCoV-OC43 and phylogenetic analysis. The RNA was converted to cDNA by a combined random priming and oligo(dT) priming strategy. The complete RdRp, S, and N genes of HCoV-OC43 from 29 NPAs were amplified and sequenced using the primers shown in Table 1 and the strategy described in our previous publications (23, 53). The nucleotide and the deduced amino acid sequences of the RdRp, S, and N genes were compared to those of HCoV-OC43 and other group 2 coronaviruses. A phylogenetic tree was constructed using the maximum-likelihood method in PhyML with GTR for RdRp and N genes and GTR+I for the S gene. The substitution models were selected according to Akaike information criterion (AIC) implemented in ModelGenerator version 0.85 (17). The robustness of branches was assessed by bootstrap analysis with 1,000 replicates.

TABLE 2. Calculated tMRCA values for S and N genes

Genotype and analysis model	Strict clock		Relaxed uncorrelated clock[a]			
			Log-normal		Exponential	
	Mean	HPD	Mean	HPD	Mean	HPD
Constant size						
S gene						
A, B, and C	48.1555	44.7545–53.5444	49.414	44.4431–58.0039	54.284	44.5851–70.9238
B and C	27.4725	22.2275–33.6636	21.3901	10.6186–33.4467	26.9695	14.7718–42.409
A	45.5514	44.6245–46.6342	45.62	44.2208–47	46.529	44.3516–50.3417
B	11.4415	9.5698–13.5421	10.7938	8.2477–14.2259	14.8138	10.0011–21.2453
C	8.121	7.2977–9.0669	8.2939	7.1553–9.9183	9.2664	7.2557–12.4169
N gene						
A, B and C	55.6007	44.102–80.1591	55.2781	44.099–79.5748	54.433	44.0908–77.5797
B and C	26.2353	13.0594–44.625	24.6398	10.2003–43.6165	21.6914	9.1696–40.8566
A	45.58	44.08–48.971	45.5361	44.08–48.9703	45.5039	44.08–49.1539
B	16.5939	8.8253–27.7325	15.3999	8.2083–26.293	13.2508	8.1759–22.8651
C	13.6795	8.2804–21.6243	13.0869	7.6379–21.0773	12.1722	7.4932–20.213
Bayesian skyline						
S gene						
A, B and C	45.0589	44.4048–45.8426	44.982	44.09–46.3717	45.0131	44.1592–46.4873
B and C	20.8376	17.939–23.7088	10.1779	8.5117–12.5169	10.5825	8.8659–12.8411
A	44.8137	44.3591–45.3259	44.4406	44.1038–44.9446	44.4342	44.1312–44.8766
B	9.5314	8.6617–10.4592	8.4944	8.0953–9.0724	8.479	8.1131–9.005
C	7.3604	7.0823–7.6852	7.3576	7.08–7.8177	7.3233	7.08–7.7445
N gene						
A, B, and C	52.9938	44.081–78.656	52.6201	44.0807–77.0595	51.413	44.0802–72.6384
B and C	23.9449	12.055–42.3612	23.1098	10.0302–41.162	20.571	8.7883–38.149
A	45.151	44.08–47.9311	45.1336	44.08–47.8964	45.119	44.08–48.0281
B	15.6571	8.8518–25.9166	15.1134	8.3088–24.9999	13.5672	8.1754–23.0218
C	13.6085	8.2093–21.2408	13.3621	7.8461–20.9611	12.7738	7.522–21.2622

[a] The adopted model is shaded gray.

Complete genome sequencing. Two complete genomes of HCoV-OC43 from two patients (HK04-01 and HK04-02) were amplified and sequenced using the RNA extracted from the original NPAs as 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 available HCoV-OC43 genome sequences using strategies described in our previous publications (24, 50, 52, 53), and the coronavirus database (CoVDB) (15) was used for sequence retrieval. These primer sequences are available on request. The 5′ ends of the viral genomes were confirmed by the rapid amplification of cDNA ends (RACE) 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 using CoVDB (15). Phylogenetic tree construction was performed using the neighbor-joining method using ClustalX 1.83, with bootstrap values calculated from 1,000 trees. The prediction of the receptor binding domain of spike protein was performed using InterProScan (2). The prediction of potential N-glycosylation sites in the spike proteins was performed using the CBS NetNGlyc 1.0 server (http://www.cbs.dtu.dk/services/NetNGlyc/). 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 was calculated using the Nei and Gojobori substitution model with Jukes-Cantor correction in MEGA 4.0 (41). Bootscan analysis was used to detect possible recombination using the nucleotide alignment of the genome sequences of HCoV-OC43 strains HK04-01 and HK04-02, the ATCC strain, and Belgium strain BE03, generated by ClustalX version 1.83, and edited manually. Bootscan analysis was performed using Simplot version 3.5.1 as described previously (21, 22, 52), with strains HK04-02 and BE04 as the query.

Estimation of divergence dates. Divergence times for the three OC43 genotypes, A, B, and C, were calculated using a Bayesian Markov chain Monte Carlo (MCMC) approach as implemented in BEAST (version 1.6.1) as described previously (7, 22, 47, 49). One parametric model (constant size) and one nonparametric model (Bayesian skyline) tree priors were used for the inference. Analyses were performed under the GTR+I and GTR substitution models for S and N gene sequence data, respectively, and using both a strict and a relaxed molecular clock. For the S gene, the MCMC run was 3×10^7 steps long, while for the N gene the MCMC runs were 4×10^8 and 6×10^8 when the Bayesian skyline and constant-size coalescent tree prior was used, respectively, with sampling every 1,000 steps. Convergence was assessed on the basis of the effective sampling size after a 10% burn-in using Tracer software, version 1.5 (7). The mean time of the most recent common ancestor (tMRCA) and the highest posterior density regions at 95% (HPDs) were calculated, and the best-fitting models were selected by a Bayes factor using marginal likelihoods implemented in Tracer (Tables 2 and 3) (40). Constant population size under a relaxed-clock model with an uncorrelated exponential distribution was adopted for making inferences, as Bayes factor analysis for the S gene indicated that this model fitted the data better than other models tested (Table 3). The trees were summarized in a target tree by the Tree Annotator program included in the BEAST package by choosing the tree with the maximum sum of posterior probabilities (maximum clade credibility) after a 10% burn-in.

Nucleotide sequence accession numbers. The nucleotide sequences of the two genomes of HCoV-OC43 have been lodged within the GenBank sequence database under accession no. JN129834 and JN129835.

RESULTS

Clinical characteristics of patients with HCoV-OC43 infections. The clinical characteristics of the 29 patients with HCoV-OC43 infections are summarized in Table 4. Sixteen were males and 13 were females. Thirteen were children and 16 were adults, with most patients at the extremes of age (median age, 36 years old; range, 24 days to 94 years old).

TABLE 3. Model selection results for the S and N genes of HCoV-OC43 through comparison of marginal likelihoods and $\log_{10}$ Bayes factors for each pair of models

Model combination[a]	Log-normal marginal likelihood		Model combination					
	P[b] (model/data)	SE[c]	Strict_CST	Strict_BSP	Uced_CST	Uced_BSP	Ucld_CST	Ucld_BSP
S gene								
Strict_CST	−7,420.992	±0.119		9.524	−9.283	−3.477	−5.761	−4.179
Strict_BSP	−7,442.921	±0.147	−9.524		−18.806	−13	−15.285	−13.702
Uced_CST	−7,399.618	±0.143	9.283	18.806	−	5.806	3.521	5.104
Uced_BSP	−7,412.987	±0.155	3.477	13	−5.806		−2.285	−0.702
Ucld_CST	−7,407.726	±0.169	5.761	15.285	−3.521	2.285		1.583
Ucld_BSP	−7,411.371	±0.167	4.179	13.702	−5.104	0.702	−1.583	
N gene								
Strict_CST	−2,183.957	±0.023		−0.929	−1.969	−2.647	−0.48	−1.289
Strict_BSP	−2,181.818	±0.031	0.929		−1.04	−1.718	0.449	−0.36
Uced_CST	−2,179.423	±0.03	1.969	1.04		−0.678	1.489	0.681
Uced_BSP	−2,177.861	±0.028	2.647	1.718	0.678		2.167	1.359
Ucld_CST	−2,182.852	±0.026	0.48	−0.449	−1.489	−2.167		−0.809
Ucld_BSP	−2,180.99	±0.032	1.289	0.36	−0.681	−1.359	0.809	

[a] Three molecular clock models, a strict clock and two relaxed clocks assuming either uncorrelated exponential (Uced) or uncorrelated lognormal distribution (Ucld) of substitution rates, were compared in combination with models of demographic history, constant size (CST), and Bayesian skyline (BSP). The adopted model is shaded gray.
[b] Marginal likelihood estimated using the program Tracer 1.5. A $\log_{10}$ Bayes factor was calculated for each pair of model combinations (i.e., model 1 in row versus model 2 in column).
[c] Standard errors for the marginal likelihoods.

