The Complete Genome Sequence of Severe Acute Respiratory Syndrome Coronavirus Strain HKU-39849 (HK-39)

F.Y. ZENG,* C.W.M. CHAN,* M.N. CHAN,* J.D. CHEN,* K.Y.C. CHOW,* C.C. HON,* K.H. HUI,* J. LI,* V.Y.Y. LI,* C.Y. WANG,* P.Y. WANG,* Y. GUAN,† B. ZHENG,† L.L.M. POON,† K.H. CHAN,† K.Y. YUEN,† J.S.M. PEIRIS,† AND F.C. LEUNG*,[1]

*Departments of *Zoology and †Microbiology, University of Hong Kong, Hong Kong, SAR China*

The complete genomic nucleotide sequence (29.7kb) of a Hong Kong severe acute respiratory syndrome (SARS) coronavirus (SARS-CoV) strain HK-39 is determined. Phylogenetic analysis of the genomic sequence reveals it to be a distinct member of the *Coronaviridae* family. 5′ RACE assay confirms the presence of at least six subgenomic transcripts all containing the predicted intergenic sequences. Five open reading frames (ORFs), namely ORF1a, 1b, S, M, and N, are found to be homologues to other CoV members, and three more unknown ORFs (X1, X2, and X3) are unparalleled in all other known CoV species. Optimal alignment and computer analysis of the homologous ORFs has predicted the characteristic structural and functional domains on the putative genes. The overall nucleotides conservation of the homologous ORFs is low (<5%) compared with other known CoVs, implying that HK-39 is a newly emergent SARS-CoV phylogenetically distant from other known members. SimPlot analysis supports this finding, and also suggests that this novel virus is not a product of a recent recombinant from any of the known characterized CoVs. Together, these results confirm that HK-39 is a novel and distinct member of the *Coronaviridae* family, with unknown origin. The completion of the genomic sequence of the virus will assist in tracing its origin. Exp Biol Med 228:866–873, 2003

Key words: SARS; coronavirus; genomic sequence; Subgenomic transcripts; 5′-RACE assay

The World Health Organization has been tracking and reporting the cumulative number of reported probable cases of severe acute respiratory syndrome (SARS) since November 2002. As of May 1, 2003, 5,865 cases of SARS have been reported from a total of 27 countries worldwide. The number of fatalities has reached 391, suggesting that this virus is highly virulent. A novel coronavirus (CoV) from respiratory specimens has been isolated from two SARS patients in Hong Kong (1). Its novelty has been confirmed by investigators from the Centers for Disease Control and Prevention (CDC), who have also isolated the CoV from patient samples (2). This CoV has not been previously identified either in humans or animals. The successful isolation of the novel CoV has not only permitted us to make a definitive diagnosis, but has also enabled us to complete the genomic sequence of the virus for further characterization.

Coronaviridae is a viral family that infects birds and mammals and causes a variety of diseases (3). Six species of CoV genomes have been completely sequenced, namely murine hepatitis virus (MHV) (4); avian infectious bronchitis virus (IBV) (5); human CoV 229E (HCV 229E) (6); bovine CoV (BCV) (7); transmissible gastroenteritis virus (TGEV) (8); and porcine epidemic diarrhea virus (PEDV) (9). The size of the genome is about 30 kb, of which more than two-thirds is occupied by open reading frames (ORF) 1a, b. This gene contains two large ORFs, ORF 1a and ORF 1b that are cotranslated with -1 ribosomal frameshifting mechanism (10–13). The gene products of both ORFs are believed to be processed into a number of functional subunits (14, 15). The putative functional domains in ORF1 include two to three papain-like domains, one 3C-like protease domain, one growth factor/receptor-like domain, one polymerase domain, one metal ion-binding domain, and a helicase domain. The remaining one-third of the genome consists mainly of three structural proteins: a surface-spike glycoprotein (S), a transmembrane protein (M), and a nucleocapsid protein (N). Some CoV genomes (Group II) also contain the hemagglutinin-esterase (HE) gene. The spike glycoprotein has major functions in virus-host cell membrane fusion and interaction with host cell surface receptors. Membrane protein is responsible for virus budding, but other viral proteins may also be involved in organizing the

This work was supported by the University of Hong Kong.

[1] To whom requests for reprints should be addressed at Department of Zoology, Kadoorie Biological Sciences Building, University of Hong Kong, Hong Kong, SAR China. E-mail: fcleung@hkucc.hku.hk

Received May 2, 2003.
Accepted May 15, 2003.

1535-3702/03/2287-0866$15.00

virus budding pre-Golgi membranes (16, 17). The nucleocapsid protein has been found to be the central hydrophilic basic domain involved in RNA binding (18). Also, it has recently been suggested that CoVs contain an internal core of helical nucleocapsid, which is composed of both M and N proteins, in which the M protein was supposed to be found only in the envelope protein of the virus (19). Small ORFs are also usually found between these structural genes, depending on the virus species (20–22).

We have extracted the genomic RNA from a tissue culture sample of the SARS CoV strain isolated from one of the earliest Hong Kong patients with SARS. This strain has been given the name HK-39. By using degenerated and specific primer PCR amplification, coupled with cDNA library screening, we were able to obtain the 29-kb complete genomic sequence. In this paper, we report the complete nucleotide sequence of a Hong Kong SARS CoV and compare and analyze its genomic organization and individual genes with those of other known SARS CoV species.

