JOURNAL OF VIROLOGY, July 2006, p. 7136–7145 Vol. 80, No. 14
0022-538X/06/$08.00+0 doi:10.1128/JVI.00509-06
Copyright © 2006, American Society for Microbiology. All Rights Reserved.

Comparative Analysis of 22 Coronavirus HKU1 Genomes Reveals a Novel Genotype and Evidence of Natural Recombination in Coronavirus HKU1

Patrick C. Y. Woo,[1,2,3]† Susanna K. P. Lau,[1,2,3]† Cyril C. Y. Yip,[1] Yi Huang,[1] Hoi-Wah Tsoi,[1] Kwok-Hung Chan,[1] and Kwok-Yung Yuen[1,2,3]*

Department of Microbiology[1] and Research Centre of Infection and Immunology,[2] Faculty of Medicine, The University of Hong Kong, and State Key Laboratory of Emerging Infectious Diseases, The University of Hong Kong, Hong Kong[3]

Received 10 March 2006/Accepted 25 April 2006

We sequenced and compared the complete genomes of 22 strains of coronavirus HKU1 (CoV HKU1) obtained from nasopharyngeal aspirates of patients with respiratory tract infections over a 2-year period. Phylogenetic analysis of 24 putative proteins and polypeptides showed that the 22 CoV HKU1 strains fell into three clusters (genotype A, 13 strains; genotype B, 3 strains and genotype C, 6 strains). However, different phylogenetic relationships among the three clusters were observed in different regions of their genomes. From nsp4 to nsp6, the genotype A strains were clustered with the genotype B strains. For nsp7 and nsp8 and from nsp10 to nsp16, the genotype A strains were clustered with the genotype C strains. From hemagglutinin esterase (HE) to nucleocapsid (N), the genotype B strains were clustered closely with the genotype C strains. Bootscan analysis showed possible recombination between genotypes B and C from nucleotide positions 11500 to 13000, corresponding to the nsp6-nsp7 junction, giving rise to genotype A, and between genotypes A and B from nucleotide positions 21500 to 22500, corresponding to the nsp16-HE junction, giving rise to genotype C. Multiple alignments further narrowed the sites of crossover to a 143-bp region between nucleotide positions 11750 and 11892 and a 29-bp region between nucleotide positions 21502 and 21530. Genome analysis also revealed various numbers of tandem copies of a perfect 30-base acidic tandem repeat (ATR) which encodes NDDEDVVTGD and various numbers and sequences of imperfect repeats in the N terminus of nsp3 inside the acidic domain upstream of papain-like protease 1 among the 22 genomes. All 10 CoV HKU1 strains with incomplete imperfect repeats (1.4 and 4.4) belonged to genotype A. The present study represents the first evidence for natural recombination in coronavirus associated with human infection. Analysis of a single gene is not sufficient for the genotyping of CoV HKU1 strains but requires amplification and sequencing of at least two gene loci, one from nsp10 to nsp16 (e.g., *pol* or helicase) and another from HE to N (e.g., spike or N). Further studies will delineate whether the ATR is useful for the molecular typing of CoV HKU1.

The recent severe acute respiratory syndrome (SARS) epidemic, the discovery of SARS coronavirus (CoV), and the identification of SARS CoV-like viruses from Himalayan palm civets and a raccoon dog from wild-animal live markets in mainland China have led to a boost in interest in the discovery of novel coronaviruses in both humans and animals (8, 23, 26, 28, 40, 42). In 2004, a novel group 1 human coronavirus (HCoV), NL63, was reported independently by two groups (6, 34). In 2005, we described the discovery, complete genome sequence, clinical features, and molecular epidemiology of a novel group 2 human coronavirus, HKU1 (genotype A) (17, 37–39, 41). This virus has also subsequently been found in patients with respiratory tract infections in other countries (1, 30, 33). Recently, we have also identified a 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 the Hong Kong Special Administrative Region (16, 27). The discovery of SARS CoV-like viruses in horseshoe bats was confirmed by another group in other provinces in China (19).

As a result of the unique mechanism of viral replication, coronaviruses have a high frequency of recombination (15). Their tendency for recombination and high mutation rates may allow them to adapt to new hosts and ecological niches. However, no convincing evidence among human coronaviruses of genetic recombination that may have contributed to their ability to reinfect humans has been documented. In our study of the phylogeny of the RNA-dependent RNA polymerase (*pol*), spike (S), and nucleocapsid (N) genes of nine isolates of CoV HKU1 recovered from patients with pneumonia, it was discovered that the sequences of the S and N genes fell into two distinct genotypes, with seven strains belonging to genotype A and two belonging to genotype B (41). On the other hand, for the *pol* gene, one of the two "genotype B" strains as determined by its S and N sequences (from patient 8) was clustered with the other seven "genotype A" strains (41). Furthermore, the same phenomenon was also observed in our subsequent prospective study of CoV HKU1-associated respiratory tract infections (17). Based on these observations, we suspected that there is an additional CoV

* Corresponding author. Mailing address: 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. Woo and S.K.P. Lau contributed equally to this paper.

TABLE 1. Characteristics of the 22 CoV HKU1 strains used in this study

Strain	Source	Mo/yr of detection	Patient characteristic					CoV HKU1 characteristic		
			Sex[a]	Age (yr)	Upper/lower respiratory tract infection	Underling disease	Clinical outcome	Genotype	No. of NDDEDVVTGD repeats	No. of imperfect repeats
N2	Patient 1 in reference 41	Mar/03	F	35	Lower	Absent	Survived	B	11	2
N3	Patient 2 in reference 41	Apr/03	M	66	Lower	Present	Died	A	14	2
N1	Patient 5 in reference 41	Jan/04	M	71	Lower	Present	Survived	A	14	2
N5	Patient 8 in reference 41	Jan/04	M	68	Lower	Absent	Survived	C	8	3
N6	Unpublished	Jan/04	F	29	Upper	Present	Survived	A	2	1.4
N7	Patient 4 in reference 41	Jan/04	M	75	Lower	Present	Survived	A	12	1.4
N9	Patient 9 in reference 41	Mar/04	F	83	Lower	Present	Survived	A	12	1.4
N10	Patient 10 in reference 41	Mar/04	M	72	Lower	Present	Died	A	13	1.4
N11	Patient 1 in reference 17	Apr/04	F	2	Upper	Absent	Survived	A	15	1.4
N13	Patient 2 in reference 17	May/04	F	7	Upper	Present	Survived	A	9	1.4
N14	Patient 3 in reference 17	Jul/04	M	84	Upper	Present	Survived	A	10	1.4
N15	Patient 4 in reference 17	Nov/04	M	3	Upper	Present	Survived	B	15	1
N16	Patient 5 in reference 17	Nov/04	M	87	Lower	Present	Survived	C	10	4
N17	Patient 6 in reference 17	Nov/04	F	4	Upper	Present	Survived	C	10	2
N18	Patient 7 in reference 17	Nov/04	M	2	Lower	Absent	Survived	A	13	1.4
N19	Patient 8 in reference 17	Dec/04	M	19	Upper	Absent	Survived	A	11	3
N20	Patient 9 in reference 17	Dec/04	M	3	Upper	Absent	Survived	C	10	4
N21	Patient 10 in reference 17	Dec/04	F	9	Upper	Present	Survived	C	10	3
N22	Patient 11 in reference 17	Jan/05	F	3	Upper	Absent	Survived	C	13	3
N24	Patient 12 in reference 17	Jan/05	M	5	Upper	Present	Survived	A	17	4.4
N23	Patient 13 in reference 17	Feb/05	M	4	Upper	Present	Survived	A	11	1.4
N25	Unpublished	Feb/05	F	5 mo	Upper	Absent	Survived	B	12	2

[a] M, male; F, female.