Fifteen patients, especially infants and young children, presented with upper respiratory tract infections. Of the 14 patients with pneumonia, three were infants or young children, nine were elderly (>60 years old), and the other two were a 36-year-old male and a 46-year-old female, respectively. The 36-year-old male patient (patient 5) did not have underlying disease and presented with chest pain and hemoptysis complicated by right pleural effusion. The 46-year-old female patient

TABLE 4. Clinical characteristics of the 29 patients with HCoV-OC43 infections

Patient no.	Mo/yr	Strain no.	Sex[a]	Age	Diagnosis[b]	Genotype
1	11/2004	HK04-01	F	9 yr	URTI	C
2	11/2004	HK04-02	M	35 mo	URTI	D
3	11/2004	HK04-03	F	1 yr	URTI, febrile convulsion	C
4	11/2004	HK04-04	M	24 m	URTI, febrile convulsion	B
5	11/2004	HK04-05	M	36 yr	Pneumonia, pleural effusion	C
6	11/2004	HK04-06	F	82 yr	Pneumonia, COPD exacerbation	C
7	11/2004	HK04-07	F	24 days	URTI	B
8	11/2004	HK04-08	F	5 yr	Pneumonia	B
9	11/2004	HK04-09	M	1 yr	URTI, febrile convulsion	C
10	11/2004	HK04-10	M	2 yr	URTI	B
11	11/2004	HK04-11	M	85 yr	URTI, urinary tract infection	C
12	11/2004	HK04-12	M	1 mo	URTI	C
13	11/2004	HK04-13	M	76 yr	URTI	C
14	12/2004	HK04-14	F	32 yr	URTI	C
15	12/2004	HK04-15	M	72 yr	URTI	C
16	12/2004	HK04-16	M	1 yr	Pneumonia	C
17	12/2004	HK04-17	F	46 yr	Pneumonia, oral herpes	B
18	12/2004	HK04-18	M	68 yr	Pneumonia, COPD exacerbation	C
19	1/2005	HK05-01	M	11 mo	URTI, NSAID-induced angioedema	C
20	12/2005	HK05-02	M	88 yr	Pneumonia	C
21	1/2006	HK06-01	F	84 yr	URTI	C
22	12/2008	HK08-01	F	27 mo	URTI	D
23	11/2008	HK08-02	M	83 yr	Pneumonia, pleural effusion, CHF	D
24	10/2009	HK09-01	M	26 mo	Pneumonia	D
25	11/2009	HK09-02	M	84 yr	Pneumonia	D
26	10/2010	HK10-01	F	82 yr	Pneumonia	D
27	10/2010	HK10-02	F	87 yr	Pneumonia, acute pulmonary edema	D
28	2/2011	HK11-01	F	72 yr	Pneumonia, CHF	D
29	2/2011	HK11-02	F	94 yr	Pneumonia	D

[a] F, female; M, male.
[b] URTI, upper respiratory tract infection; COPD, chronic obstructive pulmonary disease; NSAID, nonsteroidal antiinflammatory drug; CHF, congestive heart failure.

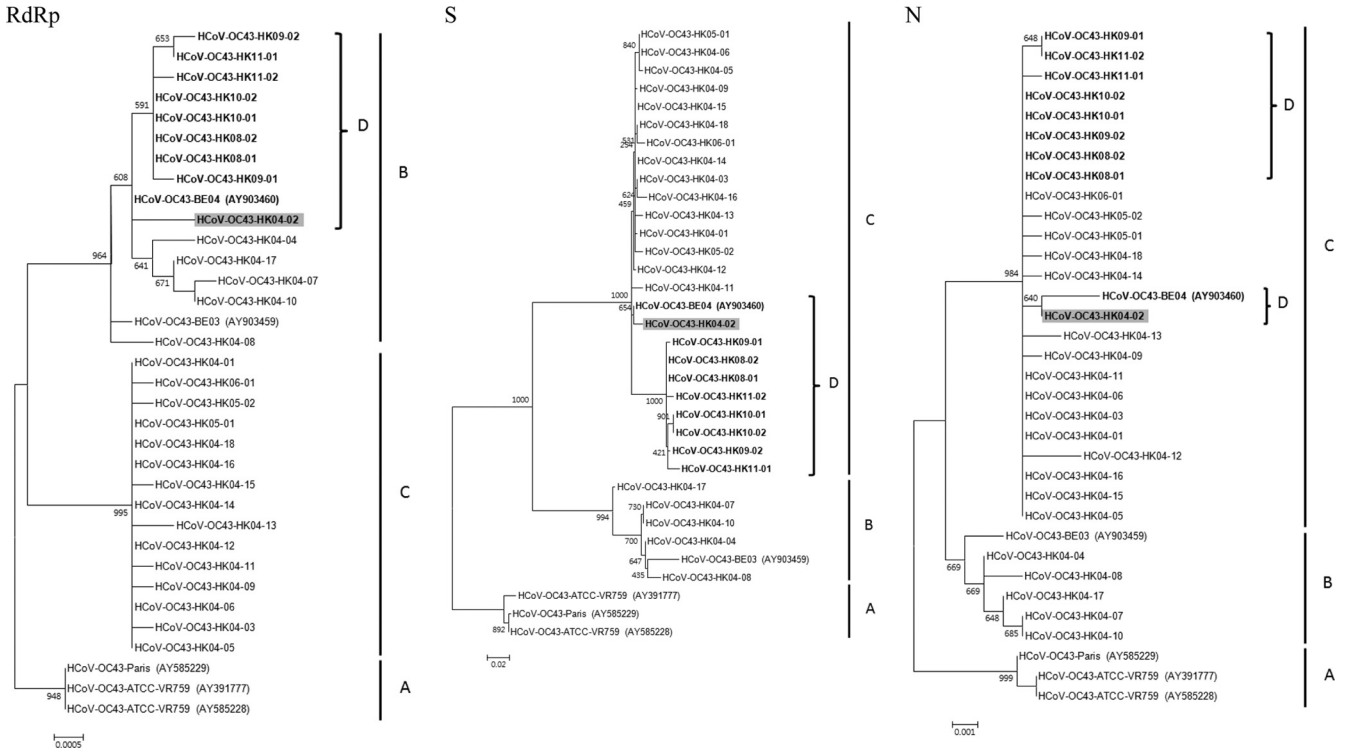


FIG. 1. Phylogenetic analysis of the complete RdRp, S, and N genes of the 21 HCoV-OC43 strains. The trees were constructed by the maximum-likelihood method, and bootstrap values were calculated from 1,000 trees. A total of 2,783, 4,089, and 1,347 nucleotide positions in RdRp, S, and N, respectively, were included in the analysis. The scale bar indicates the estimated number of substitutions per 2,000, 1,000, or 20 nucleotides as indicated. A, genotype/clade A; B, genotype/clade B; C, genotype/clade C; D, genotype D. Genotype D strains displaying incongruent phylogenetic relationships are in boldface.