Materials and Methods

Source of Materials. The initial starting material for this study was RNA isolated from fetal rhesus monkey kidney (FRhK-4) cells infected with a CoV HK-39 isolate from one of the earliest patients who died from SARS in Hong Kong (1). Total RNA was extracted using SV total RNA isolation system (Promega, Madison, WI) according to the manufacturer's instructions. The first cDNA strand was reverse-transcribed using Superscript II (Invitrogen, Carlsbad, CA) and random primers. The cDNA was then used for cDNA library construction and specific amplification of viral genomic sequences.

Construction of cDNA Library. Double-stranded DNA linkers were added to the 3′ end of the first-strand cDNA for second strand cDNA synthesis (23). The end of the double-stranded cDNA was modified by T4 polymerase. The processed DNA was than subsequently cloned into pCR2.1 vector (Invitrogen). The clones with the cDNA insert were screened by PCR and were subjected to direct sequencing analysis.

Amplification of Viral Genomic Sequences. Degenerated primers covering the whole genome of the CoV were designed based on the genomic sequences of other CoVs. With assistance from the results obtained from library screening, degenerated primers amplification, primer walking, and other public sources (e.g., SARS CoV strain Tor2, GenBank accession number NC_004718), the gaps in the genome were finally closed by specific primer PCR and sequencing.

Sequencing. DNA fragments resulting from PCRs from viral genome and library screening were purified and directly sequenced by BigDye Terminator Cycle Sequencing in an ABI Prism 3100 Genetic Analyzer (Perkin Elmer, Norwalk, CT). CoV sequences were confirmed by searching in NCBI BLAST-X.

5′-RACE and 3′-RACE. The 5′ end of RNA genome and RNA transcripts were identified by using two different 5′-RACE commercial kits: RNA ligase-mediated 5′-RACE (GeneRacer kit; Invitrogen), and end-switching 5′-RACE (SMART RACE cDNA Amplification kit; Clontech, Palo Alto, CA). One microgram of RNA extracted from virus infected cells was used for each 5′-RACE reaction in accordance with the manufacturer's instructions. Specific primers located near the 5′ end of each possible gene were used for PCR, and nested PCR was carried out if needed. For the 3′-RACE, the first-strand cDNA was reverse-transcribed with a 3′-RACE oligo, which had two primers (C1 and C2) annealed sites ahead of oligo dT. A set of specific primers at different sites of the genome were combined with 3′-RACE-anchored primers C1 or C2, respectively, and were used to amplify the possible 3′ end of poly(A)$^+$ mRNA transcripts or the RNA genome. All the PCR fragments of 5′- and 3′-RACE were subjected to sequencing directly.

Data Analysis. The 29.7-kb complete genome sequence was assembled from the sequence contigs using SeqMan of the Lasergene Package (DNASTAR, Madison, WI). The putative ORFs were predicted by EditSeq of the Lasergene Package (DNASTAR). The SARS CoV complete genome sequence was compared with those of other known CoV species. Multiple sequence analysis and optimal alignments were conducted on MegAlign of the Lasergene Package (DNASTAR). Phylogenetic tree construction and bootstrap tests were performed using MEGA 2.1 (Arizona State University, Tempe, AZ). The similarity plots of multiple sequence alignments were performed by SimPlot (Johns Hopkins University School of Medicine, Baltimore, MD). The coiled-coil motif prediction of the S protein was performed by COILS (24). Transmembrane domain topology predictions of the proteins were performed using TMHMM (CBS, Technical University of Denmark, Copenhagen, Lyngby, Denmark).

Results and Discussion

Genome Sequence of HK-39. The first genomic sequence of HK-39 obtained in this study was a 240-bp ORF 1b fragment amplified by porcine reproductive and respiratory syndrome virus (PRRSV) specific primers (Fig. 1A). PRRSV is a positive-stranded RNA virus and belongs to the order *Nidovirales*, the same order as the CoV. Six other fragments totaling about 3 kb covering different regions of the genome were successfully amplified by degenerated primers (Fig. 1A). Specific primers were designed for walking to fill up the gaps. These methods, coupled with cDNA library screening, enabled more than 30% of the genome sequence to be obtained. Work carried out in the BCCA Genome Sciences Center (British Columbia Center for Disease in Canada) enabled us to uncover the complete viral genome in 2 days. When the genome sequence of HK-39 was compared with the sequence of Tor2, a high homology was found. By using the specific primer