HKU1 genotype which has arisen from recombination between genotypes A and B of CoV HKU1.

To test this hypothesis, we performed complete genome sequencing on 21 additional strains of CoV HKU1 and compared their genomes to the CoV HKU1 genotype A strain (38). The sites of recombination were identified, and a novel CoV HKU1 genotype, genotype C, was defined.

MATERIALS AND METHODS

CoV HKU1 strains. All 22 CoV HKU1 strains were isolated from patients with respiratory tract infections in Hong Kong in a 2-year period (March 2003 to February 2005) (Table 1) (17, 38, 41).

RNA extraction. Viral RNA was extracted from the nasopharyngeal aspirates of the patients using a QIAamp viral RNA mini kit (QIAGEN, Hilden, Germany). The RNA pellet was resuspended in 10 μl of DNase-free, RNase-free double-distilled water and was used as the template for reverse transcription-PCR.

Complete genome sequencing and genome analysis. The complete genome sequence of the CoV HKU1 genotype A strain was described previously (GenBank accession no. NC_006577) (38). The complete genomes of the other 21 CoV HKU1 strains were amplified and sequenced using the RNA extracted from the nasopharyngeal aspirates of the corresponding patients as the template, by a strategy described previously (38). The RNA was converted to cDNA by a combined random-priming and oligo(dT) priming strategy. The 5′ ends of the viral genomes were confirmed by rapid amplification of cDNA ends using the 5′/3′ RACE kit (Roche, Germany). Sequences were assembled and manually edited to produce final sequences of the viral genomes. The 21 genomes were compared to that of the CoV HKU1 genotype A strain and were manually annotated.

Phylogenetic-tree construction. The nucleotide sequences for nsp1, nsp2, conserved portions of nsp3 (including papain-like protease 1 [PL1^{pro}], a member of the Appr-1-p processing enzyme family [A1pp], papain-like protease 2 [PL2^{pro}], and the hydrophobic domain [HD]), nsp4-nsp10, nsp12-nsp16, hemagglutinin esterase (HE), S, open reading frame 4 (ORF4), envelope (E), membrane (M), and N were extracted from the 22 CoV HKU1 genomes. Phylogenetic-tree construction was performed using the neighbor-joining method with ClustalX 1.83. The corresponding nucleotide sequences of human coronavirus OC43 (GenBank accession no. AY585229) were used as outgroups.

Bootscan analysis. To perform bootscan analysis, a nucleotide alignment of the genome sequences of one genotype A (38), one genotype B (patient 1 of reference 41), and one genotype C (patient 8 of reference 41) strain of CoV HKU1 and one HCoV OC43 strain (GenBank accession no. AY585229) was generated by ClustalX, version 1.83, and edited manually. Bootscan analysis was performed using Simplot version 3.5.1 (F84 model; window size, 1,000 bp; step, 200 bp) (20), with the genome sequence of HCoV OC43 as a query.

Nucleotide sequence accession numbers. The nucleotide sequences of the 21 additional genomes of CoV HKU1 strains (data not shown) have been lodged within the GenBank sequence database under accession no. AY884001, DQ339101, DQ415896, DQ415897, DQ415898, DQ415899, DQ415900, DQ415901, DQ415902, DQ415903, DQ415904, DQ415905, DQ415906, DQ415907, DQ415908, DQ415909, DQ415910, DQ415911, DQ415912, DQ415913, and DQ415914.

RESULTS

Complete genome sequence, genome organization, phylogenetic analysis, and genotypes. The sizes of the genomes of the 22 CoV HKU1 strains ranged from 29,295 to 30,097 nucleotides. The G+C contents of all 22 genomes are 32%. The overall genome organizations of the 22 CoV HKU1 strains were the same (Fig. 1A).

Phylogenetic trees using the nucleotide sequences of genes for putative proteins and polypeptides (nsp1, nsp2, conserved portions of nsp3 [PL1^{pro}, A1pp, PL2^{pro}, and HD], nsp4-nsp10, nsp12-nsp16, HE, S, ORF4, E, M, and N) of the 22 CoV HKU1 strains were constructed and are shown in Fig. 1B. In 18 of the 24 trees, the 22 CoV HKU1 strains fell clearly into three clusters, named genotype A (13 strains), genotype B (3 strains), and genotype C (6 strains). The exceptions are the five trees constructed using nsp1, nsp2, PL1^{pro}, PL2^{pro}, and HD, in which the differences among the sequences were too small, and the nsp10 tree, in which two genotype A strains, N1 and N3, were clustered with genotype C strains.