(patient 17) had a history of asthma, and her pneumonia was complicated by oral herpes. Two elderly patients were complicated by the exacerbation of chronic obstructive pulmonary diseases, and three were complicated by congestive heart failure or acute pulmonary edema. Two children were complicated by febrile convulsion. Two patients (patients 5 and 18) had recent travel history to mainland China before symptom onset. Except for a 94-year-old female patient (patient 29) who died of pneumonia with superimposed *Pseudomonas aeruginosa* infection, no other respiratory pathogens were detected in other patients and all survived.

RT-PCR and sequencing of the complete RNA-dependent RNA polymerase, spike, and nucleocapsid genes of HCoV-OC43 and phylogenetic analysis. The complete RdRp, S, and N genes of HCoV-OC43 from the 29 NPAs were amplified and sequenced. The analysis of the RdRp genes showed that the 29 HK strains possessed 99.6 and 100% nucleotide identity to those of Belgium strains BE04 and BE03, respectively. Multiple alignments revealed that there were 33 nucleotide substitutions among the RdRp genes of the 29 HK strains. Phylogenetic analysis showed the existence of three clusters of sequences, all with high bootstrap values of >900 (Fig. 1). One cluster, clade A, was formed by the ATCC and Paris strains. Fourteen HK strains were clustered with the BE03 and BE04 strains, forming clade B. The other 19 HK strains were clustered together, forming clade C.

The analysis of the S genes showed that 5 of the 29 HK strains from 2004 (HK04-04, HK04-07, HK04-08, HK04-10, and HK04-17) possessed high nucleotide identities (99.1 to 99.5%) to Belgium strain BE03, while the other 24 sequences possessed high nucleotide identities (99.2 to 99.9%) to Belgium strain BE04. Multiple alignments revealed that there were 166 nucleotide substitutions among the S genes of the 29 HK strains. Phylogenetic analysis showed that the sequences fell into three clusters, all with high bootstrap values of >990 (Fig. 1). Again, one cluster, clade A, was formed by the ATCC and Paris strains. In line with results from pairwise identities described above, five HK strains from 2004 and Belgium strain BE03 formed another cluster, clade B, while the other 24 HK strains and Belgium strain BE04 formed the third cluster, clade C.

Analysis of the N genes of the 29 HK strains showed results similar to those for the S genes. Multiple alignments revealed that there were 23 nucleotide substitutions among the N genes of the 29 HK strains. Phylogenetic analysis showed the existence of three clusters of sequences (Fig. 1). Again, the ATCC and Paris strains formed a distinct cluster, clade A. Similarly to findings of S gene analysis, five HK strains from 2004 were most closely related to the Belgium strain BE03, forming clade B. The N gene sequences of the other 24 HK strains were more closely related to those of Belgium strain BE04, forming another cluster, clade C.

From these results, 10 unusual strains, including Belgium strain BE04 and nine HK strains (one from 2004 and eight from 2008 to 2011), were found to display incongruent phylogenetic positions upon the analysis of their RdRp, S, and N

gene sequences. They belonged to clade B in the RdRp tree but belonged to clade C in the S and N trees. Multiple alignment also revealed that these 10 unusual strains displayed higher nucleotide similarity to other clade B strains in their RdRp genes, but they displayed higher nucleotide similarity to clade C strains in their S and N genes. These results suggested the presence of at least three genotypes of HCoV-OC43: genotype A (comprising only the ATCC and Paris strains), genotype B (including Belgium strain BE03 and five HK strains from 2004), and genotype C (including 15 HK strains from 2004 to 2006). Moreover, 10 unusual strains, including Belgium strain BE04 and nine HK strains, may represent an additional genotype, genotype D, which may have arisen from recombination between genotypes B and C. Interestingly, while one (HK04-02) of the nine genotype D strains from Hong Kong was detected in 2004, all eight HK strains from 2008 to 2011 included in this study belonged to genotype D, suggesting that this genotype is the predominant genotype in recent years in our population. Moreover, seven of these eight recent strains were associated with pneumonia, especially in the elderly, three of which were complicated by heart failure (Table 4).

Complete genome analysis of two HCoV-OC43 strains, HK04-01 and HK04-02. To determine if genotype D strains have arisen from recombination between genotypes B and C, the complete genome sequences of one selected strain from genotype D, HK04-02, and one selected strain from genotype C, HK04-01, were determined (complete genome sequence was available for genotype B strain BE03). The sizes of the genomes of the two HCoV-OC43 strains, HKU04-01 and HKU04-02, ranged from 30,710 to 30,722 nucleotides, with a G+C content of 37%. Their genome organizations were typical of *Betacoronavirus* (Fig. 2).

Phylogenetic trees using the nucleotide sequences of putative proteins and polypeptides (nsp1 to nsp16, NS2a, hemagglutinin-esterase [HE], S, NS5a, envelope [E], membrane [M], and N) of the two present strains and five other HCoV-OC43 complete genome sequences available (from ATCC, Paris, and Belgium strains) were constructed and are shown in Fig. 3. In most of the trees, the ATCC and Paris strains formed a distinct cluster separate from the two Belgium and two HK strains. In most nsp proteins except nsp5 and nsp15, strains HK04-02 and BE04 were more closely clustered with strain BE03 than with strain HK04-01 (in nsp1 to nsp4, nsp6, nsp8, and nsp12) or clustered closely with both strains (in nsp7, nsp9 to nsp11, nsp13, nsp14, and nsp16). However, in HE, S, E, M, and N genes, strains HK04-02 and BE04 were more closely clustered with strain HK04-01 than with strain BE03. This suggested that the two strains have arisen from recombination.

Results from bootscan analysis were in line with the observations described above. From the 5′ end of the genome to position 22500, bootscan analysis showed a number of possible recombination sites when the genome of strain HK04-02 was used as the query (Fig. 2). Upstream of position 15500, most of the region exhibits higher bootstrap support for the clustering of strain HK04-02 with BE03, except between positions 2500 and 5000, where higher bootstrap support for clustering with strain HK04-01 was observed. From positions 17000 to 22500, most of the region exhibits higher bootstrap support for clustering between strains HK04-02 and HK04-01. From position 22500 to the 3′ end of the genome, no further recombination sites were detected, with high bootstrap support for clustering between strains HK04-02 and HK04-01 observed. These indicated that recombination has taken place between nucleotide positions 2500 and 5000 (corresponding to the nsp2/nsp3 junction), between nucleotide positions 15500 and 17000 (corresponding to the nsp12/nsp13 junction), and/or around nucleotide position 22600 (corresponding to the NS2a/HE junction). Similar results were obtained when strain BE04 was subjected to bootscan analysis (Fig. 2). Among these three potential recombination sites, the one at the nsp12/nsp13 junction most likely has resulted in the incongruent phylogenetic positions observed in the RdRp, S, and N genes among the genotype D strains, as this is the junction that bridges between reciprocal clustering to genotype B and C strains observed in RdRp and S genes, respectively.

Further analysis of the present two HCoV-OC43 genomes revealed the presence of a putative transcription regulatory sequence (TRS) motif, 5′-UC(U/C)AAAC-3′, at the 3′ end of the leader sequence, and it preceded each translated ORF except for those of the NS5a and E genes, as described for HCoV-OC43 genomes previously (39). For NS5a, it has been suggested that 5′-UCUUAAG-3′ was the putative TRS in the ATCC strain (39). However, this sequence was absent from the two Belgium strains, BE03 and BE04, and the present two strains, HK04-01 and HK04-02, as a result of a 12-nucleotide deletion from the region (Fig. 4). Upstream of this deletion, a potential alternative TRS, 5′-UCUAGCA-3′, was identified 20 bp upstream to the start codon. It has been found previously that sequences in a similar region, upstream of E of HCoV-NL63 and NS5a of HCoV-HKU1, were homologous to a fragment of the corresponding leader sequences, which may serve as a compensation mechanism for the absence of an optimal TRS (34, 35). In HCoV-OC43, a 39-bp sequence upstream of NS5a also was found to be homologous to its leader sequence (Fig. 4). Further experiments are required to determine if NS5a utilizes a TRS or another mechanism for translation. As in sialodacryoadenitis virus (SDAV) and mouse hepatitis virus (MHV), the putative E gene of HCoV-OC43 may share the same TRS with NS5a, suggesting that the translation of the E protein is cap independent, possibly via an internal ribosomal entry site (IRES) (16). In MHV, an IRES element, UUUUA UUCUUUUU, has been identified upstream of the initiation codon of the E protein (16). A stretch of 13 nucleotides, CUUUAUUACCUGG, is present at this position in HCoV-OC43 according to multiple alignment. Further experiments would determine if this or another nearby sequence acts as an IRES for the E protein of HCoV-OC43.