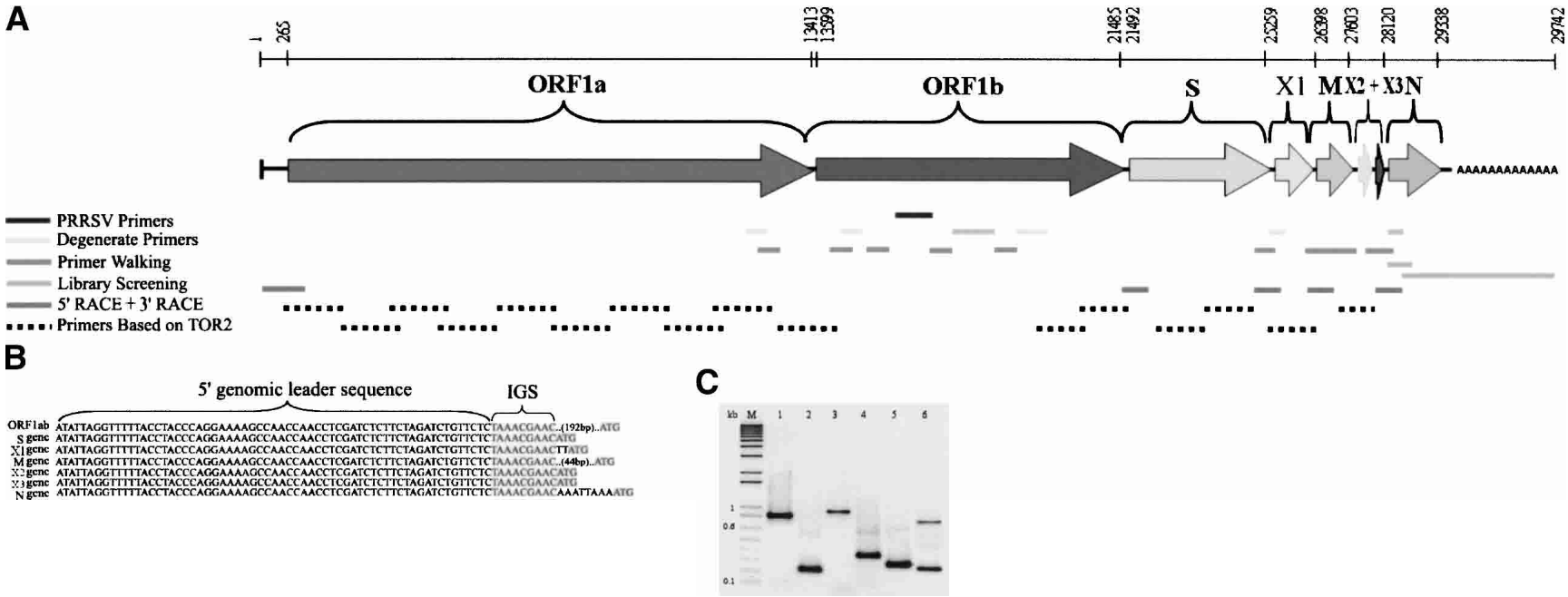

Figure 1. Genomic organization of SARS coronavirus HK-39. (A) Graphical presentation of different kinds of sequence contigs acquired by PCR using different combinations of primers, library screening, and 5′-RACE, with the confirmed ORFs. (B) Alignment 5′-RACE results of identified genes. The grey solid box highlights the nine conserved nucleotides of IGSs. (C) Agarose gel document showing the size of different 5′-RACE products using specific primer for: 1kb (Lane 1), S Gene (Lane 2), X1 (Lane 3), M (Lane 4), X2 (Lane 5), and N + X3 Gene (Lane 6).

walking (Fig. 1A), the final sequence of HK-39 was completed.

Assembly and Analysis of the Genome. In total, 125 sequence contigs were used in the complete genome assembly. The full genome length of SARS CoV HK-39 was found to be 29,749 bp, including a 264-bp 5′-untranslated region (UTR) and a 339-bp 3′-UTR with a poly $(A)_{15}$ tail. The full sequence was submitted to NCBI, and accession number AY278491 was assigned. The genomic organization of the virus was found to be similar to other virus members in the family *Coronaviridae*. SARS CoV HK-39 contains a 21-kb RNA-dependent RNA polymerase gene with two subunits (ORF1a and ORF1b) and three structural proteins, namely a surface-spike glycoprotein, a membrane protein, and a nucleocapsid protein. Three unknown ORFs (X1, X2, and X3) were also identified and confirmed by 5′-RACE (Fig.1B and C). The length and the position of the confirmed ORFs are shown in Figure 1A.

Sequence Alignment and Phylogenetic Analysis. The *Coronaviridae* family is classified into three groups according to the structural proteins that affect their antigenicity (25). Phylogenetic analysis of the whole genome and individual ORFs of HK-39 with other known CoV species showed that SARS-CoV only shares a very low level of homology with the other members of the CoV family at the level of nucleotide sequence and forms a separate group (Fig. 2A). The topology of the phylogenetic tree was similar in all analyses based on amino acid differences (Fig. 2, B–E). Although SARS-CoV showed a higher degree of amino acid sequence homology to Group 2 species, its genomic organization is more closely resembles that of Group 1 species (data not shown). Nevertheless, this controversy exists, and it is clear that SARS-CoV, including the Tor2 strain (isolated from SARS patient number 2 in Toronto [1]), Urbani strain (GenBank accession number AY278741), and CUHK-W1 strain (isolated from a patient suffering from SARS in Hong Kong, GenBank accession number AY278554 [unpublished data]), does not fall into any of the existing phylogenetic groups, and is distant from all known human CoVs such as the human 229E and human CoV OC43, although they share a common host. SARS CoV is a new member of the CoV family, and is very distinct from all CoVs characterized hitherto.

