

# Characterization and Complete Genome Sequence of a Novel Coronavirus, Coronavirus HKU1, from Patients with Pneumonia

Patrick C. Y. Woo,<sup>1,2†</sup> Susanna K. P. Lau,<sup>1,2†</sup> Chung-ming Chu,<sup>3</sup> Kwok-hung Chan,<sup>1</sup>  
Hoi-wah Tsoi,<sup>1</sup> Yi Huang,<sup>1</sup> Beatrice H. L. Wong,<sup>1</sup> Rosana W. S. Poon,<sup>1</sup>  
James J. Cai,<sup>1</sup> Wei-kwang Luk,<sup>4</sup> Leo L. M. Poon,<sup>1,2</sup> Samson S. Y. Wong,<sup>1,2</sup>  
Yi Guan,<sup>1,2</sup> J. S. Malik Peiris,<sup>1,2</sup> and Kwok-yung Yuen<sup>1,2†\*</sup>

*Department of Microbiology<sup>1</sup> and State Key Laboratory of Emerging Infectious Diseases,<sup>2</sup> The University of Hong Kong, Division of Respiratory Medicine, Department of Medicine, United Christian Hospital,<sup>3</sup> and Department of Microbiology, Tseung Kwan O Hospital,<sup>4</sup> Hong Kong*

Received 29 July 2004/Accepted 3 September 2004

Despite extensive laboratory investigations in patients with respiratory tract infections, no microbiological cause can be identified in a significant proportion of patients. In the past 3 years, several novel respiratory viruses, including human metapneumovirus, severe acute respiratory syndrome (SARS) coronavirus (SARS-CoV), and human coronavirus NL63, were discovered. Here we report the discovery of another novel coronavirus, coronavirus HKU1 (CoV-HKU1), from a 71-year-old man with pneumonia who had just returned from Shenzhen, China. Quantitative reverse transcription-PCR showed that the amount of CoV-HKU1 RNA was  $8.5$  to  $9.6 \times 10^6$  copies per ml in his nasopharyngeal aspirates (NPAs) during the first week of the illness and dropped progressively to undetectable levels in subsequent weeks. He developed increasing serum levels of specific antibodies against the recombinant nucleocapsid protein of CoV-HKU1, with immunoglobulin M (IgM) titers of 1:20, 1:40, and 1:80 and IgG titers of <1:1,000, 1:2,000, and 1:8,000 in the first, second and fourth weeks of the illness, respectively. Isolation of the virus by using various cell lines, mixed neuron-glia culture, and intracerebral inoculation of suckling mice was unsuccessful. The complete genome sequence of CoV-HKU1 is a 29,926-nucleotide, polyadenylated RNA, with G+C content of 32%, the lowest among all known coronaviruses with available genome sequence. Phylogenetic analysis reveals that CoV-HKU1 is a new group 2 coronavirus. Screening of 400 NPAs, negative for SARS-CoV, from patients with respiratory illness during the SARS period identified the presence of CoV-HKU1 RNA in an additional specimen, with a viral load of  $1.13 \times 10^6$  copies per ml, from a 35-year-old woman with pneumonia. Our data support the existence of a novel group 2 coronavirus associated with pneumonia in humans.

Since no microbiological cause can be identified for a significant proportion of patients with respiratory tract infections (18, 29), research has been conducted to identify novel agents. Of the three novel agents identified in recent 3 years, including human metapneumovirus (36), severe acute respiratory syndrome (SARS) coronavirus (SARS-CoV) (25), and human coronavirus NL63 (HCoV-NL63) (6, 37), two were coronaviruses. Coronaviruses possess the largest genomes of all RNA viruses, consisting of about 30 kb. As a result of their unique mechanism of viral replication, coronaviruses have a high frequency of recombination.

Based on genotypic and serological characterization, coronaviruses were divided into three distinct groups, with human coronavirus 229E (HCoV-229E) being a group 1 coronavirus and human coronavirus OC43 (HCoV-OC43) being a group 2 coronavirus (16). They account for 5 to 30% of human respiratory tract infections. In late 2002 and 2003, the epidemic

caused by SARS-CoV affected more than 8,000 people with 750 deaths (23–25, 44, 45, 51). We have also reported the isolation of SARS-CoV-like viruses from Himalayan palm civets, which suggested that animals could be the reservoir for the ancestor of SARS-CoV (9). On the basis of genome analysis, SARS-CoV belonged to a fourth coronavirus group or alternatively was a distant relative of group 2 coronaviruses (4, 20, 28, 31, 48). Recently, a novel group 1 human coronavirus associated with respiratory tract infections, HCoV-NL63, was discovered, and its genome was sequenced (37).

In this study, we report the discovery of a novel group 2 coronavirus in the nasopharyngeal aspirates (NPAs) of patients with pneumonia. The complete genome of the coronavirus was sequenced and analyzed. Based on the findings of this study, we propose that this new virus be designated coronavirus HKU1 (CoV-HKU1).

## MATERIALS AND METHODS

**Index patient, clinical specimens, and microbiological tests.** NPAs were collected from the index patient weekly from the first till the fifth week of illness, stool and urine were collected in the first and second weeks, and sera were collected in the first, second, and fourth weeks.

The NPAs were assessed by direct antigen detection for influenza A and B viruses, parainfluenza virus types 1, 2, and 3, respiratory syncytial virus, and adenovirus by immunofluorescence (46) and were cultured for conventional

\* 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, S. K. P. Lau, and K.-y. Yuen are all principal investigators and contributed equally to the manuscript.

respiratory viruses on MDCK (canine kidney), LLC-Mk2 (rhesus monkey kidney), HEP-2 (human epithelial carcinoma), and MRC-5 (human lung fibroblast) cells. In addition, FRhK-4 (rhesus monkey kidney), A-549 (lung epithelial adenocarcinoma), BSC-1 (African green monkey kidney), CaCO2 (human colorectal adenocarcinoma), Huh-7 (human hepatoma), and Vero E6 (African green monkey kidney) cells were added to the routine panel of cell lines. Reverse transcription (RT)-PCR for influenza A virus, human metapneumovirus, and SARS-CoV was performed directly on the NPAs (25). Serological assays for antibodies against *Mycoplasma*, *Chlamydia*, *Legionella*, and SARS-CoV were performed by using SERODIA-MYCO II (Fujirebio Inc., Tokyo, Japan), *Chlamydia pneumoniae* MIF immunoglobulin G (IgG) (Focus technologies, Cypress, Calif.), indirect immunofluorescence (MRL; San Diego, Calif.), and our recently developed enzyme-linked immunosorbent assay (ELISA), respectively (45).

**RNA extraction.** Viral RNA was extracted from the NPA, urine, and fecal specimens by using the QIAamp Viral RNA Mini kit (QIAGEN, Hilden, Germany). The RNA pellet was resuspended in 10  $\mu$ l of DNase-free, RNase-free double-distilled water and was used as the template for RT-PCR.

**RT-PCR of the *pol* gene of coronaviruses, using conserved primers and DNA sequencing.** A 440-bp fragment of the RNA-dependent RNA polymerase (*pol*) gene of coronaviruses was amplified by RT-PCR with conserved primers (5'-G GTTGGGACTATCTAAGTGTGA-3' and 5'-CCATCATCAGATAGAATC ATCATA-3') designed by multiple alignment of the nucleotide sequences of available *pol* genes of known coronaviruses. RT was performed by using the SuperScript II kit (Invitrogen, San Diego, Calif.). The PCR mixture (50  $\mu$ l) contained cDNA, PCR buffer (10 mM Tris-HCl [pH 8.3], 50 mM KCl, 3 mM MgCl<sub>2</sub>, 0.01% gelatin), 200  $\mu$ M (each) deoxynucleoside triphosphates, and 1.0 U of *Taq* polymerase (Boehringer, Mannheim, Germany). The mixtures were amplified in 40 cycles of 94°C for 1 min, 48°C for 1 min, and 72°C for 1 min and a final extension at 72°C for 10 min in an automated thermal cycler (Perkin-Elmer Cetus, Gouda, The Netherlands).