The analysis of the predicted S proteins of the present 29 HCoV-OC43 HK strains revealed a potential N-terminal signal peptide of 14 to 17 amino acids by SignalP-HMM and SignalP-NN, respectively. A potential cleavage site located after RRSRR that is identical to that of bovine CoV (BCoV) (1) and the two Belgium HCoV-OC43 strains BE03 and BE04 (48), between residues 766 and 767, where S will be cleaved into S1 and S2, was identified in the S proteins of all 29 strains. The residues RRSRG were observed upstream of the potential cleavage site in the S proteins of the ATCC and Paris strains. It has been suggested that the G-to-R amino acid change in the last position, leading to an RRSRR motif, leads to an increased cleavability compared to that of the ATCC prototype

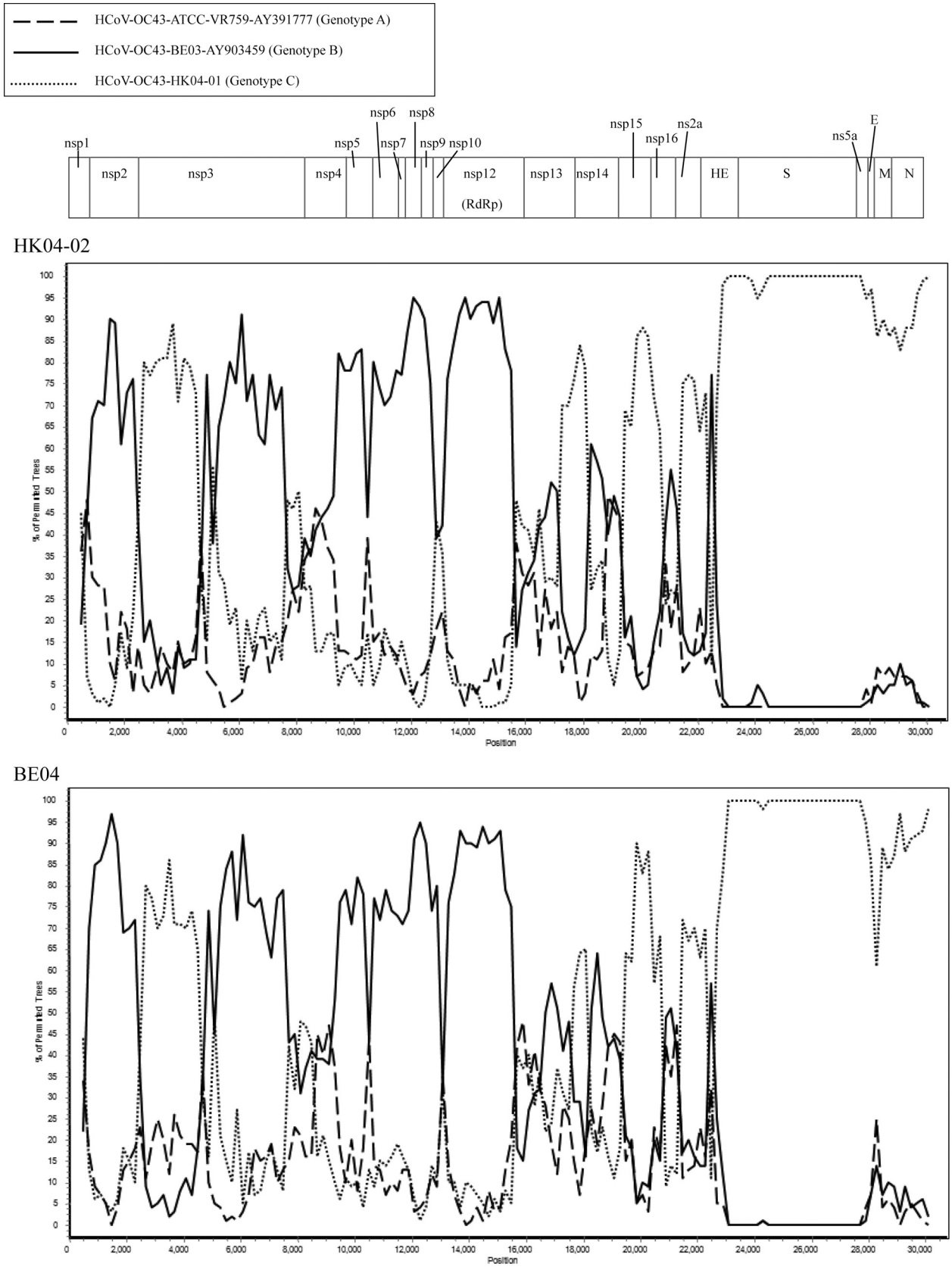
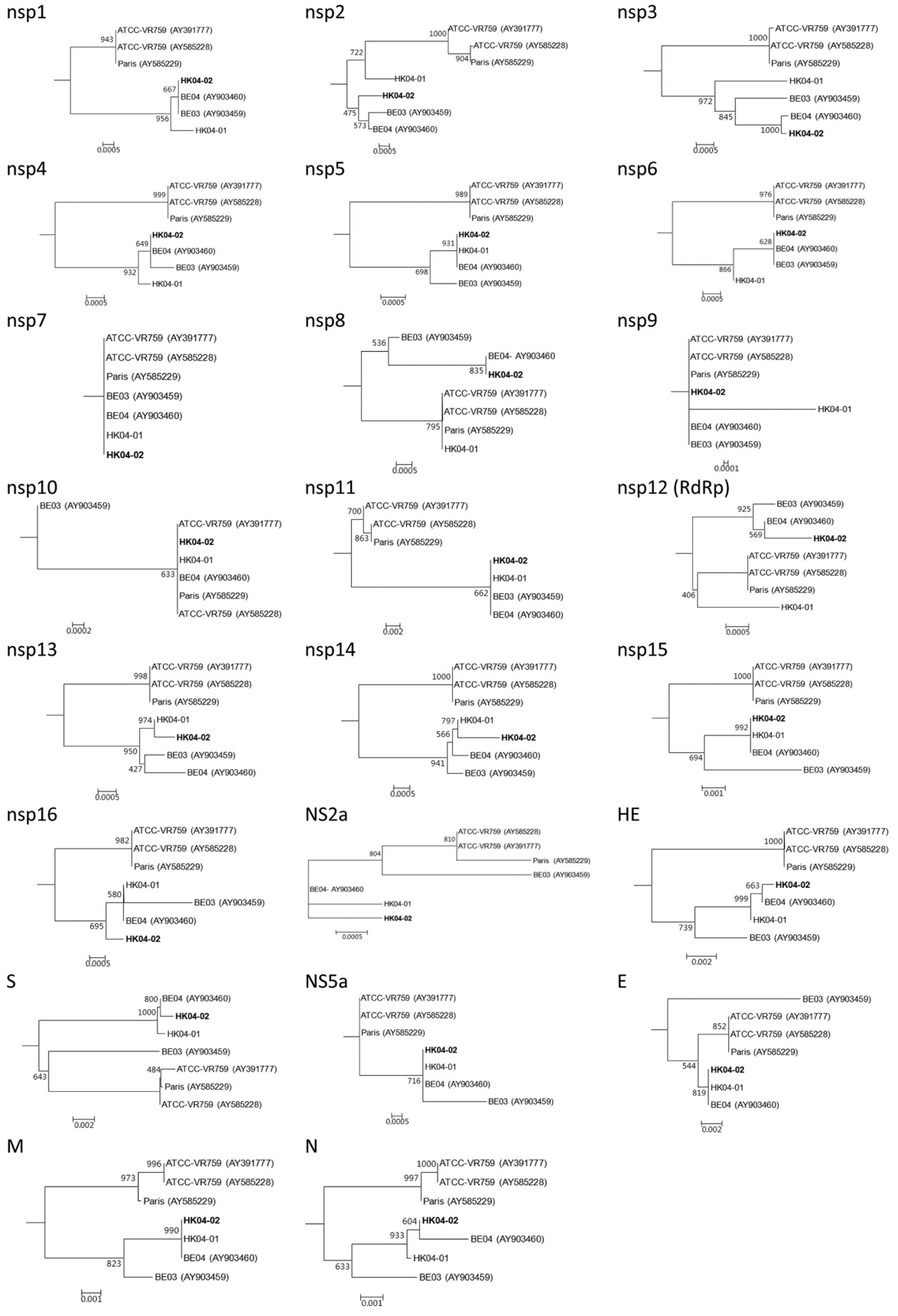