5′- and 3′-UTR of the Genome. The 5′-UTR of the genome was characterized by 5′-RACE assay (of ORF 1). We obtained the 264 bp upstream to the predicted AUG initiation codon of ORF 1, which is identical to the most updated version of the Urbani strain. Alignment of the seven 5′-RACE sequences showed a consensus of 72 bp, which is composed of a leader sequence of 61 bp and an intergenic sequence at the last 9 bp. The intergenic sequence (IGS) 5′-UAAACGAAC-3′ was identical for all of its ORFs, except X2 (Fig. 1B). Initiation codons were usually found immediately after or a few bases away from the IGS, except for ORF1 and M. An 11-codon "mini-ORF" was predicted 31 bp downstream of the IGS and 128 bp upstream of the initiation codon of AUG at the 5′-UTR, which is similar to that of IBV (26). Eighteen specific primers located at different sites of the genome were used to amplify the 3′ ends of the transcripts. Sequencing of the 3′-RACE products showed that only the regions at the 3′ end of the genome were amplified. The above results support the unique discontinuous transcription system in CoVs, which generate a nested set of transcripts with common 3′ ends and a common leader sequence on the 5′ ends. 3′-UTR (sequence downstream the N protein sequence) has been shown to be crucial in the regulation of transcription in a CoV. SimPlot analysis of 3′-UTR showed a remarkably high degree of similarity with IBV, which contradicts that of the other regions of the genome (Fig. 2F). A 32-bp conserved motif (nucleotides 29590-29621) was found in the 3′-UTR of SARS CoV HK-39. Such a motif shares a very high homology with the stem-loop II-like motif (s2m) found in IBV (27). Jonassen *et al.* (27) pointed out that such a motif was also found in some viruses that are distinct from IBV, like some animal astroviruses and picornavirus, and that it may be due to the consequence of the RNA transferring event that occurred between these viruses. Their findings in IBV, together with the identified putative s2m motif in HK-39, imply that these two CoVs are evolutionarily related.

Putative Functional Domains of ORF 1. Sequence alignment on the predicted amino acid sequence of ORF1 revealed the uniqueness of this novel SARS CoV. In general, it shows an average of less than 50% similarity with any other groups of CoVs. In ORF1a, one papain-like (PL) domain, a 3C-like (3CL) protease, and a growth factor/receptor-like (GFL) domain were predicted with the reference to that of TGEV (Fig. 3A) (8). The organization is similar to other members of the virus family (12, 13, 28). Two characteristic and remarkably hydrophobic regions located at both sides of the 3CL domain were identified by computer predictions. In total, 10 putative 3CL cleavage sites were predicted in ORF1ab, and their locations are shown in Figure 3A. One of the necessary elements for the ribosomal frame-shifting mechanism (10, 13), a ribosomal slippage site UUAAAC, was also identified at 13392 bp, 15 bp upstream of the stop codon of ORF1a. Alignment of the predicted amino acid sequence of ORF1b to known strains revealed the presence of conserved putative domains, including RNA polymerase domain (POL), metal ion-binding (MIB), and helicase (Hel) domain (Fig. 3A). These findings support the conclusion that the SARS CoV is a typical member of the CoV family.

Topologies of Structural Protein. The putative 5′ region of the spike protein of SARS CoV shares good homology with the bovine S1 region and contains a possible polybasic cleavage site of SLLR at amino acid 667. Two coiled-coil structures were predicted in the putative S2 region (Fig. 3, B2). Normally, two to three clusters of heptad repeats are found in the S2 region in other CoVs. The heptad