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

**Complete genome sequencing and genome analysis.** The complete genome of CoV-HKU1 was amplified and sequenced by using the RNA extracted from the NPAs as a template. The RNA was converted to cDNA by a combined random-priming and oligo(dT) priming strategy. As the initial results obtained from sequencing the 440-bp fragment revealed that the polymerase (*Pol*) of CoV-HKU1 is homologous to those of other group 2 coronaviruses, the cDNA was amplified by degenerate primers designed by multiple alignment of the genomes of murine hepatitis virus (MHV) (GenBank accession no. AF201929), HCoV-OC43 (GenBank accession no. NC\_005147), bovine coronavirus (BCoV) (GenBank accession no. NC\_003045), rat sialodacryoadenitis coronavirus (SDAV) (GenBank accession no. AF207551), equine coronavirus NC99 (ECoV) (GenBank accession no. AY316300), and porcine hemagglutinating encephalomyelitis virus (PHEV) (GenBank accession no. AY078417) and additional primers designed from the results of the first and subsequent rounds of sequencing. These primer sequences are available on request. The 5' end of the viral genome was confirmed by rapid amplification of cDNA ends using the 5'/3' rapid amplification of cDNA ends kit (Roche, Mannheim, Germany). Sequences were assembled and manually edited to produce a final sequence of the viral genome. The nucleotide sequence of the genome and the deduced amino acid sequences of the open reading frames (ORFs) were compared to those of other coronaviruses. Phylogenetic tree construction was performed by using the PileUp method with GrowTree (Genetics Computer Group, Inc.). Prediction of signal peptides and their cleavage sites was performed by using SignalP (21). Protein family analysis was performed by using PFAM and InterProScan (1, 2). Prediction of transmembrane domains was performed by using TMpred and TMHMM (11, 32). PHDhtm was also used when there was disagreement between the results obtained by using TMpred and TMHMM (3). Potential N-glycosylation sites were predicted by using ScanProsite (7).

**Quantitative RT-PCR.** For real-time quantitative PCR assays, cDNA was amplified in SYBR Green I fluorescence reactions (Roche) (23). Briefly, 20  $\mu$ l of reaction mixtures containing 2  $\mu$ l of cDNA, 3.5 mM MgCl<sub>2</sub>, and 0.25 M (each) forward and reverse specific primers (5'-GGTGGGATTATCTCAATGTG A-3' and 5'-CCATCATCACTCAAAATCATCATA-3') were subjected to thermal cycling at 95°C for 10 min followed by 50 cycles of 95°C for 10 s, 55°C for 4 s, and 72°C for 18 s, using a Light cycler (Roche). A plasmid with the target sequence was used to generate the standard curve. At the end of the assay, PCR

products (440-bp fragment of *pol*) were subjected to a melting curve analysis (65 to 95°C, 0.1°C/s) to confirm the specificity of the assay.

**Cloning and purification of His<sub>6</sub>-tagged recombinant N protein of CoV-HKU1.** To produce a plasmid for protein purification, primers (5'-TTTTCCTTTTGCG GCCGCTTAAGCAACAGAGTCTTCTA-3' and 5'-CGGAATTCGATGTCT TATACCTCCCGGT-3') were used to amplify the gene encoding the N protein of CoV-HKU1 by RT-PCR. The sequence coding for amino acid residues 1 to 441 of the N protein was amplified and cloned into the EcoRI and NotI sites of expression vector pET-28b(+) (Novagen, Madison, Wis.) in frame and downstream of the series of six histidine residues. The recombinant N protein was expressed and purified by using the Ni<sup>2+</sup>-loaded HiTrap chelating system (Amersham Pharmacia) according to the manufacturer's instructions.

**Western blot analysis.** Western blot analysis was performed according to our published protocol (45). Briefly, 600 ng of purified His<sub>6</sub>-tagged recombinant N protein of CoV-HKU1 was loaded into each well of a sodium dodecyl sulfate–10% polyacrylamide gel and subsequently electroblotted onto a nitrocellulose membrane (Bio-Rad, Hercules, Calif.). The blot was cut into strips, and the strips were incubated separately with a 1:2,000 dilution of serum samples obtained during the first, second, and fourth weeks of the patient's illness. Serum samples from two healthy blood donors were used as controls. Antigen-antibody interaction was detected with an ECL fluorescence system (Amersham Life Science, Buckinghamshire, United Kingdom).

**ELISA with recombinant N protein of CoV-HKU1.** Sera from 100 healthy blood donors were used to set up a baseline for the N protein ELISA-based IgG and IgM antibody tests. The ELISA-based IgG and IgM antibody tests were modified from our previous publication (45). Briefly, each well of a Nunc (Roskilde, Denmark) immunoplate was coated with purified His<sub>6</sub>-tagged recombinant N protein (20 ng for IgG and 80 ng for IgM) for 1 h and then blocked in phosphate-buffered saline with 5% skim milk. The serum samples obtained from the patient during the first, second, and fourth weeks of the illness were serially diluted and were added to the wells of the His<sub>6</sub>-tagged recombinant N protein-coated plates in a total volume of 100  $\mu$ l and incubated at 37°C for 2 h. After five washes with washing buffer, 100  $\mu$ l of diluted horseradish peroxidase-conjugated goat antihuman IgG (1:4,000) and mouse antihuman IgM (1:1,000) antibodies (Zymed Laboratories Inc., South San Francisco, Calif.) was added to the wells and incubated at 37°C for 1 h. After washing with washing buffer five times, 100  $\mu$ l of diluted 3,3',5,5'-tetramethylbenzidine (Zymed Laboratories, Inc.) was added to each well and incubated at room temperature for 15 min. One hundred microliters of 0.3 M H<sub>2</sub>SO<sub>4</sub> was added, and the absorbance at 450 nm of each well was measured. Each sample was tested in duplicate, and the mean absorbance for each serum was calculated.

**Screening of NPAs collected during the SARS period.** Four hundred NPAs negative for SARS-CoV by RT-PCR, obtained from patients with respiratory tract infections during the SARS period in 2003 (median age 35, range 2 to 87), were screened for the presence of CoV-HKU1 RNA using the protocol described above.

**Nucleotide sequence accession number.** The nucleotide sequence of CoV-HKU1 has been lodged within the GenBank sequence database under accession no. AY597011.

## RESULTS

**Index patient and microbiological tests.** A 71-year-old Chinese man was admitted to hospital in January 2004 because of fever and productive cough with purulent sputum for 2 days. He had a history of pulmonary tuberculosis more than 40 years ago complicated by cicatrization of the right upper lobe and bronchiectasis with chronic *Pseudomonas aeruginosa* colonization of airways. He was a chronic smoker and also had chronic obstructive airway disease, hyperlipidemia, and asymptomatic abdominal aortic aneurysm. He had just returned from Shenzhen, China, 3 days before admission. A chest radiograph showed patchy infiltrates over the left lower zone. NPA for direct antigen detection of respiratory viruses, RT-PCR of influenza A virus, human metapneumovirus, and SARS-CoV, and viral cultures were negative. After the virus was determined to be a coronavirus, the NPAs were inoculated into RD (human rhabdomyosarcoma), I13.35 (murine macrophage), L929 (murine fibroblast), HRT-18 (colorectal adenocarci-

TABLE 1. Comparison of genomic features of CoV-HKU1 and other coronaviruses and amino acid identities

Coronavirus <sup>a</sup>	Genome features <sup>c</sup>		Pairwise amino acid identity (%) <sup>b</sup>							
	Size (bases)	G+C content	3CL <sup>pro</sup>	Pol	Hel	HE	S	E	M	N
Group 1										
HCoV-229E	27,317	0.38	45	54	55	NP <sup>e</sup>	31	26	35	28
PEDV	28,033	0.42	44	56	55	NP	30	34	37	37
PTGV	28,586	0.38	45	57	57	NP	32	34	37	27
CCoV	NA <sup>c</sup>	NA	NA	NA	NA	NP	31	32	36	27
HCoV-NL63	27,553	0.34	43	54	54	NP	30	28	32	28
Group 2										
CoV-HKU1	29,926	0.32	na <sup>d</sup>	na	na	na	na	na	na	na
HCoV-OC43	30,738	0.37	82	87	88	57	60	54	76	58
MHV	31,357	0.42	85	90	89	50	61	57	84	68
BCoV	31,028	0.37	84	88	88	56	61	55	76	57
SDAV	NA	NA	NA	NA	NA	50	61	60	77	62
ECoV	NA	NA	NA	NA	NA	53	61	56	78	59
PHEV	NA	NA	NA	NA	NA	54	61	54	77	57
Group 3										
IBV	27,608	0.38	41	60	57	NP	32	28	38	27
SARS-CoV	29,751	0.41	48	65	63	NP	33	27	34	31

<sup>a</sup> HCoV-229E, human coronavirus 229E; PEDV, porcine epidemic diarrhea virus; PTGV, porcine transmissible gastroenteritis virus; CCoV, canine enteric coronavirus; HCoV-NL63, human coronavirus NL63; HCoV-OC43, human coronavirus OC43; MHV, murine hepatitis virus; BCoV, bovine coronavirus; SDAV, rat sialodacryoadenitis coronavirus; ECoV, equine coronavirus NC99; PHEV, porcine hemagglutinating encephalomyelitis virus; IBV, infectious bronchitis virus; SARS-CoV, SARS coronavirus.