FIG. 2. Genome organization and bootscan analysis of the HCoV-OC43 genomes. Bootscanning was conducted with Simplot version 3.5.1 (F84 model; window size, 1,000 bp; step, 200 bp) on a gapless nucleotide alignment, which was generated with ClustalX with the genome sequences of strains HK04-02 (upper) and BE04 (lower) as the query sequences.



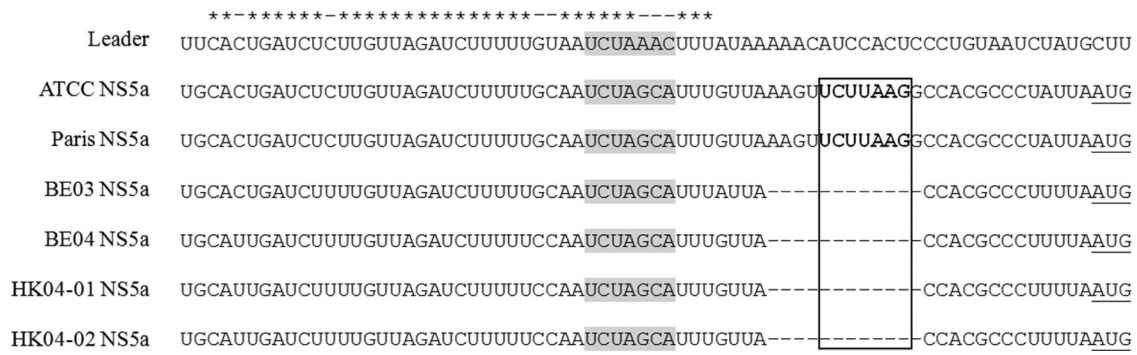


FIG. 4. Alignment of the leader sequence and the homologous sequence upstream of the NS5a ORF. Sequence homology between the sequences is marked by asterisks. The putative TRS is highlighted in gray. The TRS suggested by St.-Jean et al. (39) are in boldface, and the corresponding deletions in BE03, BE04, HK04-01, and HK04-02 are indicated by the box. The initiation codons of NS5a are underlined.

HCoV-OC43 spike protein (48). The analysis of the S proteins of HCoV-OC43 also reveals different numbers of potential N-linked glycosylation sites in different strains, with six NXS and eight NXT sites in the ATCC and Paris strains, six NXS and nine NXT sites in strain BE03, and seven NXS and nine NXT sites in strains BE04, HK04-01, and HK04-02. Sialic acid was known to be the receptor for S protein binding in HCoV-OC43, although the receptor-binding domain is not well defined (18). The S gene of the 29 HK strains contains a putative receptor binding domain (amino acid positions 339 to 549) with 90 to 96.5% amino acid identities to sequences of ATCC and Paris strains.

Using all seven available HCoV-OC43 genome sequences for analysis, the K_a/K_s ratios for the various coding regions were calculated (Table 5). The highest K_a/K_s ratios in HCoV-OC43 genomes were observed at nsp5 (0.599), followed by HE (0.542), S (0.402), and N (0.369), suggesting that these regions in HCoV-OC43 are under higher selection pressure. However, when using the 29 HK strains (from 2004 to 2011) for analysis, the K_a/K_s ratio for the S gene dropped to 0.270. This is likely due to the very high K_a/K_s ratio observed at S genes among the three genotype A (5.134) strains compared to ratios for genotype B (0.332) and genotype C (0.316) strains.

Estimation of divergence dates. Using the constant population size under a relaxed-clock model with an uncorrelated exponential distribution, the mean evolutionary rate of HCoV-OC43 was estimated at 6.7×10^{-4} and 3.6×10^{-4} nucleotide substitutions per site per year for S and N genes, respectively (Fig. 5). These estimates were comparable to previous findings of 6.1×10^{-4} and 3.6×10^{-4} nucleotide substitutions per site per year for S and N genes of pigeon herpes encephalomyelitis virus, BCoV, and HCoV-OC43 (49). Molecular clock analysis using the S gene showed that the tMRCA of all HCoV-OC43 genotypes was estimated at 1,956.8 (HPDs, 1,940.16 to 1,966.49) 54 years ago; that of genotype A at 1,964.55 (HPDs, 1,960.74 to 1,966.73) 47 years ago; that of genotypes B and C at 1,984.11 (HPDs, 1968.67 to 1996.31) 27 years ago; that of genotype B at 1,996.27 (HPDs, 1,989.83 to 2,001.08) 15 years ago; and that of genotype C at 2,001.81 (HPDs, 1,998.66 to 2,003.82) 9 years ago. Molecular clock analysis using the N gene showed that the tMRCA of all HCoV-OC43 genotypes was estimated at 1,956.65 (HPDs, 1,933.5 to 1,966.99), also 54 years ago; that of genotype A at 1,965.58 (HPDs, 1,961.92 to 1,967) 46 years ago; that of genotypes B and C at 1,989.39 (HPDs, 1,970.22 to 2,001.91) 22 years ago; that of genotype B at 1,997.93 (HPDs, 1,988.21 to 2,002.9) 13 years ago; and that of genotype C at 1,998.91 (HPDs, 1,990.87 to 2,003.59) 12 years ago. These estimates using the two different genes were consistent and also were in line with a previous study demonstrating the tMRCA of HCoV-OC43 strains at 1,944 and 1,957 by S and N gene analysis, respectively (49).

DISCUSSION

The present study represents the first report of possible natural recombination among HCoV-OC43 strains, which has resulted in the emergence of strains of potentially novel genotypes. Although HCoV-OC43 was first discovered in 1967 (30), genomic studies of HCoV-OC43 have been scarce, with the first complete genomes coming from a laboratory strain from the ATCC and a clinical isolate, designated Paris, reported in 2004 (39). This was followed by genomic studies on two HCoV-OC43 strains detected in 2003 and 2004 in Belgium, showing that the Belgium strains were genetically distinct and that HCoV-OC43 could have originated from recent zoonotic transmission (47, 48). It also was found that the Paris isolate may be cross-contaminated with the ATCC strain, which explains their close genetic relatedness (48). A later study from France analyzing the S1 genes of seven HCoV-OC43 strains also showed high genetic diversity (43). In this study, we

FIG. 3. Phylogenetic analysis of nsp1 to nsp16, NS2a, HE, S, NS5a, E, M, and N genes of seven HCoV-OC43 genomes. The trees were constructed by the neighbor-joining method using Kimura's two-parameter correction, and bootstrap values were calculated from 1,000 trees. A total of 738, 1,815, 5,697, 1,488, 909, 861, 267, 591, 330, 411, 46, 2,784, 1,809, 1,563, 1,125, 900, 837, 1,244, 4,092, 330, 248, 678, and 1,347 nucleotide positions in nsp1, nsp2, nsp3, nsp4, nsp5, nsp6, nsp7, nsp8, nsp9, nsp10, nsp12, nsp13, nsp14, nsp15, nsp16, NS2a, HE, S, NS5a, E, M, and N, respectively, were included in the analysis. The scale bar indicates the estimated number of substitutions per 50 or 100 nucleotides as indicated. The corresponding nucleotide sequences of HCoV-HKU1 were used as the outgroups.