The three genotypes exhibited different relationships to each

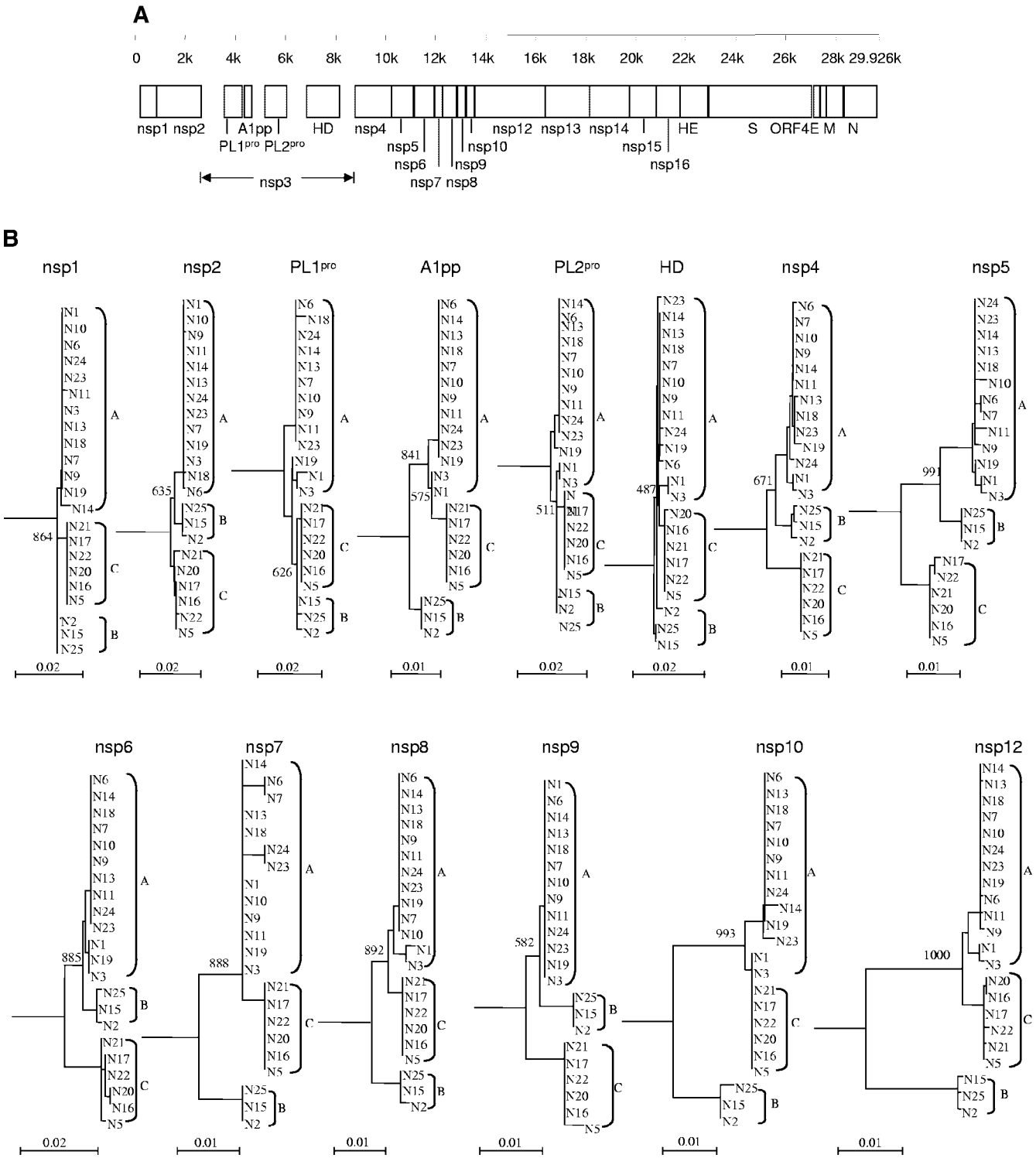

FIG. 1. Genome organization and phylogenetic analysis of CoV HKU1. (A) Genome organization of CoV HKU1. The regions of the genome used for phylogenetic tree construction are labeled. (B) Phylogenetic analysis of nsp1, nsp2, conserved portions of nsp3 ($PL1^{pro}$, A1pp, $PL2^{pro}$, and HD), nsp4-nsp10, nsp12-nsp16, HE, S, ORF4, E, M, and N of the 22 CoV HKU1 genomes. The trees were constructed by the neighbor-joining method using Jukes-Cantor correction and bootstrap values calculated from 1,000 trees. Seven hundred forty, 1,821, 655, 333, 895, 1,263, 1,488, 909, 861, 276, 582, 330, 411, 2,783, 1,809, 1,563, 1,125, 900, 1,276, 4,126, 330, 253, 687, and 1,358 nucleotide positions in nsp1, nsp2, $PL1^{pro}$, A1pp, $PL2^{pro}$, HD, nsp4, nsp5, nsp6, nsp7, nsp8, nsp9, nsp10, nsp12, nsp13, nsp14, nsp15, nsp16, HE, S, ORF4, 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 OC43 were used as the outgroups. A, genotype A; B, genotype B and C, genotype C.

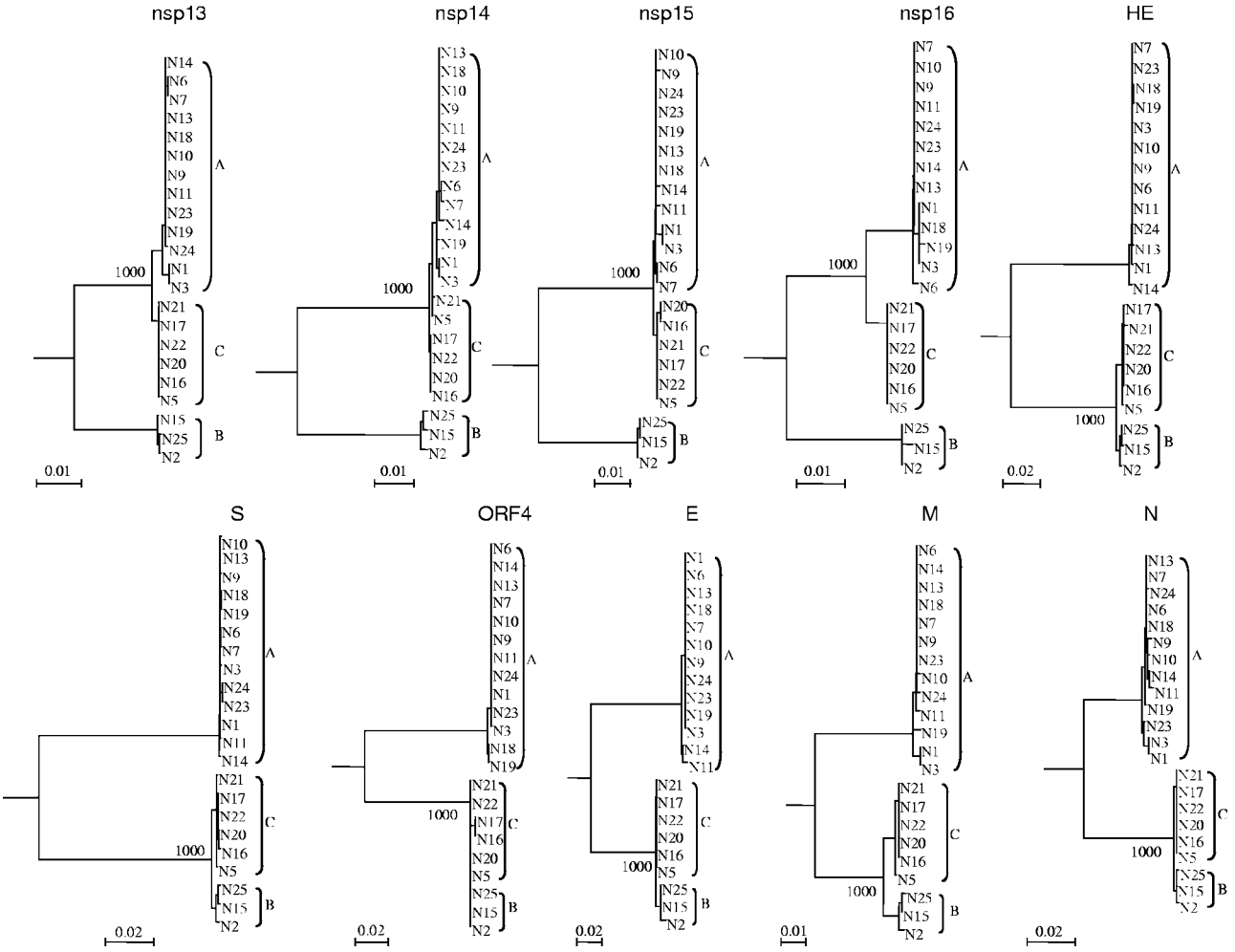

FIG. 1—*Continued.*

other in different regions of their genomes. From nsp4 to nsp6, the genotype A strains were clustered with the genotype B strains, but for nsp7 and nsp8, the genotype A strains were clustered with the genotype C strains. From nsp10 to nsp16, the genotype A strains were clustered closely with the genotype C strains, with high bootstrap values, but from HE to N, the genotype B strains were clustered closely with the genotype C strains, with bootstrap values of 1,000 in all cases. No association was observed between the genotypes and the time of detection or the age, sex, clinical disease, presence of underlying disease, or outcome of the patients (Table 1).