<sup>b</sup> Amino acid identities between the predicted chymotrypsin-like protease (3CL<sup>pro</sup>), RNA-dependent RNA polymerase (Pol), helicase (Hel), hemagglutinin-esterase (HE), spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins of CoV-HKU1 and the corresponding proteins of other coronaviruses.

<sup>c</sup> NA, not available.

<sup>d</sup> na, not applicable.

<sup>e</sup> NP, not present.

noma), and B95a (marmoset B-lymphoblastoid) cell lines and mixed neuron-glia culture. No cytopathic effect was observed. Quantitative RT-PCR, using the culture supernatants and cell lysates to monitor the presence of viral replication, also showed negative results. Moreover, intracerebrally inoculated suckling mice remained healthy after 14 days. Sputum was negative for bacterial and mycobacterial pathogens. Paired sera for antibodies against *Mycoplasma*, *Chlamydia*, *Legionella*, and SARS-CoV were negative. His symptoms improved, and he was discharged after 5 days of hospitalization.

**RT-PCR of the *pol* gene of coronaviruses by using conserved primers and DNA sequencing.** RT-PCR of the *pol* gene from the patient's NPA showed a band of about 440 bp. Sequencing of the band showed 91% amino acid and 84% nucleotide identity to the corresponding sequence in MHV (GenBank accession no. AF201929), 89% amino acid and 82% nucleotide identity to HCoV-OC43 (GenBank accession no. NC\_005147),

and 89% amino acid and 82% nucleotide identity to BCoV (GenBank accession no. NC\_003045).

**Genome analysis.** The genome of CoV-HKU1 is a 29,926-nucleotide, polyadenylated RNA. The G+C content is 32%, the lowest among all known coronaviruses with genome sequence available (Table 1). The genome organization is the same as that of other coronaviruses, with the characteristic gene order 5'-replicase, spike (S), envelope (E), membrane (M), nucleocapsid (N)-3'. Both 5' and 3' ends contain short untranslated regions. The 5' end of the genome consists of a putative 5' leader sequence (17, 19). A putative transcription regulatory sequence (TRS) motif, 5'-AAUCUAAAC-3' (as in MHV and BCoV), or alternatively, 5'-UAAUUAUUAUG, was found at the 3' end of the leader sequence and precedes each translated ORF except ORF5 (Table 2). As in SDAV and MHV, ORF5, which encodes the putative E protein, may share the same TRS with ORF4, suggesting that the translation of

TABLE 2. Coding potential and putative transcription regulatory sequences of the CoV-HKU1 genome sequence

ORF	Start to end (nucleotide position)	No. of nucleotides	No. of amino acids	Frame	Nucleotide position in genome	Putative TRS
						TRS sequence <sup>a</sup>
ORF 1a	206–13600	13,395	4,465	+2	63	UUA <u>AAU</u> CUAAACUUUUUAA (127) AUG
ORF 1b	13600–21753	8,154	2,717	+1		
ORF 2 (HE)	21773–22933	1,161	386	+2	21763	UUA <u>AAU</u> CUAAACU <u>AUG</u>
ORF 3 (S)	22942–27012	4,071	1,356	+1	22933	UUA <u>AAU</u> CUAAAC <u>AUG</u>
ORF 4	27051–27380	330	109	+3	27035	UUA <u>AAU</u> CUAAACUUUAUUU <u>AUG</u>
ORF 5 (E)	27373–27621	249	82	+1		
ORF 6 (M)	27633–28304	672	223	+3	27621	CUA <u>AAU</u> CUAAACAUUA <u>AUG</u>
ORF 7 (N)	28320–29645	1,326	441	+3	28304	UUA <u>AAU</u> CUAAACUAAUAGGA <u>AUG</u>
ORF 8	28342–28959	618	205	+1	28304	UUA <u>AAU</u> CUAAACUAAUAGGAUGUCUUAUACUCCGGUCAUUA <u>AUG</u>

<sup>a</sup> Boldface type indicates putative initiation codon. Underlining indicates core sequence of TRS motif identical to the 3' end of the leader sequence.

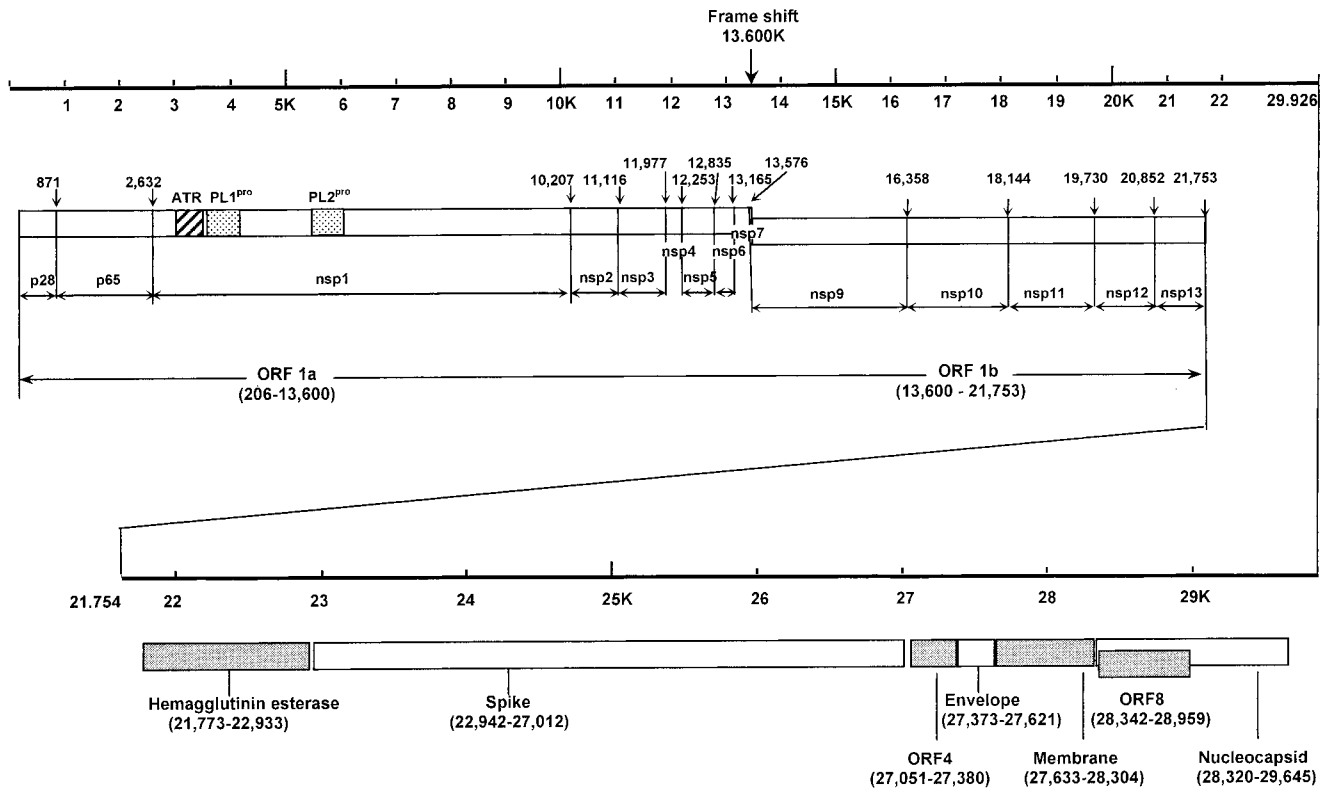


FIG. 1. Genome organization of CoV-HKU1. Overall organization of the 29,926-nucleotide CoV-HKU1 genomic RNA. Predicted ORFs 1a and 1b, encoding the nonstructural polyproteins (p28, p65, and nsp1 to -13) and those encoding the hemagglutinin-esterase, spike, envelope, membrane and nucleocapsid structural proteins are indicated. Arrows indicate putative cleavage sites (with the corresponding nucleotide positions) of the replicase polyprotein encoded by ORF 1a and ORF 1b. ATR and PL1<sup>pro</sup> and PL2<sup>pro</sup> represent the acidic tandem repeat and the two papain-like proteases, respectively, in nsp1.