TABLE 5. Estimation of nonsynonymous and synonymous substitution rates in the seven genomes of HCoV-OC43

Gene	Function	HCoV-OC43 substitution rates		
		K_a	K_s	K_a/K_s
nsp1	Unknown	0.003	0.01	0.24
nsp2	Unknown	0.003	0.012	0.245
nsp3	Papain-like protease	0.002	0.014	0.17
nsp4	Unknown	0.002	0.011	0.142
nsp5	3C-like protease	0.002	0.004	0.599
nsp6	Unknown	0	0.015	0
nsp7	Predicted replicase	0	0	
nsp8	Predicted replicase	0.002	0.008	0.277
nsp9	RNA synthesis protein	0	0.004	0
nsp10	RNA synthesis protein	0	0.003	0
nsp11	Unknown	0	0.06	0
nsp12	RdRp	0.001	0.009	0.066
nsp13	Helicase	0	0.014	0.015
nsp14	Unknown	0.001	0.015	0.065
nsp15	Endoribonuclease-like	0.003	0.015	0.169
nsp16	Putative methyltransferase	0	0.016	0
ns2a	Predicted phosphoesterase	0	0.015	0
HE	Helicase	0.011	0.02	0.542
S	Spike	0.014	0.035	0.402[a]
ns5a	Unknown	0.003	0	
E	Envelope	0.003	0.01	0.299
M	Membrane	0.004	0.019	0.223
N	Nucleocapsid	0.005	0.013	0.369

[a] The K_a/K_s ratio for S genes of the 29 HK strains from 2004 to 2011 was 0.270.

showed that there were at least three distinct clusters of HCoV-OC43 strains upon RdRp, S, and N gene analysis. One cluster, clade A, was formed by the ATCC and Paris strains. The other two clusters, clade B and clade C, were formed by the present HK strains and Belgium strains BE03 and BE04. However, 10 unusual strains displayed incongruent phylogenetic positions and belonged to clade B upon RdRp gene analysis and to clade C upon S and N gene analysis. These results suggested the presence of four different genotypes of HCoV-OC43, genotype A (comprising the ATCC and Paris strains), genotype B (including Belgium strain BE03 and five HK strains from 2004), genotype C (including 15 HK strains from 2004 to 2006), and genotype D (including the 10 unusual strains: Belgium strain BE04 and 9 HK strains, 1 from 2004 and 8 from 2008 to 2011). Moreover, genotype D is likely a recombinant genotype which has arisen from recombination between genotype B and C strains at a region between the RdRp and S genes within the genome. To investigate the suspected recombination event, complete genome sequences of two strains, HK04-01 (genotype C) and HK04-02 (genotype D), were determined. Both phylogenetic and bootscan analyses showed possible recombination events between genotypes B and C in the generation of genotype D strains, a situation similar to that reported for HCoV-HKU1 (52). The analysis of more HCoV-OC43 strains from other countries also will reveal the relative prevalence of the different genotypes in different localities and the presence of additional genotypes arising from other recombination events.

The recombinant genotype D strains may represent an emerging HCoV-OC43 genotype associated with human infections. The present study, the first molecular epidemiology study on HCoV-OC43 infections with clinical characteristics presented, revealed genetic evolution into different genotypes over time. In a previous study from Belgium, three phylogenetic clusters were identified based on S gene analysis, the ATCC cluster and two clusters containing four 2003 strains and three 2004 strains, respectively, suggesting different temporal patterns among different clusters (48). In this study, 29 HK strains collected during a 7-year period were included to better elucidate the genetic evolution of HCoV-OC43 over time. None of the contemporary strains belong to genotype A, which consisted only of ATCC and Paris strains that likely were isolated 44 years ago. Five of the HK strains from 2004, together with Belgium strain BE03 from 2003, belonged to genotype B. Fifteen HK strains from 2004 to 2006 belonged to genotype C. One HK strain from 2004, eight HK strains from 2008 to 2011, and Belgium strain BE04 belonged to genotype D. While only 1 of the 18 HK strains from 2004 belonged to genotype D, all 8 HK strains from 2008 to 2011 belonged to this recombinant genotype. This suggests that new genotypes of HCoV-OC43 have evolved over time, with the most recent HCoV-OC43 strains circulating in our population being dominated by genotype D, which likely has arisen from recombination as early as 2004. Molecular clock analysis using S and N gene sequences suggested that the most recent common ancestor of all HCoV-OC43 genotypes emerged in the 1950s (mean, 1957), while genotype B and C emerged in the 1980s (means, 1984 and 1989 by S and N gene analysis, respectively), genotype B emerged in the 1990s (means, 1996 and 1998 by S and N gene analysis, respectively), and genotype C emerged in the late 1990s to early 2000s (means, 1999 and 2002 by S and N gene analysis, respectively). Although the tMRCA of the recombinant genotype D strains could not be studied by molecular clock analysis, the detection of a genotype D HK strain and the reported Belgium strain BE04 from 2004 suggested that this genotype has emerged no later than that year. Moreover, seven of the eight genotype D HK strains from 2008 to 2011 were associated with pneumonia, especially in the elderly, suggesting that this emerging, recombinant genotype is associated with more severe disease. However, molecular epidemiology studies involving a larger number of strains and from different geographical areas are required to better understand the molecular evolution of HCoV-OC43 and the relative pathogenicity of the different genotypes. Continuous studies also are warranted to detect the emergence of new genotypes and recombinants of HCoV-OC43 as well as other human coronaviruses and to assess their significance and potential in causing future epidemics. Nevertheless, it should noted that the amplification and sequencing of a single gene may not be sufficient to define the genotypes of HCoV-OC43, HCoV-HKU1, HCoV-NL63, and probably other coronaviruses (36, 52). Given that recombination events are not uncommon among these human coronaviruses, the amplification and sequencing of at least two gene loci, probably one from ORF1ab (e.g., RdRp or helicase) and one from HE to N (e.g., S or N), should be performed to more accurately understand their molecular epidemiology and reveal novel genotypes due to recombination events.

Although MHV is, historically, the most well-studied coronavirus for recombination in *in vitro* studies (10, 28), there is increasing evidence for natural recombination in other coronaviruses, some of which lead to the generation of new strains