The putative transcription regulatory sequence motif, 5′-AAUCUAAAC-3′ (as in mouse hepatitis virus [MHV] and bovine coronavirus) (22) or, alternatively, 5′-UAAAUCUAAAC-3′, that was found at the 3′ end of the leader sequence and precedes each translated ORF except ORF5 described to occur in the genome of the CoV HKU1 genotype A strain, was also present in all of the other 21 CoV HKU1 genomes. On the other hand, the sequence of the putative internal ribosomal entry site (IRES) (32) for the ORF of the envelope protein in the genomes of all three CoV HKU1 genotype B strains and all six genotype C CoV HKU1 strains were UUUUAUCGCUUGG, instead of AUUUAUUGUUUGG in all 13 CoV HKU1 genotype A strains, although both sequences were similar to the IRES element, UUUUAUUCUUUUU, in MHV (10).

The 22 genomes differed in their numbers of tandem copies of the 30-base acidic tandem repeat (ATR) in the N terminus of nsp3 inside the acidic domain upstream of $PL1^{pro}$ (Tables 1 and 2). All 22 genomes had tandem copies of a perfect 30-base repeat which encodes NDDEDVVTGD and various numbers of imperfect repeats. The median number of tandem copies of the perfect 30-base repeat was 11.5 (range, 2 to 17), and the median number of imperfect repeats was 2 (range, 1 to 4). All of the 10 CoV HKU1 strains with incomplete imperfect repeats (1.4 and 4.4) belonged to genotype A.

Bootscan analysis. Bootscan analysis showed that from the 5′ end of the genome to position 12000, there could be a number of possible recombination sites in the genomes of the three genotypes (Fig. 2). Right upstream to position 11500, high bootstrap support for clustering between the CoV HKU1 genotype A strain and the CoV HKU1 genotype B strain was observed. From positions 13000 to 21500, high bootstrap support for clustering between the CoV HKU1 genotype A strain and the CoV HKU1 genotype C strain was observed. From

TABLE 2. Amino acid sequences of acidic tandem repeats of the 22 CoV HKU1 strains[a]

Genotype	Strain	Amino acid sequence[a]
A	N1	$(NDDEDVVTGD)_{14}$(NNDEEIVTGD) (NDDQIVVTGD)
	N3	$(NDDEDVVTGD)_{14}$(NNDEEIVTGD) (NDDQIVVTGD)
	N6	$(NDDEDVVTGD)_{2}$(NDDD) (NDDQIVVIGD)
	N7	$(NDDEDVVTGD)_{12}$(NDDD) (NDDQIVVIGD)
	N9	$(NDDEDVVTGD)_{12}$(NDDD) (NDDQIVVIGD)
	N10	$(NDDEDVVTGD)_{13}$(NDDD) (NDDQIVVIGD)
	N11	$(NDDEDVVTGD)_{15}$(NDDD) (NDDQIVVIGD)
	N13	$(NDDEDVVTGD)_{9}$(NDDD) (NDDQIVVIGD)
	N14	$(NDDEDVVTGD)_{10}$(NDDD) (NDDQIVVIGD)
	N18	$(NDDEDVVTGD)_{13}$(NDDD) (NDDQIVVIGD)
	N19	$(NDDEDVVTGD)_{10}$(NNDEEIVTGD) (NDDEDVVTGD)$_{1}$(NNDEEIVTGD) (NDDQIVVTGD)
	N23	$(NDDEDVVTGD)_{11}$(NDDD) (NDDQIVVIGD)
	N24	$(NDDEDVVTGD)_{1}$(NDDEHVVTGD) (NDDEHVVTGD) $(NDDEDVVTGD)_{9}$(NDDEHVVTGD) $(NDDEDVVTGD)_{7}$(NDDD) (NDDQIVVIGD)
B	N2	$(NDDEDVVTGD)_{11}$(NDDEEIVTGD) (NDDQIVVTGD)
	N15	$(NDDEDVVTGD)_{15}$(ND-QIVVTGD)
	N25	$(NDDEDVVTGD)_{12}$(NDDEEIVTGD) (NDDQIVVTGD)
C	N5	$(NDDEDVVTGD)_{8}$(NNDEDVVTGD) (NNDEESVTGD) (NDDQIVVTGD)
	N16	$(NDDEDVVTGD)_{10}$(NNDEDVVTGD) (NNGEDVVTGD) (NNDEESVTGD) (NDDQIVVTGD)
	N17	$(NDDEDVVTGD)_{10}$(NNDEESVTGD) (NDDQIVVTGD)
	N20	$(NDDEDVVTGD)_{10}$(NNDEDVVTGD) (NNGEDVVTGD) (NNDEESVTGD) (NDDQIVVTGD)
	N21	$(NDDEDVVTGD)_{10}$(NNDEDVVTGD) (NNDEESVTGD) (NDDQIVTGD) (NDDQIVVTGD)
	N22	$(NDDEDVVTGD)_{13}$(NNDEDVVTGD) (NNDEESVTGD) (NDDQIVVTGD)

[a] The amino acids underlined denote those that are different from the NDDEDVVTGD tandem repeats.

position 22500 to the 3′ end of the genome, high bootstrap support for clustering between the CoV HKU1 genotype B strain and CoV HKU1 genotype C strain was observed. These findings indicate that recombination has possibly taken place between nucleotide positions 11500 and 13000, corresponding to the nsp6-nsp7 junction, and between nucleotide positions 21500 and 22500, corresponding to the nsp16-HE junction.

Comparative sequence analysis of the nsp6-nsp7 junction and nsp13-HE gene junction. Since both phylogenetic trees and bootscan analysis showed that there was a possible recombination site at the nsp6-nsp7 junction and the nsp16-HE gene junction, multiple alignments among the nucleotide sequences of the 22 genomes were performed to ascertain the exact sites of recombination.

Upstream of nucleotide position 11750 of the CoV HKU1 genotype A genome (227 bases before the end of nsp6), there was high nucleotide identity between the sequences of the CoV HKU1 genotype A and genotype B strains, whereas downstream to nucleotide position 11892 of the CoV HKU1 genotype A genome (85 bases before the end of nsp6), there was high nucleotide identity between the sequences of the CoV HKU1 genotype A and genotype C strains (Fig. 3). This indicates that the site of crossover was probably within a 143-bp region between nucleotide positions 11750 and 11892.