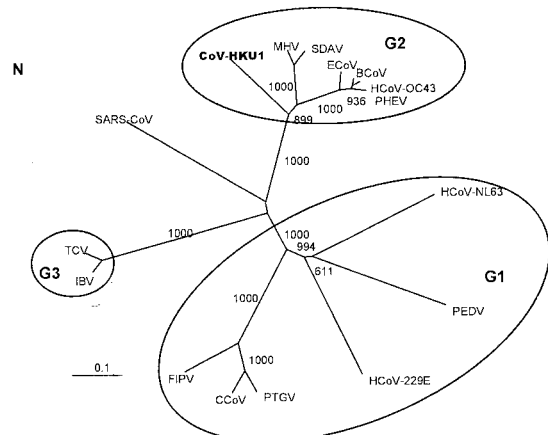
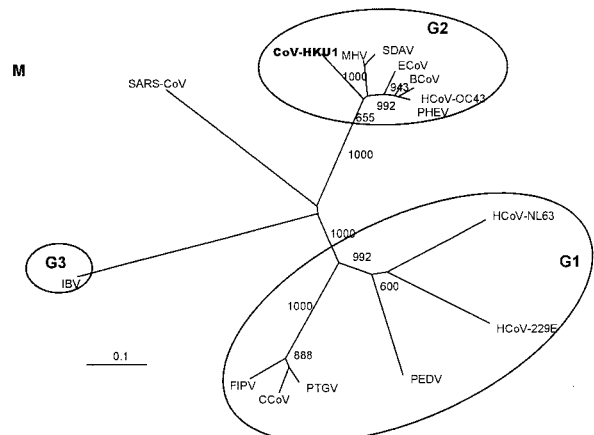
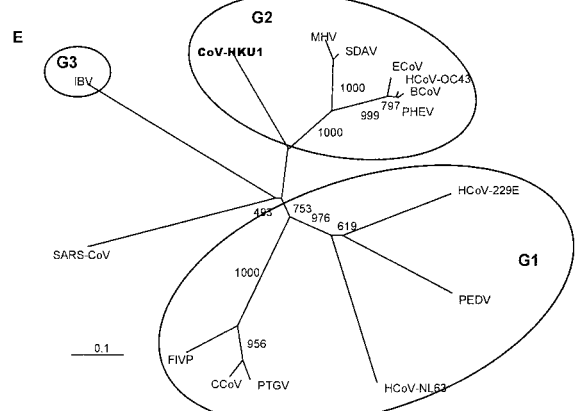
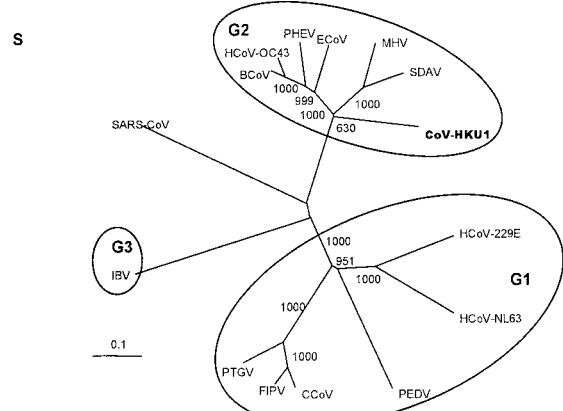
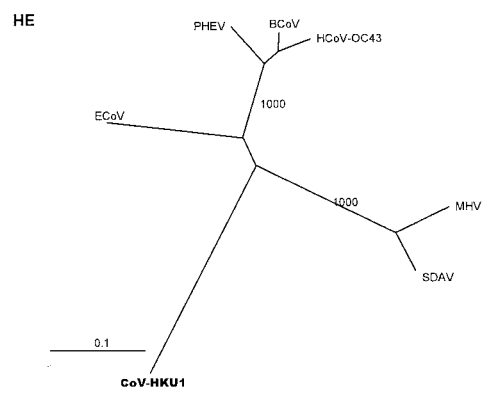
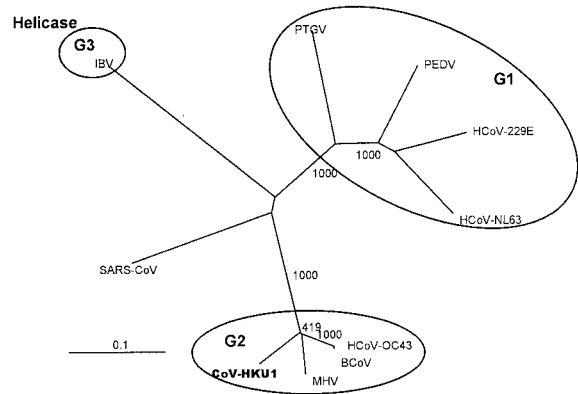
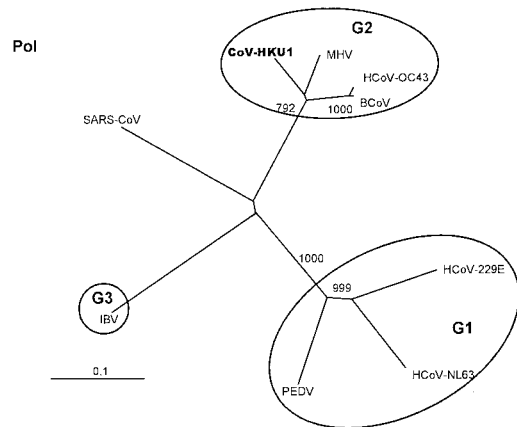
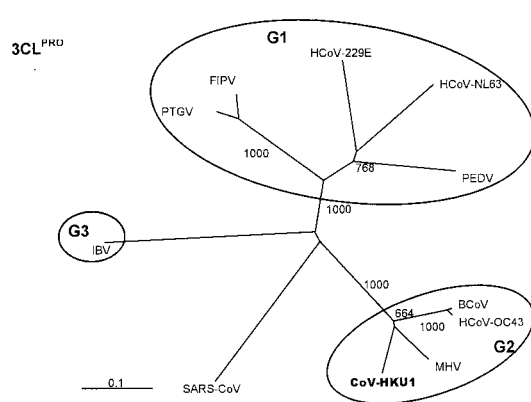
the E protein is cap independent, possibly via an internal ribosomal entry site (IRES) (34). A stretch of 13 nucleotides, AUUUAUUGUUUGG (similar to the IRES element, UUUUAUUCUUUUU, in MHV), upstream of the initiation codon of the E protein is present in CoV-HKU1 (12). Further experiments would determine if this sequence acts as an IRES for this ORF and whether 5'-UAAUCUAAAC-3' or 5'-AAUCUAAAC-3' is the real TRS for CoV-HKU1. Of note is that 5'-AAUCUAAAC-3' and 5'-UAAUCUAAAC-3' are also observed at nucleotide positions 19528 and 22518 of the genome, respectively, neither of which precedes an ORF of obvious significance. Analysis of more genomes of CoV-HKU1 would reveal whether this is a consistent feature and its possible role in recombination of the CoV-HKU1 genome. The 3' untranslated region contains a predicted bulged stem-loop structure 2 to 66 nucleotides downstream of N gene (nucleotide position 29647 to 29711). This bulged stem-loop structure is conserved in group 2 coronaviruses (8). Downstream to the bulged stem-loop structure, 63 to 115 nucleotides downstream of the N gene (nucleotide position 29708 to 29760), a pseudoknot structure is present. This pseudoknot structure is conserved among coronaviruses and plays a role in coronavirus RNA replication (42).