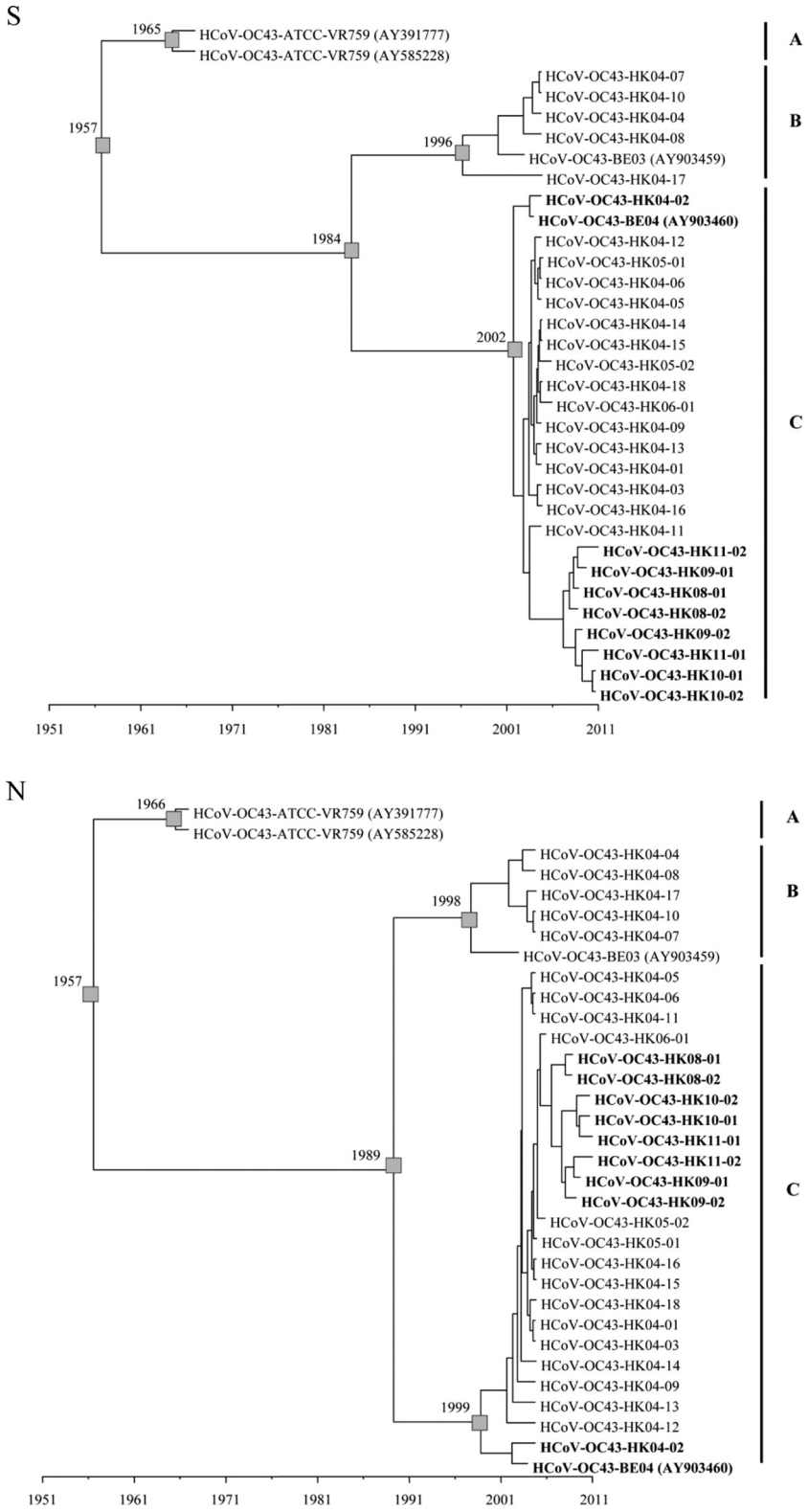


FIG. 5. Estimation of the time to the most recent common ancestor for HCoV-OC43 genotypes. The time-scaled phylogeny was summarized from all MCMC phylogenies of the S and N gene data set, which were analyzed under the relaxed-clock model with an uncorrelated exponential distribution in BEAST (version 1.6.1). A, genotype A; B, genotype B; C, genotype C. Genotype D strains are in boldface.

or genotypes. In feline coronavirus (FCoV), FCoV type II strains have been found to have originated from a double recombination between FCoV type I and canine coronavirus (CCoV) in the M gene (14). Novel CCoV type II strains also have been suggested recently to have originated from double recombination with porcine transmissible gastroenteritis virus in the S gene (6). In our previous study on the natural recombination of HCoV-HKU1 in the generation of different genotypes, the sites of major recombination were localized at nsp16 just upstream of the stop codon of ORF1ab, at the end of nsp6, and upstream of nsp5 (52). Recently, we also have demonstrated natural recombination events among different bat coronaviruses, including those which could have accounted for the emergence of civet SARSr-CoV (22). It was found that civet SARSr-CoV strain SZ3 was a potential recombinant of horseshoe bat SARSr-Rh-BatCoV strains Rp3 from Guangxi Province and Rf1 from Hubei Province, with a recombination breakpoint identified at the nsp16/S intergenic region (22). In the study on Rousettus bat coronavirus HKU9, which belongs to novel *Betacoronavirus* subgroup D, recombination was identified at nsp3 and at the nsp15/16 junction (21). In this study, possible recombination was observed mainly within ORF1ab at the nsp2/nsp3 junction, nsp12/nsp13 junction, and NS2a/HE junction of HCoV-OC43 genomes. ORF1ab is the region most susceptible to recombination in the coronavirus genomes. Further studies are required to better understand the role of recombination in various coronaviruses and the common sites of recombination in coronavirus genomes.

The present results also revealed previously undescribed features in the HCoV-OC43 genomes. Although the TRS preceding most ORFs in HCoV-OC43 genomes has been described previously, the putative TRS shared between NS5a and E suggested for the ATCC strain (39) was absent from the genomes of BE03, BE04, HK04-01, and HK04-02 as a result of a 12-nucleotide deletion. While we identified a potential alternative TRS, a 39-bp sequence homologous to the leader sequence also was found upstream of NS5a. Such a homologous sequence has been suggested as a compensation mechanism for the absence of TRS in HCoV-NL63 and HCoV-HKU1 (34, 35). Further experiments are required to ascertain the mechanism for the translation of NS5a and E protein in HCoV-OC43. The analysis of the K_a/K_s ratios for different regions of the HCoV-OC43 genome also revealed interesting findings. The K_a/K_s ratios were relatively high at nsp5, HE, S, and N regions of the HCoV-OC43 genome, suggesting that these regions were under higher selective pressure. The nsp5 of HCoV-OC43 encodes the chymotrypsin-like protease 3C-like protease ($3CL^{pro}$) that is important for the proteolytic cleavage of the large polyprotein encoded by ORF1ab. The $3CL^{pro}$ of SARS-CoV also has been found to induce cellular apoptosis (27). Both HE and S are major viral membrane glycoproteins which may be important for tissue tropism, cell attachment, and eliciting neutralizing antibodies (20, 59, 60). The coronavirus N protein, a highly phosphorylated protein required for viral replication, is also an immunogenic protein which elicits subgroup-specific antibody responses (12, 21, 51, 58). Interestingly, the K_a/K_s ratios at the S gene had dropped from 5.134 among genotype A strains isolated in the 1960s to 0.270 among the 29 HK strains from 2004 to 2011. This may reflect the rapid evolution of spike protein to adapt to a new host soon after interspecies transmission, as HCoV-OC43 was thought to have originated from zoonotic transmission, sharing a common ancestor with bovine coronavirus dating back to 1890 (47). Further studies on more ancient and contemporary strains are required to better understand the selective pressure at different regions of the HCoV-OC43 genome and its significance in terms of protein function and evolution.

ACKNOWLEDGMENTS

We are grateful for the generous support of Carol Yu, Richard Yu, Hui Hoy, and Hui Ming with the genomic sequencing platform.

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

REFERENCES

1. **Abraham, S., T. E. Kienzle, W. Lapps, and D. A. Brian.** 1990. Deduced sequence of the bovine coronavirus spike protein and identification of the internal proteolytic cleavage site. Virology **176:**296–301.
2. **Apweiler, R., et al.** 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. Nucleic Acids Res. **29:**37–40.
3. **Birch, C. J., et al.** 2005. Human coronavirus OC43 causes influenza-like illness in residents and staff of aged-care facilities in Melbourne, Australia. Epidemiol. Infect. **133:**273–277.
4. **Brian, D. A., and R. S. Baric.** 2005. Coronavirus genome structure and replication. Curr. Top. Microbiol. Immunol. **287:**1–30.
5. **Cheng, V. C., S. K. Lau, P. C. Woo, and K. Y. Yuen.** 2007. Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection. Clin. Microbiol. Rev. **20:**660–694.
6. **Decaro, N., et al.** 2009. Recombinant canine coronaviruses related to transmissible gastroenteritis virus of Swine are circulating in dogs. J. Virol. **83:** 1532–1537.
7. **Drummond, A. J., and A. Rambaut.** 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. **7:**214.
8. **Esposito, S., et al.** 2006. Impact of human coronavirus infections in otherwise healthy children who attended an emergency department. J. Med. Virol. **78:**1609–1615.
9. **Fouchier, R. A., et al.** 2004. A previously undescribed coronavirus associated with respiratory disease in humans. Proc. Natl. Acad. Sci. U. S. A. **101:**6212–6216.
10. **Fu, K., and R. S. Baric.** 1992. Evidence for variable rates of recombination in the MHV genome. Virology **189:**88–102.
11. **Gerna, G., et al.** 2006. Impact of human metapneumovirus and human cytomegalovirus versus other respiratory viruses on the lower respiratory tract infections of lung transplant recipients. J. Med. Virol. **78:**408–416.
12. **Gill, E. P., et al.** 1994. Development and application of an enzyme immunoassay for coronavirus OC43 antibody in acute respiratory illness. J. Clin. Microbiol. **32:**2372–2376.
13. **Guan, Y., et al.** 2003. Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. Science **302:**276–278.
14. **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.
15. **Huang, Y., S. K. Lau, P. C. Woo, and K. Y. Yuen.** 2008. CoVDB: a comprehensive database for comparative analysis of coronavirus genes and genomes. Nucleic Acids Res. **36:**D504–D511.
16. **Jendrach, M., V. Thiel, and S. Siddell.** 1999. Characterization of an internal ribosome entry site within mRNA 5 of murine hepatitis virus. Arch. Virol. **144:**921–933.
17. **Keane, T. M., C. J. Creevey, M. M. Pentony, T. J. Naughton, and J. O. Mclnerney.** 2006. Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. **6:**29.
18. **Krempl, C., B. Schultze, and G. Herrler.** 1995. Analysis of cellular receptors for human coronavirus OC43. Adv. Exp. Med. Biol. **380:**371–374.
19. **Kuypers, J., et al.** 2007. Clinical disease in children associated with newly described coronavirus subtypes. Pediatrics **119:**e70–e76.
20. **Langereis, M. A., A. L. van Vliet, W. Boot, and R. J. de Groot.** 2010. Attachment of mouse hepatitis virus to O-acetylated sialic acid is mediated