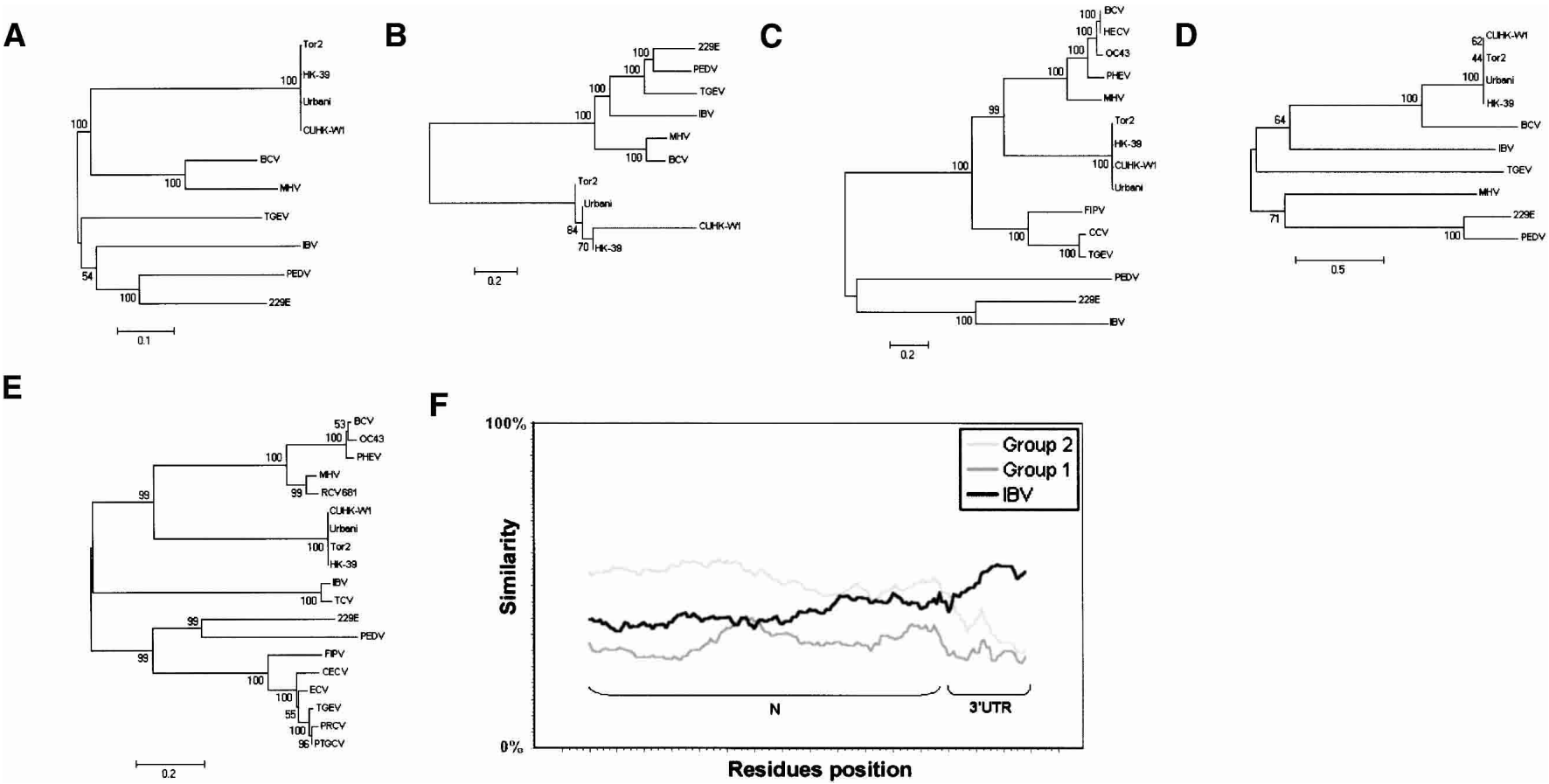

Figure 2. The SARS coronavirus is a new coronavirus member. Phylograms showing the relationship of the four SARS coronaviruses based on full genome nucleotide differences (A) or amino acid sequence differences (B,C,D,E), using neighbor-joining with Kimura 2-parameter distance (scale bar). Numbers at the node are bootstra *P* values (500 replicates) conducted by MEGA version 2.1 (31). GenBank accession number of full genomic nucleotide sequence (A) BCV (Bovine coronavirus): NC_003045; MHV (Murine hepatitis virus): NC_001846; TGEV (Transmissible gastroenteritis virus): NC_002306; IBV (Avian infectious bronchitis virus): NC_001451; PEDV (Porcine epidemic diarrhea virus): NC_003436; 229E (Human coronavirus 229E): NC_002645; and of protein sequence (B) 1ab gene - 229E: NP_073549; PEDV: NP_598309; TGEV: NP_058422; IBV: NP_066134; MHV: NP_068668; BCV: NP_150073; (C) S gene – BCV: AAL40400; HEC (Human enteric coronavirus 4408): L07748; OC43 (Human coronavirus OC43): L14643; PHEV (Porcine hemagglutinating encephalomyelitis virus): AAL80031; MHV: P11225; FIPV (Feline infectious peritonitis virus): AB088223; CCV (Canine coronavirus): AB017789; TGEV: NP_058424; PEDV: CAA80971; 229E: AAK32191; IBV: AAO34396; (D) M gene – BCV: NP_150082; IBV: NP_040835; TGEV: NP_058427; MHV: AAF36439; 229E: NP_073555; PEDV: NP_598313; (E) N gene – BCV: NP_150083; OC43: P33469; PHEV: AAL80036; MHV: P18446; RCV681 (Rat coronavirus strain 681): Q02915; IBV: NP_040838; TCV (Turkey coronavirus): AF111997.1; 229E: NP_073556; PEDV: NP_598314; FIPV: P25909; CECV (Canine enteric coronavirus strain INSAVC-1): P36298; ECV (Enteric coronavirus): CAA47246; TGEV: NP_058428; PRCV (Porcine respiratory coronavirus strain 86/137004 / British isolate): P33463; and PTGCV (Porcine transmissible gastroenteritis coronavirus strain FS772/70): P05991. (F) Simplot analysis based on nucleotide residues homology of SARS CoV HK-39 with other coronaviruses on N Gene and 3′-untranslated region. *x* axis is the nucleotide location from the start codon of N gene of SARS CoV HK-39 to the end of the genome. *Y* axis is the percentage of homology. Grouping of the coronaviruses: Group 1, CCV, FIPV, 229E, PEDV, PTGCV; Group 2, BCV, OC43, MHV, PHEV, RCV681; Group 3, IBV, TCV.