The coding potential of the CoV-HKU1 genome is shown in Fig. 1 and Table 2, and the phylogenetic analysis of the chy-

motrypsin-like protease (3CL<sup>pro</sup>), Pol, helicase, hemagglutinin-esterase (HE), S, E, M, and N is shown in Fig. 2.

The replicase 1a ORF (nucleotide position 206 to 13600) and replicase 1b ORF (nucleotide position 13600 to 21753) occupy 21.5 kb of the CoV-HKU1 genome. Similar to the case with other coronaviruses, a frame shift interrupts the protein-coding regions and separates ORFs 1a and 1b. This ORF encodes a number of putative proteins, including nsp1 (which contains the putative papain-like proteases), nsp2 (the putative 3CL<sup>pro</sup>), nsp9 (the putative Pol), nsp10 (the putative helicase), and other proteins with unknown functions. These proteins are produced by proteolytic cleavage of the large replicase polyprotein. The arrangement of the resulting putative proteins is the same as that in the MHV genome (Fig. 3). This polyprotein is synthesized by a -1 ribosomal frameshift at a conserved site (UUUAAAC) upstream of a pseudoknot structure at the junction of ORF 1a and ORF 1b. This ribosomal frameshift would result in a polyprotein of 7,182 amino acids, which has 75 to 77% amino acid identities with the polyproteins of other group 2 coronaviruses and 43 to 47% amino acid identities with the polyproteins of non-group 2 coronaviruses. The Pol of CoV-HKU1, with 928 amino acids, has 87 to 90% amino acid identities with the Pol of other group 2 coronaviruses and 54 to 65% amino acid identities with the Pol of non-group 2 coronaviruses (Table 1 and Fig. 2). The catalytic





histidine and cysteine amino acid residues, conserved among the 3CL<sup>pro</sup> in all coronaviruses, are present in the predicted 3CL<sup>pro</sup> of CoV-HKU1 (amino acids His<sup>3375</sup> and Cys<sup>3479</sup> of ORF 1a). nsp1, which corresponds to p210 in MHV, contains two papain-like proteases (PL<sup>pro</sup>), PL1<sup>pro</sup> and PL2<sup>pro</sup>. In the N terminus of nsp1 (amino acid residues 945 to 1104 of ORF 1a), there are 14 tandem copies of a 30-base repeat which encodes NDDEDVVTGD, followed by two 30-base regions that encode NNDEEIVTGD and NDDQIVVTGD, located inside the acidic domain upstream of PL1<sup>pro</sup> (Fig. 3). This acidic tandem repeat (ATR) is not observed in other coronaviruses. The presence of this ATR is confirmed by sequencing the corresponding part of the genome from two NPAs collected 1 week apart. The presence of the repeat does not result in a marked change in the isoelectric point of the acidic domain (3.31 in CoV-HKU1 versus 3.92 in MHV) or the predicted secondary structure (random coil in both CoV-HKU1 and MHV). Moreover, the characteristic amino acid residues for proteolytic cleavage by the two PL<sup>pro</sup>, determined by mutagenesis studies, located at the junctions of p28/p65, p65/nsp1, and nsp1/nsp2 in MHV, are all present in the corresponding positions in CoV-HKU1 (13). Furthermore, the zinc finger domain proposed to possess nonproteolytic activity in other coronaviruses is also present in PL1<sup>pro</sup> of CoV-HKU1 (10).

ORF 2 (nucleotide position 21773 to 22933) encodes the predicted HE glycoprotein with 386 amino acids. HE is present in group 2 coronaviruses and influenza C virus. The HE of CoV-HKU1 has 50 to 57% amino acid identities with the HE of other group 2 coronaviruses (Table 1 and Fig. 2). PFAM and InterProScan analysis of the ORF shows that amino acid residues 1 to 349 of the predicted protein constitute a member of the hemagglutinin esterase family (PFAM accession no. PF03996 and INTERPRO accession no. IPR007142). Furthermore, PFAM and InterProScan analysis shows that amino acid residues 122 to 236 of the predicted protein constitute the hemagglutinin domain of the HE fusion glycoprotein family (PFAM accession no. PF02710 and INTERPRO accession no. IPR003860). SignalP analysis reveals a signal peptide probability of 0.738, with a cleavage site between residues 13 and 14. Although TMPred and TMHMM analysis of the ORF shows four and three transmembrane domains, respectively, PHDhtm analysis shows only one transmembrane domain, at positions 354 to 376. This concurs with only one transmembrane region reported in the C terminus of the HE of BCoV and puffinosis virus (14). PrositeScan analysis of the HE protein of CoV-HKU1 reveals eight potential N-linked glycosylation (six NXS and two NXT) sites. These are located at positions 83 (NYT), 110 (NGS), 145 (NVS), 168 (NYS), 193 (NFS), 286 (NSS), 314 (NVS), and 328 (NFT). The putative active site for neuraminidase *O*-acetyl-esterase activity, FGDS, is

located at positions 31 to 34 (39). In BCoV, it has been shown that HE is required for viral replication in one study (38) but is not essential for viral infection under some specific experimental conditions (26). In MHV, the expression of HE is heterogeneous, depending on the number of copies of UCUGAA in the leader sequence, the presence of initiation codon, upstream promoter, and a complete ORF with C-terminal transmembrane anchor (49), and appears to be related to central nervous system tropism (50). In CoV-HKU1, the initiation codon and a complete ORF are present. Since the HE of CoV-HKU1 is quite distantly related to the HE of MHV and BCoV/HCoV-OC43 (Fig. 2), further experiments have to be performed to determine the essentiality and function of HE in CoV-HKU1.

ORF 3 (nucleotide position 22942 to 27012) encodes the predicted S glycoprotein (PFAM accession no. PF01601) with 1,356 amino acids. The S protein of CoV-HKU1 has 60 to 61% amino acid identities with the S proteins of other group 2 coronaviruses but less than 35% amino acid identities with the S proteins of non-group 2 coronaviruses (Table 1 and Fig. 2). InterProScan analysis predicts it as a type I membrane glycoprotein. Important features of the S protein of CoV-HKU1 are depicted in Fig. 4. PrositeScan of the S protein of CoV-HKU1 revealed 28 potential N-linked glycosylation (12 NXS and 16 NXT) sites. SignalP analysis revealed a signal peptide probability of 0.909, with a cleavage site between residues 13 and 14. By multiple alignments with the S proteins of other group 2 coronaviruses, a potential cleavage site located after RRKR, between residues 760 and 761, where S will be cleaved into S1 and S2, was identified. Immediately upstream to RRKR, there is a series of five serine residues that are not present in any other known coronaviruses (Fig. 4). Most of the S protein (residues 15 to 1300) is exposed on the outside of the virus, with a transmembrane domain at the C terminus (TMHMM analysis of the ORF shows one transmembrane domain at positions 1301 to 1356), followed by a cytoplasmic tail rich in cysteine residues. Two heptad repeats, located at residues 982 to 1083 (HR1) and 1250 to 1297 (HR2), identified by multiple alignments with other coronaviruses, are present. The receptor for S protein binding in MHV and HCoV-OC43 are CEACAM1 and sialic acid, respectively (15, 41, 43). While the three conserved regions (sites I, II, and III) and amino acid residues (Thr<sup>62</sup>, Thr<sup>212</sup>, Tyr<sup>214</sup>, and Tyr<sup>216</sup>) in the N-terminal of the MHV S protein important for receptor-binding activity (33) are present in CoV-HKU1 (Fig. 4), the amino acid residues on the S protein of HCoV-OC43 that are important for receptor binding are not well defined. Further experiments should be performed to delineate the receptor for CoV-HKU1.