Upstream of nucleotide position 21502 of the CoV HKU1 genotype A genome (249 bases before the stop codon of ORF1ab), there was high nucleotide identity between the sequences of the CoV HKU1 genotype A and genotype C strains, whereas downstream of nucleotide position 21530 of the CoV HKU1 genotype A genome (221 bases before the stop codon of ORF1ab), there was high nucleotide identity between the sequences of the CoV HKU1 genotype B and genotype C strains, including a 13-bp insertion just downstream of the stop codon of ORF1ab (Fig. 4). This indicates that the site of crossover was probably within a 29-bp region between nucleotide positions 21502 and 21530.

DISCUSSION

This is the first time that evidence for natural recombination is documented for coronavirus associated with human infection. Coronaviruses are unique in having a high frequency of homologous RNA recombination, as a result of random template switching during RNA replication, thought to be mediated by a "copy choice" mechanism (2, 4, 13, 14, 21, 35). In feline coronavirus (FCoV), it has been documented that FCoV type II strains originated from a double recombination between FCoV type I and canine coronavirus, and the site of recombination has been pinpointed to a region of about 50 nucleotides in the M gene by multiple alignment (9). As for recombination between different strains of MHV, in vitro studies have shown variations in both sites and rates of recombination, with the S gene having a frequency threefold that of the polymerase gene (7, 21). In the present study, by comparing the sequences of 22 complete genomes of CoV HKU1 strains, we documented that major recombination has occurred among the three CoV HKU1 genotypes. Both phylogenetic and bootscan analysis showed that the nucleotide sequences of the six genotype C strains were almost identical to those of the 13 genotype A strains from nsp10 to nsp16 (Fig. 1B and 4). Interestingly, the topologies of the phylogenetic trees changed dramatically starting from the HE gene. From HE to N, the nucleotide sequences of the six genotype C strains were almost identical to those of the three genotype B strains (Fig. 1B). This is also in

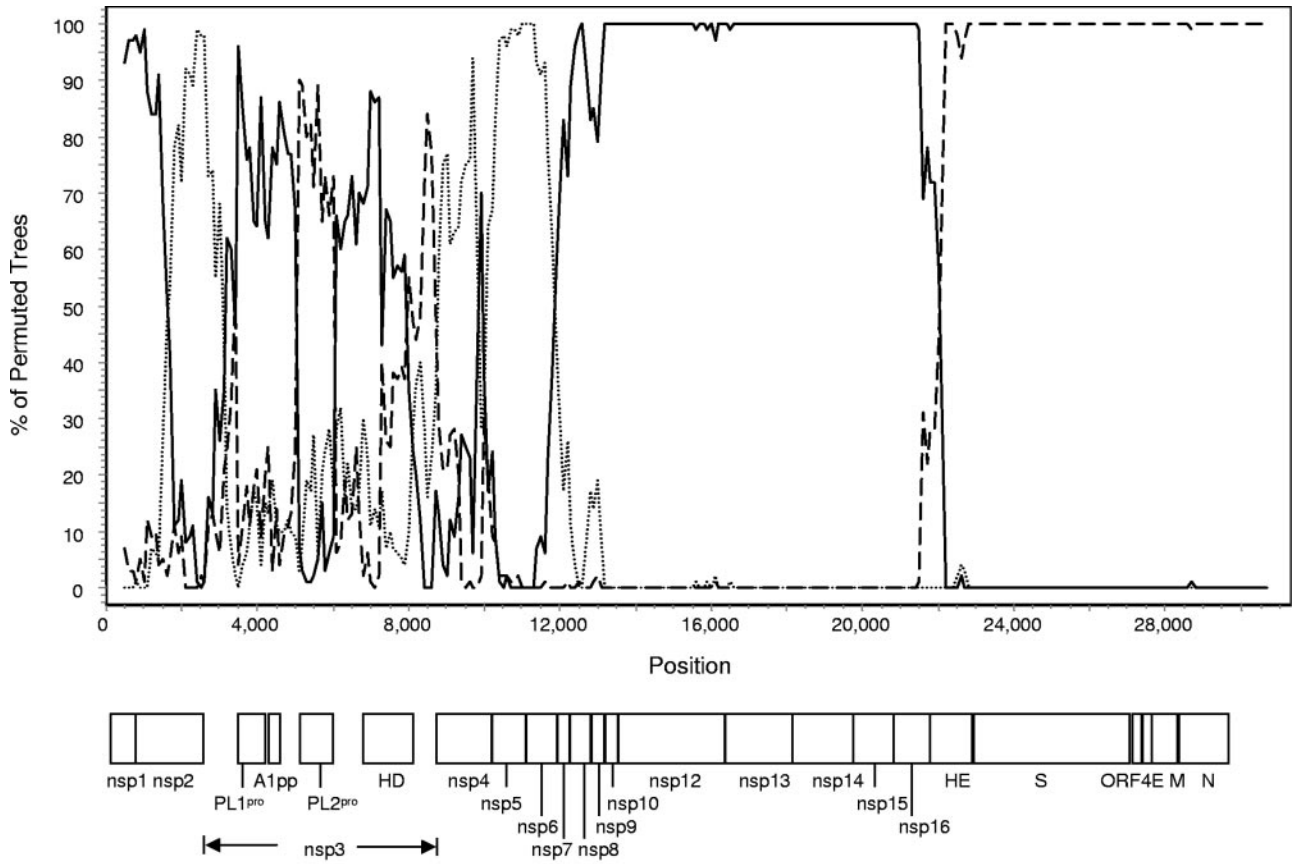

FIG. 2. Bootscan analysis of the CoV HKU1 genomes. Bootscanning was conducted with Simplot version 3.5.1 (F84 model; window size, 1,000 bp; step, 200 bp) on a gapless nucleotide alignment, generated with ClustalX, with the genome sequence of HCoV OC43 (GenBank accession no. AY585229) as the query sequence. The dashed line denotes CoV HKU1 genotype A (NC_006577), the solid line denotes CoV HKU1 genotype B (AY884001), and the dotted line denotes CoV HKU1 genotype C (DQ339101).

line with results of bootscan analysis, suggesting recombination between genotypes A and B, giving rise to genotype C (Fig. 2). Multiple alignments of the nucleotide sequences of the nsp16-HE regions of the three genotypes confirmed our suspicion, and results localized the site of recombination to a stretch of 29 nucleotides in nsp16, just upstream to the stop codon of ORF1ab (Fig. 4). This is in keeping with the finding that the putative IRES of all the genotype B and genotype C strains were the same but different from those of the genotype A strains, as it is located downstream of HE. In addition to the recombination site in nsp16, there was another one at the end of nsp6, also evidenced by a shift in clustering in the phylogenetic trees, bootscan analysis, and multiple-alignment results (Fig. 1B, 2 and 3). In contrast to the nsp16 recombination site, recombination has occurred between genotypes B and C in this region, giving rise to genotype A. Furthermore, as shown in bootscan and phylogenetic analyses, additional recombination events might have occurred in ORF1ab upstream of nsp5 (Fig. 1B and 2). However, due to the relatively small variations in the sequences among the three genotypes, these putative recombination sites were difficult to ascertain with multiple alignments.