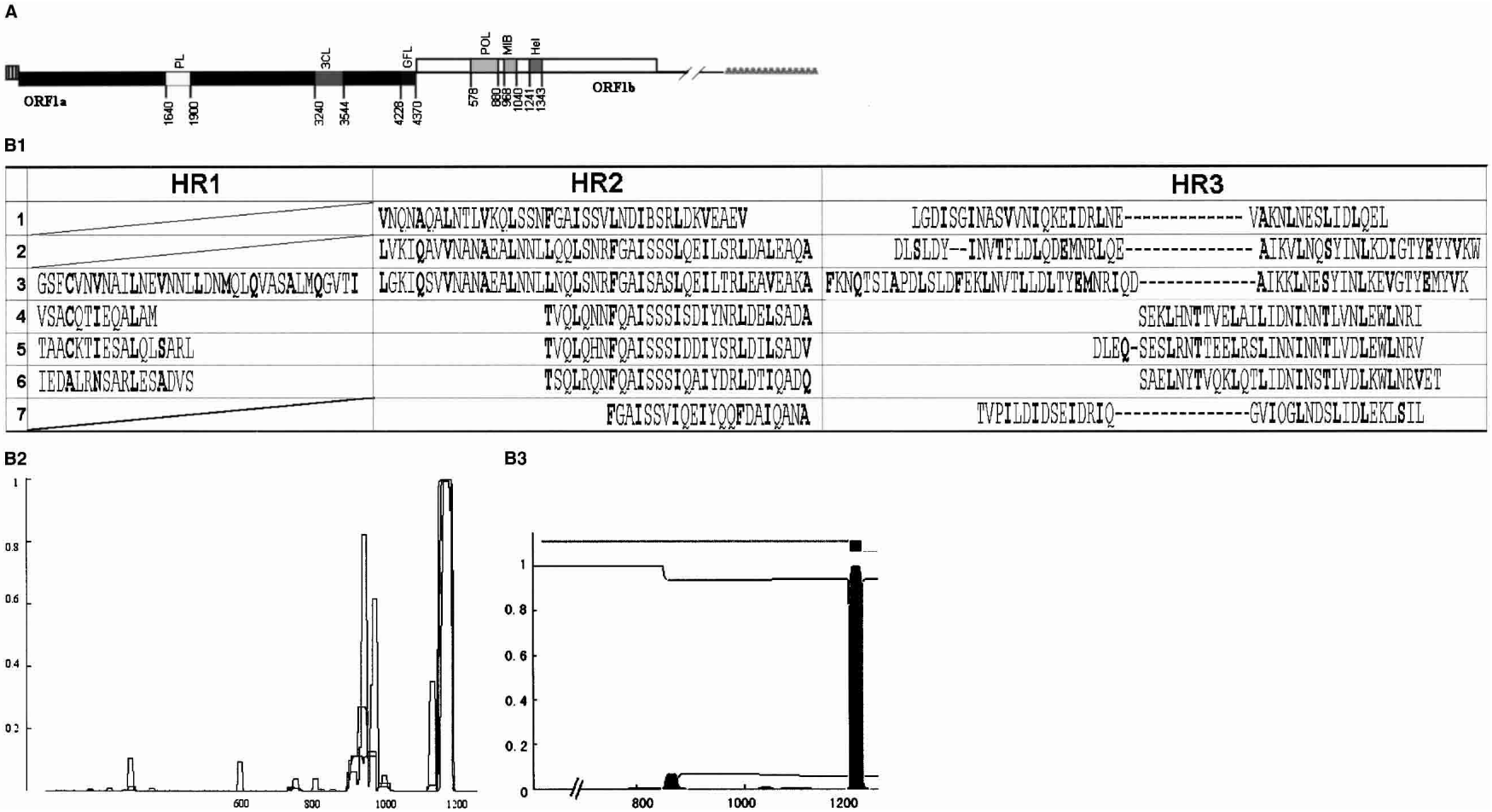

Figure 3. Molecular studies of the SARS HK-39 coronavirus. (A) Potential functional domains of ORF1ab. The first striped box on the left represents leader sequences, arrows indicate putative 3C-like cleavage sites, and other boxes in each ORF are the putative functional domains (PL, Papain-like; 3CL, 3C-like protease domain; GFL, growth factor/receptor-like domain; POL, RNA-dependent RNA polymerase; MIB, metal ion binding domain; Hel, Helicase). (B) 1. Comparison of heptad repeats in different coronavirus with SARS coronavirus HKU39; hydrophobic residues are shown in bold font (1, SARS CoV HK-39; 2, BCV; 3, MHV; 4, TGEV; 5, PEDV; 6, 229E; 7, IBV). 2. Coiled coil structure prediction of S gene based on the appearance of heptad repeats. Peaks indicate the position of the coiled coil and its width corresponds to the length of the coil. *y* axis is the possibility value and *x* axis is the amino acid location according to S gene only. 3. Transmembrane domain prediction of S and (C) M gene based on the hydrophobicity of amino acid residues of the gene. *y* axis is the possibility value and *x* axis is the amino acid location. Dark solid bar and peak indicates potential transmembrane domain. (D) Potential RNA interaction site of HK-39 as predicted from the N gene amino acid alignment, where highly conserved residues (highlighted) of the 16 coronaviruses in the middle region of the gene fall into the region believed to be involved in 5′ sequence interaction, as in the case of IBV (32). Amino acid position is located on the top and number is according to the consensus sequences shown above position bar.