ORF 4 (nucleotide position 27051 to 27380) encodes a predicted protein with 109 amino acids. This ORF overlaps with

FIG. 2. Phylogenetic analysis of chymotrypsin-like protease (3CL<sup>pro</sup>), RNA-dependent RNA polymerase (Pol), helicase, hemagglutinin-esterase (HE), spike (S), envelope (E), membrane (M), and nucleocapsid (N) of CoV-HKU1. The trees were constructed by the neighbor-joining method, using Jukes-Cantor correction and bootstrap values calculated from 1,000 trees. Three hundred three, 928, 595, 418, 1356, 75, 225, and 406 amino acid positions in 3CL<sup>pro</sup>, Pol, helicase, HE, S, E, M and N, respectively, were included in the analysis. The scale bar indicates the estimated number of substitutions per 10 amino acids. HCoV-229E, human coronavirus 229E; PEDV, porcine epidemic diarrhea virus; PTGV, porcine transmissible gastroenteritis virus; CCoV, canine enteric coronavirus; HCoV-NL63, human coronavirus NL63; HCoV-OC43, human coronavirus OC43; MHV, murine hepatitis virus; BCoV, bovine coronavirus; SDAV, rat sialodacryoadenitis coronavirus; ECoV, equine coronavirus NC99; PHEV, porcine hemagglutinating encephalomyelitis virus; IBV, infectious bronchitis virus; SARS-CoV, SARS coronavirus.

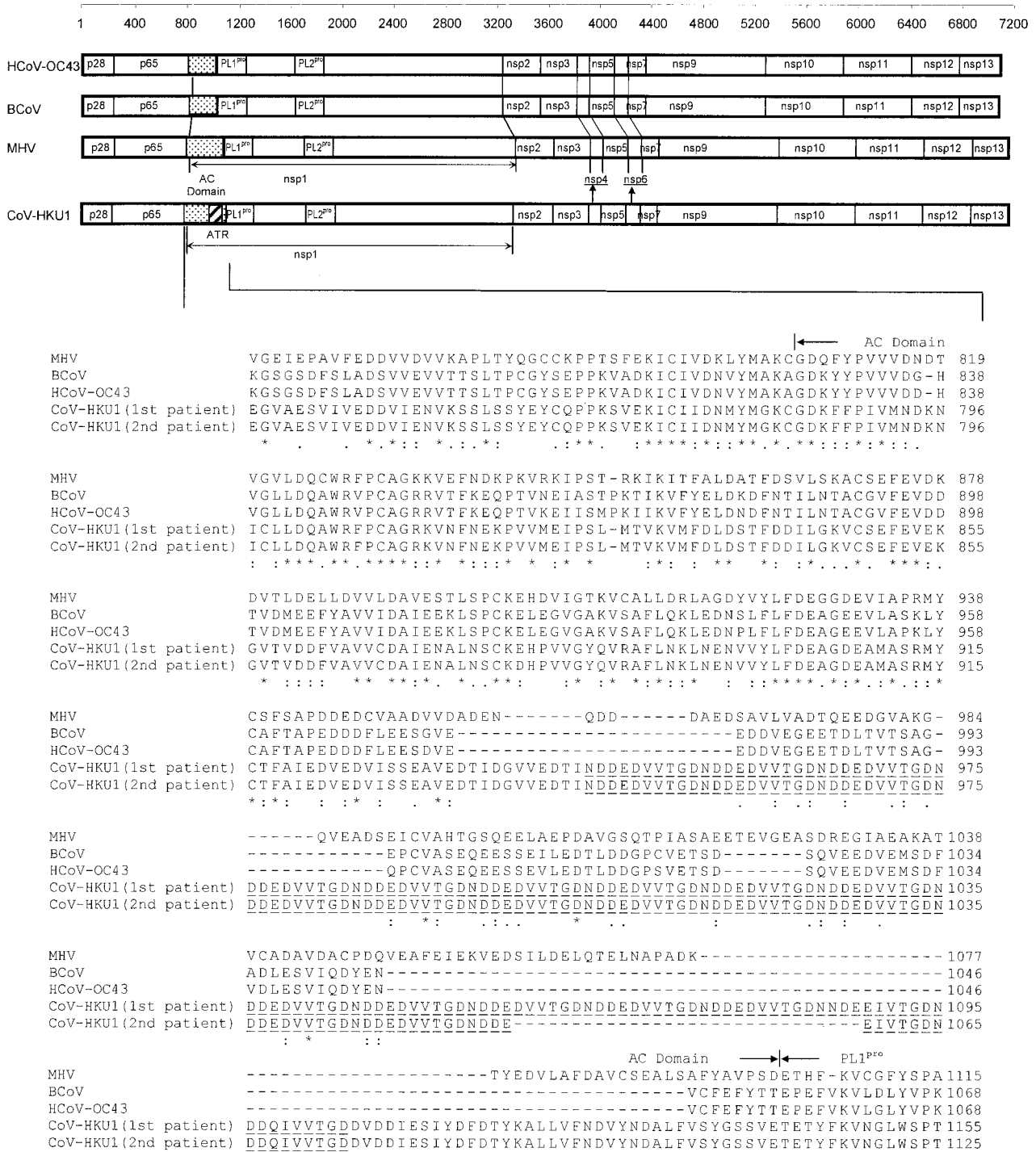


FIG. 3. Arrangements of proteins in replicase polyprotein in HKU1 compared with those in HCoV-OC43, BCoV, and MHV. Alignment of the AC domains of HCoV-OC43, BCoV, and MHV and the AC domains and ATR (underlined) of CoV-HKU1 in the two patients was generated with ClustalX 1.83. AC domain, acidic domain. GenBank accession numbers are as follows: MHV, NC\_001846; BCoV, NC\_003045; HCoV-OC43, AY585229.

the ORF that encodes the E protein. PFAM analysis of the ORF shows that the predicted protein is a member of the coronavirus nonstructural protein NS2 family (PFAM accession no. PF04753). TMpred and TMHMM analysis does not

reveal any transmembrane helix. This predicted protein of CoV-HKU1 has 44 to 51% amino acid identities with the corresponding proteins of other group 2 coronaviruses.