A novel genotype, genotype C, of CoV HKU1 is defined. It has been well known that recombination is an important mechanism for the generation and evolution of virus genotypes (12, 29, 31). In our previous study, we showed that seven of the nine CoV HKU1 strains were of genotype A and one of the nine strains was of genotype B by *pol*, S gene, and N gene sequence analysis (41). In the present study, we showed that the latter half of the genomes of the six genotype C strains probably represents a result of recombination between genotypes A and B. Analysis of the complete genomes of more CoV HKU1 strains from other countries will reveal the relative prevalences of the different genotypes in different localities. From the results of the present study, no association was observed between the genotypes and clinical characteristics of the patients. Furthermore, amplification and sequencing of a single gene is not sufficient to define the genotype of CoV HKU1. It would require amplification and sequencing of at least two gene loci, one from nsp10 to nsp16 (e.g., *pol* or helicase) and another from HE to N (e.g., S or N).

The origin and function of the ATR located inside the acidic domain upstream of $PL1^{pro}$, unique to CoV HKU1, remain enigmatic. Significant variations were observed among the ATRs of CoV HKU1 strains. Only two pairs of CoV HKU1 strains (N1 and N3, N7, and N9) possessed the same nucleotide sequence in their ATRs. No relationship was found between the number of repeats and the genotype or virulence of

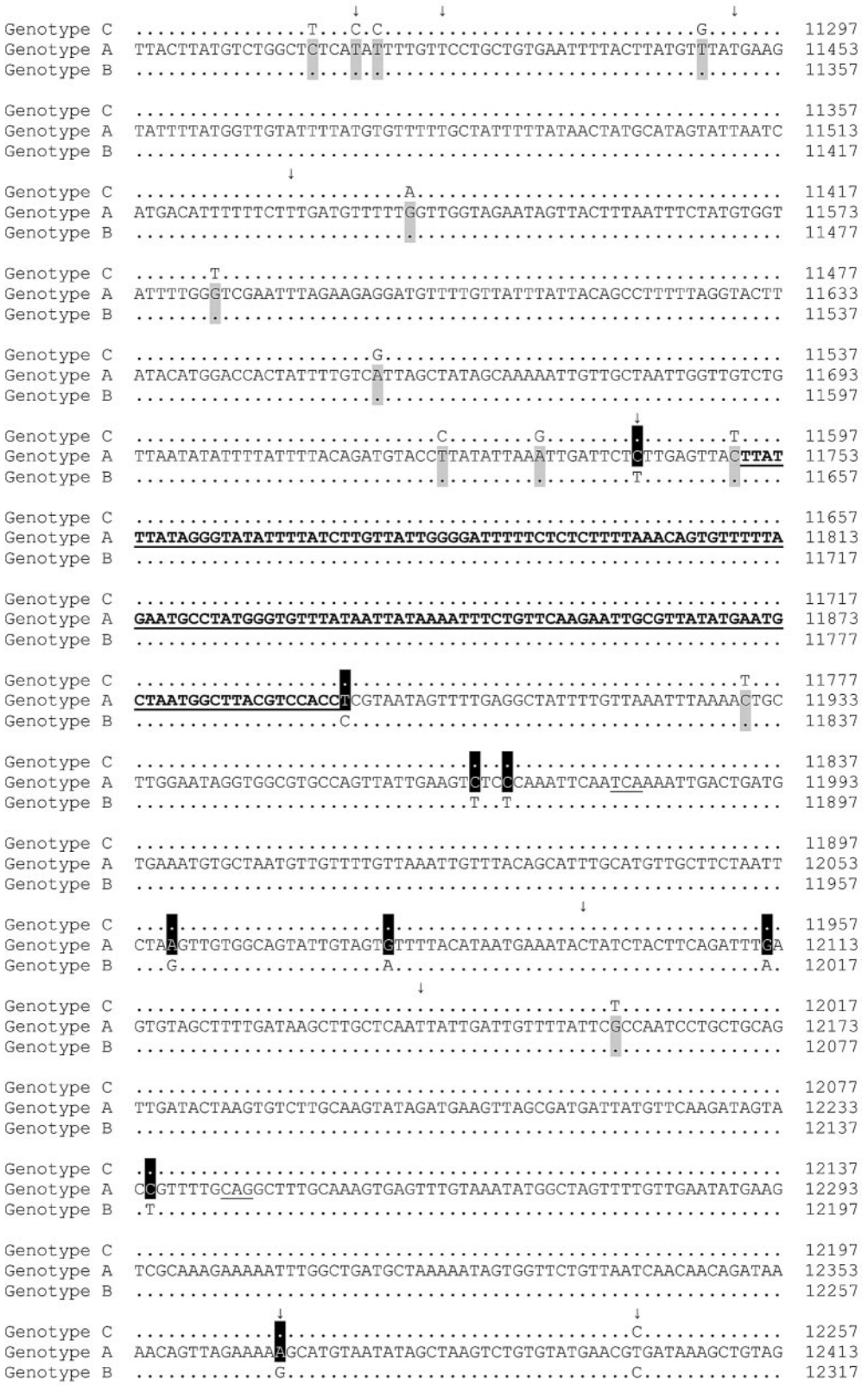

FIG. 3. Comparative sequence analysis of the nsp6-nsp7 junction. Multiple alignment of the nucleotide sequences of CoV HKU1 genotypes A, B, and C. In CoV HKU1 genotype B and CoV HKU1 genotype C, only the nucleotides differing from those in CoV HKU1 genotype A are depicted. The nucleotides in CoV HKU1 genotype C that are the same as those in CoV HKU1 genotype A but different from those in CoV HKU1 genotype B are highlighted in black, and those in CoV HKU1 genotype B that are the same as those in CoV HKU1 genotype A but different from those in CoV HKU1 genotype C are highlighted in gray. The putative template switching region is underlined and bold. The first (TCA) and last (CAG) codons of nsp7 are also underlined. The arrows denote positions with nucleotide polymorphism (at position 11414, N16, N17, N20, N21, and N22 [genotype C] were T instead of C; at 11422, N16 and N20 [genotype C] were C instead of T; at 11449, N25 [genotype B] was C instead of T; at 11528, N16, N17, N20, and N22 [genotype C] were C instead of T; at 11740, N6, N7, N9, N10, N11, N13, N14, N18, N23, and N24 [genotype A] were T instead of C; at 12095, N6 and N7 [genotype A] were T instead of C; at 12140, N23 and N24 [genotype A] were C instead of T; at 12367, N15 and N25 [genotype C] were A instead of G; and at 12400, N6, N7, N9, N10, N11, N13, N14, N18, N23, and N24 [genotype A] were C instead of T).