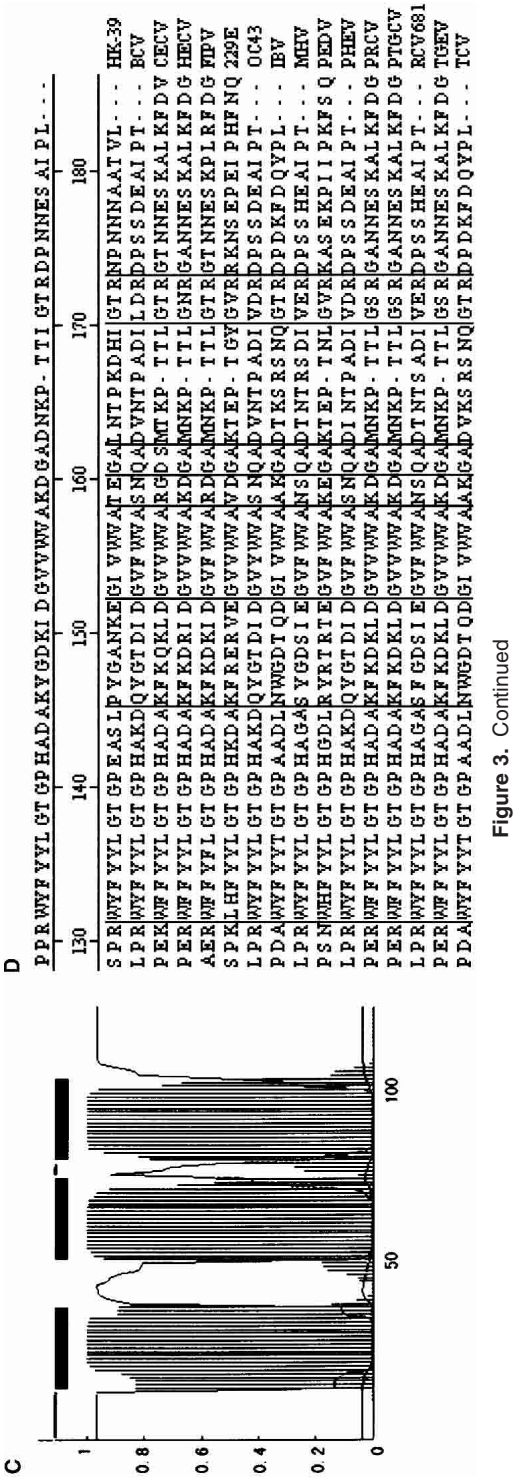

Figure 3. Continued

repeats of these coiled-coil regions were identified in other CoVs (Fig. 3, B1). A conserved transmembrane domain was predicted in the C-terminal of the S2 region (Fig. 3, B3). Amino acid sequence alignment of the transmembrane domain has shown that this region is highly conserved. All coronavirus M proteins show to posses a triple membrane-spanning protein with a Nexo-Cendo configuration (17, 29, 30). TMHMM predicts three transmembrane domains on the N-terminal of the M gene of SARS CoV (Fig. 3C). It is observed in the alignment of N protein sequence that there are three stretches of amino acid residues that are highly conserved among the 16 CoV species. Such residues are believed to be involved in its structural maintenance and direct interaction with RNA in the case of IBV (Fig. 3D) (25).

Conclusion. We have completed the sequencing of a novel CoV HK-39. The distinctive molecular genomic and phylogenetic characteristics of this novel virus seem to warrant its assignment to a new and distinct group IV of the *Coronaviridae* family.

1. Peiris J, Lai S, Poon L, Guan Y, Yam L, Lim W, Nicholls J, Yee W, Yan W, Cheung M, Cheng V, Chan K, Tsang D, Yung R, Ng T, Yuen K. Coronavirus as a possible cause of severe acute respiratory syndrome. Lancet **361:**1319–1325, 2003.
2. Ksiazek TG, Erdman D, Goldsmith CS, Zaki SR, Peret T, Emery S, Tong S, Urbani C, Comer JA, Lim W, Rollin PE, Dowell SF, Ling AE, Humphrey CD, Shieh WJ, Guarner J, Paddock CD, Rota P, Fields B, DeRisi J, Yang JY, Cox N, Hughes JM, LeDuc JW, Bellini WJ, Anderson LJ. Novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med **348:** 1953–1966, 2003.
3. Enjuanes L, Brian D, Cavanagh D, Holmes K, Lai MMC, Laude H, Masters P, Rottier P, Siddell SG, Spaan WJM, Taguchi F, Talbot P. *Coronaviridae*. In: van Regenmortel MHV, Fauquet CM, Bishop DHL, Carsten EB, Esters MK, Lemon SM, McGeoch DJ, Maniloff J, Mayo MA, Pringle CR, Wickner RB, Eds. Virus Taxonomy. Classification and Nomenclature of Viruses. New York: Academic Press, pp835–849, 2000.
4. Pachuk CJ, Bredenbeek PJ, Zoltick PW, Spaan WJ, Weiss SR. Molecular cloning of the gene encoding the putative polymerase of mouse hepatitis coronavirus, strain A59. Virology **171:**141–148, 1989.
5. Boursnell ME, Brown TD, Foulds IJ, Green PF, Tomley FM, Binns MM. Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. J Gen Virol **68:**57–77, 1987.
6. Herold J, Raabe T, Schelle-Prinz B, Siddell SG. Nucleotide sequence of the human coronavirus 229E RNA polymerase locus. Virology **195:**680–691, 1993.
7. Yoo D, Pei Y. Full-length genomic sequence of bovine coronavirus (31 kb). Completion of the open reading frame 1a/1b sequences. Adv Exp Med Biol **494:**73–76, 2001.
8. Eleouet JF, Rasschaert D, Lambert P, Levy L, Vende P, Laude H. Complete sequence (20 kilobases) of the polyprotein-encoding gene 1 of transmissible gastroenteritis virus. Virology **206:**817–822, 1995.
9. Kocherhans R, Bridgen A, Ackermann M, Tobler K. Completion of the porcine epidemic diarrhoea coronavirus (PEDV) genome sequence. Virus Genes **23:**137–144, 2001.
10. Brierley I, Digard P, Inglis SC. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell **19:**537–547, 1989.
11. Bredenbeek PJ, Pachuk CJ, Noten AF, Charite J, Luytjes W, Weiss SR, Spaan WJ. The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-