ORF 5 (nucleotide position 27373 to 27621) encodes the



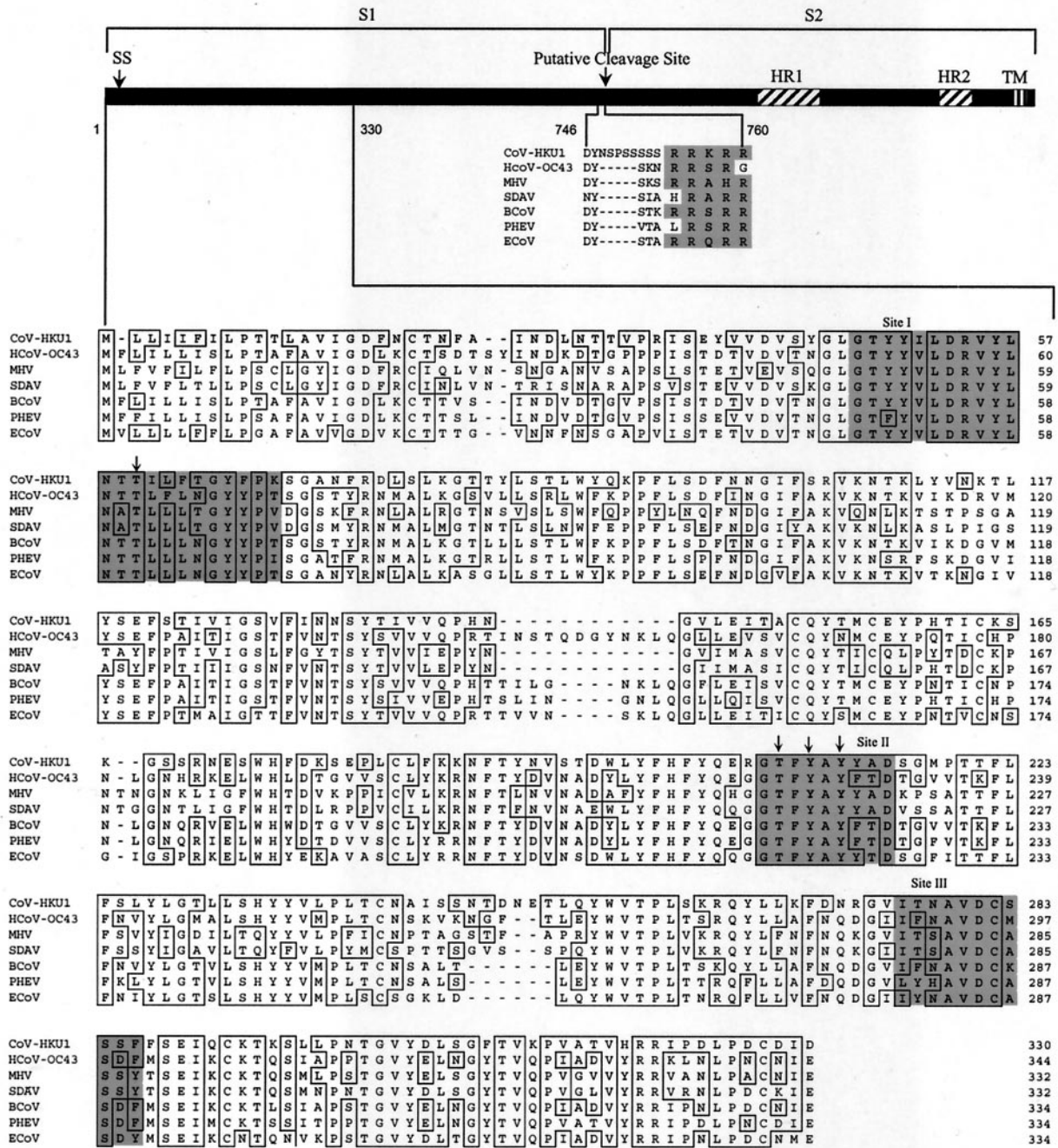


FIG. 4. Spike protein of CoV-HKU1. The spike protein (1,356 amino acids) of CoV-HKU1 is depicted by the horizontal bar. SS, N terminal signal sequence (amino acid residues 1 to 13); HR1, heptad repeat 1 (amino acid residues 982 to 1083); HR2, heptad repeat 2 (amino acid residues 1250 to 1297); TM, transmembrane domain (amino acid residues 1301 to 1323). Alignment of the N-terminal region important for receptor binding (amino acid residues 1 to 330) and the region upstream of the cleavage site between S1 and S2 of CoV-HKU1 and other group 2 coronaviruses was done with ClustalX 1.83. Residues that match the CoV-HKU1 sequence exactly are boxed. The three conserved regions (sites I, II, and III) for receptor binding in MHV are shaded. The positions of the four conserved amino acids important for receptor binding in MHV are indicated with arrows. GenBank accession numbers were as follows: MHV, P11224; BCoV, NP\_150077; HCoV-OC43, NP\_937950; SDAV, AAF97738; PHEV, AAL80031; ECoV, AAQ67205.

predicted E protein with 82 amino acids. The E protein of CoV-HKU1 has 54 to 60% amino acid identities with the E proteins of other group 2 coronaviruses but less than 35% amino acid identities with the E proteins of non-group 2 coronaviruses (Table 1 and Fig. 2). PFAM and InterProScan anal-

ysis of the ORF shows that the predicted E protein is a member of the nonstructural protein NS3/small envelope protein E family (PFAM accession no. PF02723). SignalP analysis predicts the presence of a transmembrane anchor (probability 0.995). TMpred analysis of the ORF shows two transmem-



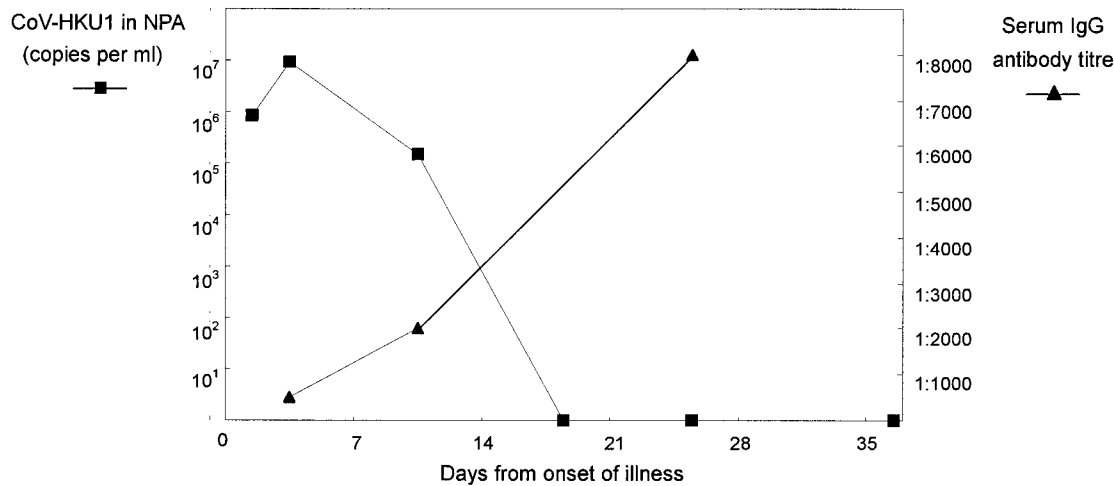


FIG. 5. Sequential quantitative RT-PCR for CoV-HKU1 in NPAs and serum IgG titers against N protein of CoV-HKU1.

brane domains at positions 16 to 34 and 39 to 59, and TMHMM analysis of the ORF shows two transmembrane domains at positions 10 to 32 and 39 to 58, consistent with the anticipated association of the E protein with the viral envelope.

ORF 6 (nucleotide position 27633 to 28304) encodes the predicted M protein with 223 amino acids. The M protein of CoV-HKU1 has 76 to 84% amino acid identities with the M proteins of other group 2 coronaviruses but less than 40% amino acid identities with the M proteins of non-group 2 coronaviruses (Table 1 and Fig. 2). PFAM analysis of the ORF shows that the predicted M protein is a member of the coronavirus matrix glycoprotein family (PFAM accession no. PF01635). SignalP analysis predicts the presence of a transmembrane anchor (probability, 0.926). TMPred analysis of the ORF shows three transmembrane domains at positions 21 to 42, 53 to 74, and 77 to 98. TMHMM analysis of the ORF shows three transmembrane domains at positions 20 to 39, 46 to 68, and 78 to 100. The N-terminal 19 to 20 amino acids are located on the outside, and the C-terminal 123- to 125-amino-acid hydrophilic domain is located on the inside of the virus.

ORF 7 (nucleotide position 28320 to 29645) encodes the predicted N protein (PFAM accession no. PF00937) with 441 amino acids. The N protein of CoV-HKU1 has 57 to 68% amino acid identities with the N proteins of other group 2 coronaviruses but less than 40% amino acid identities with the N proteins of non-group 2 coronaviruses (Table 1 and Fig. 2).

ORF 8 (nucleotide position 28342 to 28959) encodes a hypothetical protein (N2) of 205 amino acids within the ORF that encodes the predicted N protein. PFAM analysis of the ORF shows that the predicted protein is a member of the coronavirus nucleocapsid I protein family (PFAM accession no. PF03187). This hypothetical N2 protein of CoV-HKU1 has 32 to 39% amino acid identities with the N2 proteins of other group 2 coronaviruses. This protein has been shown to be nonessential for viral replication in MHV (5).

**Quantitative RT-PCR.** Quantitative RT-PCR showed that the amounts of CoV-HKU1 RNA were  $8.5 \times 10^5$  and  $9.6 \times 10^6$  copies per ml in two NPAs collected in the first week of the illness and  $1.5 \times 10^5$  copies per ml in the NPA collected in the second week of the illness, but CoV-HKU1 RNA was unde-

tectable in the NPAs collected in the third, fourth, and fifth weeks of the illness (Fig. 5). CoV-HKU1 RNA was undetectable in all urine and stool specimens.