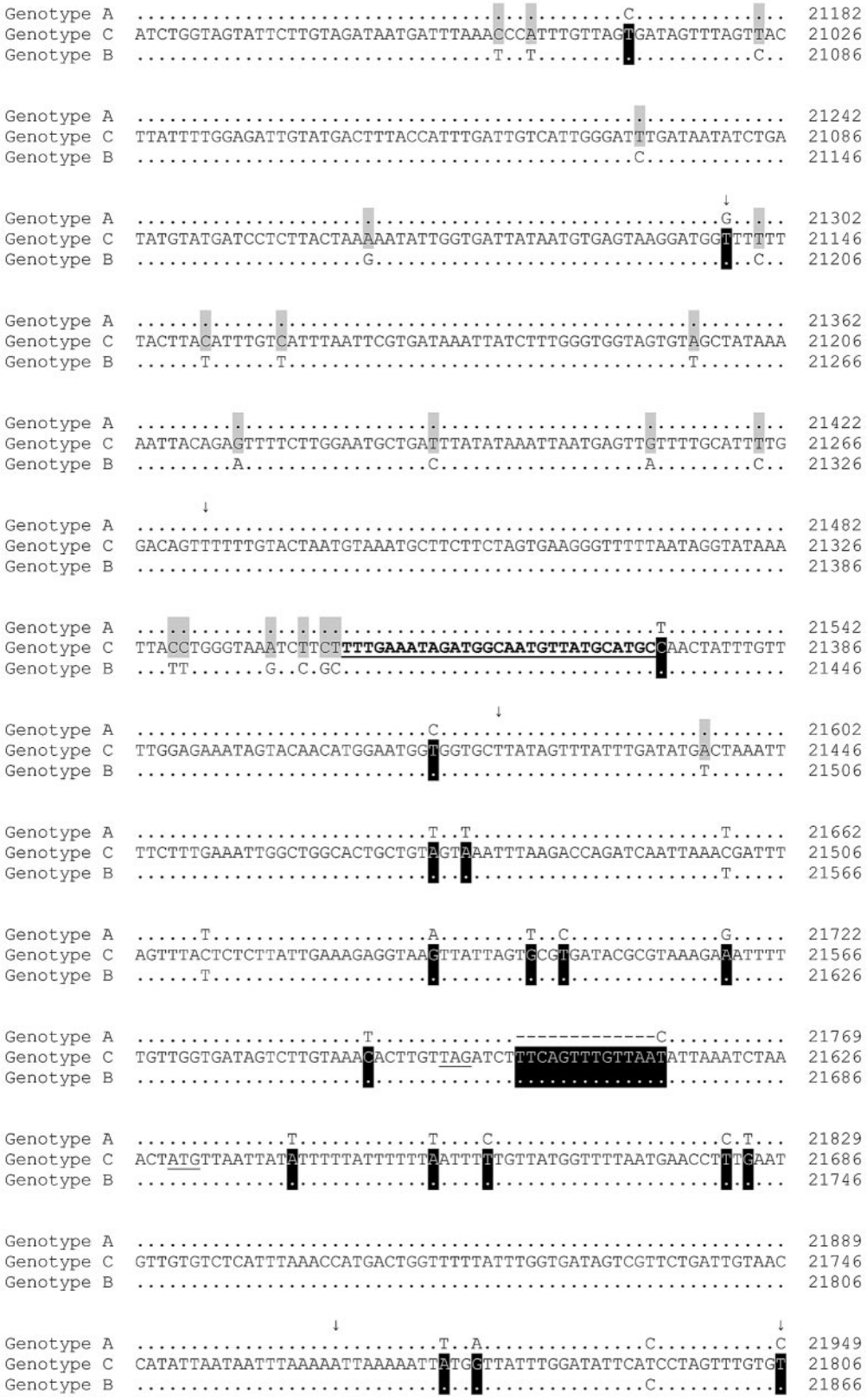

FIG. 4. Comparative sequence analysis of the nsp16-HE gene junction. Multiple alignments of the nucleotide sequences of CoV HKU1 genotypes A, B, and C. In CoV HKU1 genotype A and CoV HKU1 genotype B, only the nucleotides differing from those in CoV HKU1 genotype C are depicted. The nucleotides in CoV HKU1 genotype C that are the same as those in CoV HKU1 genotype A but different from those in CoV HKU1 genotype B are highlighted in gray, and those in CoV HKU1 genotype C that are the same as those in CoV HKU1 genotype B but different from those in CoV HKU1 genotype A are highlighted in black. The putative template switching region is underlined and bold. The stop codon of ORF1ab (TAG) and the start codon of the HE gene (ATG) are also underlined. The arrows denote positions with nucleotide polymorphisms (at 21297, N6, N7, N9, N10, N11, N13, N14, N23, and N24 [genotype A] were T instead of G; at 21429, N6 [genotype A] was C instead of T; at 21576, N15 [genotype B] was C instead of T; at 21908, N15 and N25 [genotype B] were G instead of A; and at 21949, N14 was T instead of C).

the strains. We speculate that this "independent evolution" of the number of repeats was due to the random expansion or deletion of part of the repeat region during the process of viral replication as a result of inaccurate replication by the viral polymerase or recombination between the repeat regions of different CoV HKU1 strains, a phenomenon widely observed in tandem repeats of genomes in all domains of life (3, 24). On the other hand, the sequence of the imperfect repeats seemed to coevolve with the rest of the genomes, most notably that all 10 CoV HKU1 strains with incomplete imperfect repeats (NDDD) were of genotype A (Tables 1 and 2). This could be due to the deletion of part of a repeat in one genotype A strain and subsequent expansion or deletion of whole repeats in its descendants. Further studies will delineate whether this ATR is useful for the molecular typing of CoV HKU1 strains.

This high frequency of recombination has resulted in the generation of a high diversity of coronaviruses in different animals. Before the SARS epidemic in 2003, a total of 19 coronaviruses were known, including 2 human, 13 mammalian, and 4 avian coronaviruses. After the SARS epidemic, within a short period of 3 years, 20 additional novel coronaviruses were described (5, 6, 11, 16, 19, 25, 26, 34, 36, 38, 43). These include 3 human coronaviruses, 11 mammalian coronaviruses, and 6 avian coronaviruses. Notably, there was a recent discovery of at least eight different species of coronaviruses in bats in Hong Kong, including SARS CoV-like viruses and a probable novel subgroup, group 2c, of coronavirus (16, 43). The high frequency of recombination in such a high diversity of coronaviruses may easily result in the generation of novel coronavirus species or genotypes that can cross host species barriers, leading to major zoonotic outbreaks with disastrous consequences. The potential of generation of novel species leading to zoonotic outbreaks and major consequences is analogous to the situation of avian and human influenza epidemiology, although the mechanism of generation of novel types and variants is by reassortment, which is different from recombination in coronaviruses (18, 44). Amplification of conserved regions in coronaviruses using RNA extracted from various animal specimens will lead to the discovery of more coronaviruses and subsequent complete genome sequencing, and comparative genome analysis will reveal the intricate relationships among the various coronaviruses.

ACKNOWLEDGMENTS

In memory of the late Ted Sun and with gratitude to his family for their generous support for research on emerging infections. This work is partly supported by a Research Grant Council grant (7616/05 M); the Research Fund for the Control of Infectious Diseases of the Health, Welfare, and Food Bureau of the Hong Kong Special Administrative Region Government; Commercial Radio's Fund; the Suen Chi Sun Charitable Foundation; and the William Benter Infectious Disease Fund.