by hemagglutinin-esterase and not by the spike protein. J. Virol. **84:**8970–8974.
21. **Lau, S. K. P., et al.** 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.
22. **Lau, S. K., et al.** 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.
23. **Lau, S. K., et al.** 2006. Coronavirus HKU1 and other coronavirus infections in Hong Kong. J. Clin. Microbiol. **44:**2063–2071.
24. **Lau, S. K., et al.** 2005. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc. Natl. Acad. Sci. U. S. A. **102:**14040–14045.
25. **Lau, S. K., et al.** 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.
26. **Li, W., et al.** 2005. Bats are natural reservoirs of SARS-like coronaviruses. Science **310:**676–679.
27. **Lin, C. W., K. H. Lin, T. H. Hsieh, S. Y. Shiu, and J. Y. Li.** 2006. Severe acute respiratory syndrome coronavirus 3C-like protease-induced apoptosis. FEMS Immunol. Med. Microbiol. **46:**375–380.
28. **Makino, S., J. G. Keck, S. A. Stohlman, and M. M. Lai.** 1986. High-frequency RNA recombination of murine coronaviruses. J. Virol. **57:**729–737.
29. **McIntosh, K., et al.** 1970. Seroepidemiologic studies of coronavirus infection in adults and children. Am. J. Epidemiol. **91:**585–592.
30. **McIntosh, K., W. B. Becker, and R. M. Chanock.** 1967. Growth in suckling-mouse brain of "IBV-like" viruses from patients with upper respiratory tract disease. Proc. Natl. Acad. Sci. U. S. A. **58:**2268–2273.
31. **Mihindukulasuriya, K. A., G. Wu, J. St. Leger, R. W. Nordhausen, and D. Wang.** 2008. Identification of a novel coronavirus from a beluga whale by using a panviral microarray. J. Virol. **82:**5084–5088.
32. **Patrick, D. M., et al.** 2006. An outbreak of human coronavirus OC43 infection and serological cross-reactivity with SARS coronavirus. Can. J. Infect. Dis. Med. Microbiol. **17:**330–336.
33. **Peiris, J. S., et al.** 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet **361:**1319–1325.
34. **Pyrc, K., et al.** 2010. Culturing the unculturable: human coronavirus HKU1 infects, replicates, and produces progeny virions in human ciliated airway epithelial cell cultures. J. Virol. **84:**11255–11263.
35. **Pyrc, K., M. F. Jebbink, B. Berkhout, and L. van der Hoek.** 2004. Genome structure and transcriptional regulation of human coronavirus NL63. Virol. J. **1:**7.
36. **Pyrc, K., et al.** 2006. Mosaic structure of human coronavirus NL63, one thousand years of evolution. J. Mol. Biol. **364:**964–973.
37. **Rota, P. A., et al.** 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science **300:**1394–1399.
38. **Sloots, T. P., et al.** 2006. Evidence of human coronavirus HKU1 and human bocavirus in Australian children. J. Clin. Virol. **35:**99–102.
39. **St.-Jean, J. R., et al.** 2004. Human respiratory coronavirus OC43: genetic stability and neuroinvasion. J. Virol. **78:**8824–8834.
40. **Suchard, M. A., R. E. Weiss, and J. S. Sinsheimer.** 2001. Bayesian selection of continuous-time Markov chain evolutionary models. Mol. Biol. Evol. **18:**1001–1013.
41. **Tamura, K., J. Dudley, M. Nei, and S. Kumar.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24:**1596–1599.
42. **Tang, X. C., et al.** 2006. Prevalence and genetic diversity of coronaviruses in bats from China. J. Virol. **80:**7481–7490.
43. **Vabret, A., et al.** 2006. Inter- and intra-variant genetic heterogeneity of human coronavirus OC43 strains in France. J. Gen. Virol. **87:**3349–3353.
44. **Vabret, A., T. Mourez, S. Gouarin, J. Petitjean, and F. Freymuth.** 2003. An outbreak of coronavirus OC43 respiratory infection in Normandy, France. Clin. Infect. Dis. **36:**985–989.
45. **van der Hoek, L., et al.** 2004. Identification of a new human coronavirus. Nat. Med. **10:**368–373.
46. **van der Most, R. G., L. Heijnen, W. J. Spaan, and R. J. de Groot.** 1992. Homologous RNA recombination allows efficient introduction of site-specific mutations into the genome of coronavirus MHV-A59 via synthetic co-replicating RNAs. Nucleic Acids Res. **20:**3375–3381.
47. **Vijgen, L., et al.** 2005. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. J. Virol. **79:**1595–1604.
48. **Vijgen, L., et al.** 2005. Circulation of genetically distinct contemporary human coronavirus OC43 strains. Virology **337:**85–92.
49. **Vijgen, L., et al.** 2006. Evolutionary history of the closely related group 2 coronaviruses: porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, and human coronavirus OC43. J. Virol. **80:**7270–7274.
50. **Woo, P. C., et al.** 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.
51. **Woo, P. C., et al.** 2004. False-positive results in a recombinant severe acute respiratory syndrome-associated coronavirus (SARS-CoV) nucleocapsid enzyme-linked immunosorbent assay due to HCoV-OC43 and HCoV-229E rectified by Western blotting with recombinant SARS-CoV spike polypeptide. J. Clin. Microbiol. **42:**5885–5888.
52. **Woo, P. C., et al.** 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.
53. **Woo, P. C., et al.** 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J. Virol. **79:**884–895.
54. **Woo, P. C., et al.** 2009. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. J. Virol. **83:**908–917.
55. **Woo, P. C., et al.** 2004. Relative rates of non-pneumonic SARS coronavirus infection and SARS coronavirus pneumonia. Lancet **363:**841–845.
56. **Woo, P. C., et al.** 2005. Clinical and molecular epidemiological features of coronavirus HKU1-associated community-acquired pneumonia. J. Infect. Dis. **192:**1898–1907.
57. **Woo, P. C., et al.** 2006. Molecular diversity of coronaviruses in bats. Virology **351:**180–187.
58. **Wu, C. H., et al.** 2009. Glycogen synthase kinase-3 regulates the phosphorylation of severe acute respiratory syndrome coronavirus nucleocapsid protein and viral replication. J. Biol. Chem. **284:**5229–5239.
59. **Yokomori, K., et al.** 1995. Neuropathogenicity of mouse hepatitis virus JHM isolates differing in hemagglutinin-esterase protein expression. J. Neurovirol. **1:**330–339.
60. **Yoo, D., et al.** 1992. Synthesis and processing of the haemagglutinin-esterase glycoprotein of bovine coronavirus encoded in the E3 region of adenovirus. J. Gen. Virol. **73:**2591–2600.