A59: a highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. Nucleic Acids Res **18:**1825–1832, 1990.
12. Lee HJ, Shieh CK, Gorbalenya AE, Koonin EV, La Monica N, Tuler J, Bagdzhadzhyan A, Lai MM. The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. Virology **180:**567–582, 1991.
13. Herold J, Siddell SG. An "elaborated" pseudoknot is required for high frequency frameshifting during translation of HCV 229E polymerase mRNA. Nucleic Acids Res **21:**5838–5842, 1993.
14. Denison MR, Zoltick PW, Leibowitz JL, Pachuk CJ, Weiss SR. Identification of polypeptides encoded in open reading frame 1b of the putative polymerase gene of the murine coronavirus mouse hepatitis virus A59. J Virol **85:**3072–3082, 1991.
15. Liu DX, Brierley I, Tibbles KW, Brown TD. A 100-kilodalton polypeptide encoded by open reading frame (ORF) 1b of the coronavirus infectious bronchitis virus is processed by ORF 1a products. J Virol **68:**5772–5780, 1994.
16. Opstelten DJ, de Groote P, Horzinek MC, Rottier PJ. Folding of the mouse hepatitis virus spike protein and its association with the membrane protein. Arch Virol Suppl **9:**319–328, 1994.
17. Rottier PJM. The coronavirus nucleocapsid protein. In: Siddell SG, Ed. The Coronaviridae. New York: Plenum Press, pp115–139, 1995.
18. Laude H, Masters PS. The coronavirus nucleocapsid protein. In: Siddell SG, Ed. The Coronaviridae. New York: Plenum Press, pp141–163, 1995.
19. Risco C, Anton IM, Enjuanes L, Carrascosa JL. The transmissible gasteroenteritis coronavirus contains a spherical core shell consisting of M and N proteins. J Virol **70:**4773–4777, 1996.
20. Jacobs L, van der Zeijst BAM, Horzinek MC. Characterization and translation of transmissible gastroenteritis virus mRNAs. J Virol **57:**1010–1015, 1986.
21. Rasschaert D, Gelfi J, Laude H. Enteric coronavirus TGEV: partial sequence of the genomic RNA, its organization and expression. Biochimie **69:**591–600, 1987.
22. Wesley RD, Cheung AK, Michael DD, Woods RD. Nucleotide sequence of coronavirus TGEV genomic RNA: evidence for three mRNA species between the peplomer and matrix protein genes. Virus Res **13:**87–100, 1989.
23. Shibata Y, Carninci P, Watahiki A, Shiraki T, Konno H, Muramatsu M, Hayashizaki Y. Cloning full-length, cap-trapper-selected cDNAs by using the single-strand linker ligation method. BioTechniques **30:**1250–1253, 2001.
24. Lupas A, Van Dyke M, Stock J. Predicting coiled coils from protein sequences. Science **252:**1162–1164, 1991.
25. Lee CW, Jackwood MW. Evidence of genetic diversity generated by recombination among avian coronavirus IBV. Arch Virol **145:**2135–2148, 2000.
26. Brown JDK, Boursnell MEG, Binns MM, Tomley FM. Cloning and sequencing of the 5′ terminal sequences from avian infectious bronchitis virus genomic RNA. J Gen Virol **67:**221–228, 1989.
27. Jonassen CM, Jonassen TO, Grinde B. A common RNA motif in the 3′ end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. J Gen Virol **79:**715–718, 1998.
28. Gorbalenya AE, Koonin EV, Donchenko AP, Blinov VM. Coronavirus genome: prediction of putative functional domains in the nonstructural polyprotein by comparative amino acid sequence analysis. Nucleic Acids Res **17:**4846–4861, 1989.
29. Den Boon JA, Snijder EJ, Chimside ED, De Vries AAF, Horzinek MC, Spaan WJM. Equine arteritis virus is not a togaviurs but belongs to the coronavirus-like superfamily. J Virol **65:**2910–2920, 1991.
30. Faaberg KS, Plagemann PGW. The envelope proteins of lactate dehydrogenase-elevating virus and their membrane topology. Virology **212:**512–525, 1995.
31. Zhow ML, Collisson EW. The amino and carboxyl domains of the infectious bronchitis virus nucleocapsid protein interact with 3′genome RNA. Virus Res **67:**31–39, 2000.
32. Kumar S, Tamura K, Jakobsen IB, Nei B. MEGA2: Molecular Evolutionary Genetics Analysis Software. Bioinformatics, 2001.