REFERENCES

1. **Allander, T., M. T. Tammi, M. Eriksson, A. Bjerkner, A. Tiveljung-Lindell, and B. Andersson.** 2005. Cloning of a human parvovirus by molecular screening of respiratory tract samples. Proc. Natl. Acad. Sci. USA **102:**12891–12896.
2. **Banner, L. R., and M. M. Lai.** 1991. Random nature of coronavirus RNA recombination in the absence of selection pressure. Virology **185:**441–445.
3. **Bierne, H., M. Seigneur, S. D. Ehrlich, and B. Michel.** 1997. uvrD mutations enhance tandem repeat deletion in the *Escherichia coli* chromosome via SOS induction of the RecF recombination pathway. Mol. Microbiol. **26:**557–567.
4. **Copper, P. D., A. Steiner-Pryor, P. D. Scotti, and D. Delong.** 1974. On the nature of poliovirus genetic recombinants. J. Gen. Virol. **23:**41–49.
5. **East, M. L., K. Moestl, V. Benetka, C. Pitra, O. P. Honer, B. Wachter, and H. Hofer.** 2004. Coronavirus infection of spotted hyenas in the Serengeti ecosystem. Vet. Microbiol. **102:**1–9.
6. **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.
7. **Fu, K., and R. S. Baric.** 1992. Evidence for variable rates of recombination in the MHV genome. Virology **189:**88–102.
8. **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.
9. **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.
10. **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.
11. **Jonassen, C. M., T. Kofstad, I. L. Larsen, A. Lovland, K. Handeland, A. Follestad, and A. Lillehaug.** 2005. Molecular identification and characterization of novel coronaviruses infecting graylag geese (*Anser anser*), feral pigeons (*Columbia livia*) and mallards (*Anas platyrhynchos*). J. Gen. Virol. **86:**1597–1607.
12. **Kalinina, O., H. Norder, S. Mukomolov, and L. O. Magnius.** 2002. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. J. Virol. **76:**4034–4043.
13. **Kirkegaard, K., and D. Baltimore.** 1986. The mechanism of RNA recombination in poliovirus. Cell **47:**433–443.
14. **Lai, M. M., R. S. Baric, S. Makino, J. G. Keck, J. Egbert, J. L. Leibowitz, and S. A. Stohlman.** 1985. Recombination between nonsegmented RNA genomes of murine coronaviruses. J. Virol. **56:**449–456.
15. **Lai, M. M., and D. Cavanagh.** 1997. The molecular biology of coronaviruses. Adv. Virus Res. **48:**1–100.
16. **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.
17. **Lau, S. K. P., P. C. Y. Woo, C. C. Y. Yip, H. Tse, H.-W. Tsoi, V. C. C. Cheng, P. Lee, B. S. F. Tang, C. H. Y. 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, China. J. Clin. Microbiol. **44:**2063–2071.
18. **Li, K. S., Y. Guan, J. Wang, G. J. Smith, K. M. Xu, L. Duan, A. P. Rahardjo, P. Puthavathana, C. Buranathai, T. D. Nguyen, A. T. Estoepangestie, A. Chaisingh, P. Auewarakul, H. T. Long, N. T. Hanh, R. J. Webby, L. L. Poon, H. Chen, K. F. Shortridge, K. Y. Yuen, R. G. Webster, and J. S. Peiris.** 2004. Genesis of a highly pathogenic and potentially pandemic H5N1 influenza virus in eastern Asia. Nature **430:**209–213.
19. **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.
20. **Lole, K. S., R. C. Bollinger, R. S. Paranjape, D. Gadkari, S. S. Kulkarni, N. G. Novak, R. Ingersoll, H. W. Sheppard, and S. C. Ray.** 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. **73:**152–160.
21. **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.
22. **Makino, S., S. A. Stohlman, and M. M. Lai.** 1986. Leader sequences of murine coronavirus mRNAs can be freely reassorted: evidence for the role of free leader RNA in transcription. Proc. Natl. Acad. Sci. USA **83:**4204–4208.
23. **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.
24. **Nakamura. Y., M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujimoto, M. Hoff, and E. Kumlin.** 1987. Variable number of

tandem repeat (VNTR) markers for human gene mapping. Science **235:** 1616–1622.
25. **Pearks Wilkerson, A. J., E. C. Teeling, J. L. Troyer, G. K. Bar-Gal, M. Roelke, L. Marker, J. Pecon-Slattery, and S. J. O'Brien.** 2004. Coronavirus outbreak in cheetahs: lessons for SARS. Curr. Biol. **14:**R227–R228.
26. **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.
27. **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.
28. **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.
29. **Simmonds, P., and S. Midgley.** 2005. Recombination in the genesis and evolution of hepatitis B virus genotypes. J. Virol. **79:**15467–15476.
30. **Sloots, T. P., P. McErlean, D. J. Speicher, K. E. Arden, M. D. Nissen, and I. M. Mackay.** 2006. Evidence of human coronavirus HKU1 and human bocavirus in Australian children. J. Clin. Virol. **35:**99–102.
31. **Steininger, C., B. Schmied, M. Sarcletti, M. Geit, and E. Puchhammer-Stockl.** 2005. Cytomegalovirus genotypes present in cerebrospinal fluid of HIV-infected patients. AIDS **19:**273–278.
32. **Thiel, V., and S. G. Siddell.** 1994. Internal ribosome entry in the coding region of murine hepatitis virus mRNA 5. J. Gen. Virol. **75:**3041–3046.
33. **Vabret, A., J. Dina, S. Gouarin, J. Petitjean, S. Corbet, and F. Freymuth.** 2006. Detection of the new human coronavirus HKU1: a report of 6 cases. Clin. Infect. Dis. **42:**634–639.
34. **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. Spaargaren, and B. Berkhout.** 2004. Identification of a new human coronavirus. Nat. Med. **10:**368–373.
35. **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.
36. **Wise, A. G., M. Kiupel, and R. K. Maes.** 2006. Molecular characterization of a novel coronavirus associated with epizootic catarrhal enteritis (ECE) in ferrets. Virology **349:**164–174.
37. **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.
38. **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.
39. **Woo, P. C., S. K. Lau, Y. Huang, H. W. Tsoi, K. H. Chan, and K. Y. Yuen.** 2005. Phylogenetic and recombination analysis of coronavirus HKU1, a novel coronavirus from patients with pneumonia. Arch. Virol. **150:**2299–2311.
40. **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.
41. **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.
42. **Woo, P. C., S. K. Lau, B. H. Wong, H. W. Tsoi, A. M. Fung, K. H. Chan, V. K. Tam, J. S. Peiris, and K. Y. Yuen.** 2004. Detection of specific antibodies to severe acute respiratory syndrome (SARS) coronavirus nucleocapsid protein for serodiagnosis of SARS coronavirus pneumonia. J. Clin. Microbiol. **42:** 2306–2309.
43. **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.** 26 April 2006. Molecular diversity of coronaviruses in bats. Virology [Epub ahead of print].
44. **Yuen, K. Y., P. K. Chan, M. Peiris, D. N. Tsang, T. L. Que, K. F. Shortridge, P. T. Cheung, W. K. To, E. T. Ho, R. Sung, and A. F. Cheng.** 1998. Clinical features and rapid viral diagnosis of human disease associated with avian influenza A H5N1 virus. Lancet **351:**467–471.