**Purification of His<sub>6</sub>-tagged recombinant N protein and Western blot analysis.** To produce recombinant N protein of CoV-HKU1, the recombinant N protein was expressed in *Escherichia coli* and subsequently purified. The purified recombinant N protein was separated on sodium dodecyl sulfate-polyacrylamide gels followed by Western blot analysis with serum samples. Several prominent immunoreactive bands were visible for serum samples collected during the second and fourth weeks of the patient's illness (Fig. 6, lanes 2 and 3). The sizes of the largest bands were about 53 kDa, consistent with the expected size of 52.8 kDa for the full-length His<sub>6</sub>-tagged

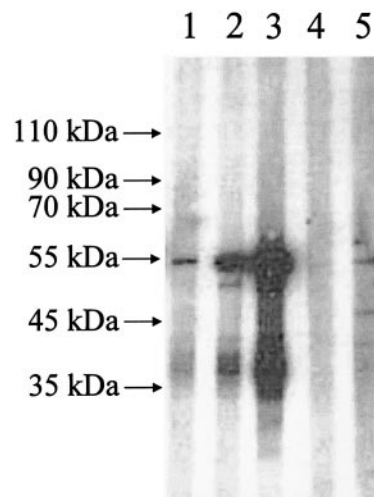


FIG. 6. Western blot analysis of purified recombinant CoV-HKU1 N protein antigen. Prominent immunoreactive protein bands of about 53 kDa were visible on the Western blot that used recombinant N protein as the antigen during the second and fourth weeks of the patient's illness (lanes 2 and 3). Only very faint bands were observed for serum samples obtained from the patient during the first week of the illness (lane 1) and two healthy blood donors (lanes 4 and 5).

recombinant N protein, whereas the other bands were probably its degradation products. Only very faint bands were observed for serum samples obtained from the patient during the first week of the illness (Fig. 6, lane 1) and two healthy blood donors (Fig. 6, lanes 4 and 5).

**ELISA using recombinant N protein of CoV-HKU1.** An ELISA-based antibody test was developed with this recombinant N protein for the detection of specific antibodies against this protein. Box titration was carried out with serial dilutions of recombinant N protein coating antigen (in one axis) and serum (in the other axis) obtained from the fourth week of the patient's illness. The results identified 20 and 80 ng of purified recombinant N protein per well as the ideal amounts for plate coating and 1:1,000 and 1:20 as the most optimal serum dilutions for IgG and IgM detection, respectively.

To establish the baseline for the ELISA tests, serum samples (diluted at 1:1,000 and 1:20 for IgG and IgM, respectively) from 100 healthy blood donors were tested. The mean ELISA optical densities at 450 nm for IgG and IgM detection were 0.178 and 0.224, with standard deviations of 0.070 and 0.117, respectively. Absorbance values of 0.387 and 0.576 were selected as the cutoff values (means plus three standard deviations) for IgG and IgM, respectively. Using these cutoffs, the titers for IgG of the patient's sera obtained during the first, second, and fourth weeks of the illness were <1:1,000, 1:2,000, and 1:8,000, respectively, and those for IgM were 1:20, 1:40, and 1:80, respectively (Fig. 5).

**Screening of NPAs during the SARS period.** Among the 400 NPAs that were negative for SARS-CoV by RT-PCR, obtained during the SARS period in 2003, one was positive for RNA of CoV-HKU1. The NPA was obtained from a 35-year-old, previously healthy woman with pneumonia of unknown etiology in March 2003, 10 months earlier than the index case. There was no direct relationship or contact between the two cases. The detection of several unique features upon sequencing confirmed the presence of CoV-HKU1. Sequencing of the 2,784-bp fragment that encodes Pol revealed 87 base (3.1%) and seven (0.8%) amino acid differences between the Pol of this virus and that of the virus from the index patient. Sequencing of the fragment that encodes nsp1 showed that 11 ATR are present, compared to 14 ATR in the fragment from the index patient (Fig. 3). This indicates that the ATR is probably a consistent feature in nsp1 of CoV-HKU1 and may also be a region of frequent insertion and deletion. Sequencing of the replicase polyprotein/HE junction revealed that NS2a, absent from the virus of the index patient, is also absent from this virus. The amount of CoV-HKU1 RNA in the NPA was  $1.13 \times 10^6$  copies per ml. Since the convalescent-phase serum is not available from this patient, antibody response cannot be determined.

## DISCUSSION

We report the characterization and complete genome sequence of a novel coronavirus detected in the NPAs of patients with pneumonia. The clinical significance of the virus in the index patient was made evident by the high viral loads in the patient's NPAs during the first week of his illness, which coincided with his acute symptoms. The viral load decreased during the second week of the illness and was undetectable in the

third week. In addition, the fall in viral load was accompanied by the recovery from the illness and development of a specific antibody response to the recombinant N protein of the virus. The fact that the present virus could not be recovered from cell cultures could be related to the lack of a susceptible cell line for CoV-HKU1 or the inherently low recovery rate of some coronaviruses. Many decades after the recognition of HCoV-229E and HCoV-OC43, the other non-SARS human respiratory coronaviruses known to cause pneumonia at low frequencies (27, 35, 40), there are still only a few primary virus isolates available, and organ culture is required for primary isolation of HCoV-OC43. In our experience, SARS-CoV can be recovered only from less than 20% of patients with serologically and RT-PCR-documented SARS-CoV pneumonia. After the discovery of CoV-HKU1 in the index patient, we conducted a preliminary study on 400 NPAs that were collected last year during the SARS period. Among these 400 NPAs, CoV-HKU1 was detected in one specimen, with a viral load comparable to that of the index patient. These results suggested that CoV-HKU1 is not only an incidental finding in an isolated patient but a previously unrecognized coronavirus associated with pneumonia.

Genomic analysis reveals that CoV-HKU1 is a group 2 coronavirus. The genome organization of CoV-HKU1 concurs with those of other coronaviruses, with the characteristic gene order 5'-replicase, S, E, M, N-3', short untranslated regions in both 5' and 3' ends, 5' conserved coronavirus core leader sequence, putative TRS upstream of multiple ORFs, and conserved pseudoknot in the 3' untranslated region. CoV-HKU1 contains certain features that are characteristic of group 2 coronaviruses, including the presence of HE, ORF 4, and N2. Phylogenetic analysis of the 3CL<sup>pro</sup>, Pol, helicase, S, E, M, and N proteins showed that these genes of CoV-HKU1 were clustered with the corresponding genes in other group 2 coronaviruses. However, the proteins of CoV-HKU1 formed distinct branches in the phylogenetic trees, indicating that CoV-HKU1 is a distinct member within the group and is not very closely related to any other known members of group 2 coronaviruses (Fig. 2).

CoV-HKU1 exhibits additional features that are distinct from those of other group 2 coronaviruses. Compared to other group 2 coronaviruses, there is a deletion of about 800 bp between the replicase ORF 1b and the HE ORF in CoV-HKU1. In other group 2 coronaviruses, including MHV, SDAV, HCoV-OC43, and BCoV, an ORF of 798 to 837 bp (273 to 278 amino acids) is present between the replicase ORF 1b and the HE ORF. This ORF encodes protein of the coronavirus nonstructural protein NS2a family (PFAM accession no. PF05213). Further experiments will reveal if this is a non-essential gene in other coronaviruses, as in MHV (30), and if it serves virus-specific functions in different group 2 coronaviruses. In addition to the deletion, upstream to PL1<sup>pro</sup> in ORF 1a, there are 14 tandem copies of a 30-base repeat that codes for a highly acidic domain. Similar repeats, with different amino acid compositions, have been found in the genomes of human, rat, and parasites but not in other coronaviruses (22, 47). The function of these repeats is not well understood, although some authors have suggested that they could be important antigens, and their biological role may be related to their special three-dimensional structure. The vitellaria anti-

genic protein of *Clonorchis sinensis* contains 23 tandem copies of a 30-bp repeat that codes for DGGAPPKSG (47). In the case of *Plasmodium falciparum*, it has been shown that the antigenicity of the circumsporozoite protein is due to its repeating epitope structure (22). It has also been suggested that the tandemly repeated peptide may induce a strong humoral immune response in the infected host and thus may also be useful in serological diagnosis. Further experiments should be performed to delineate the antigenic properties, biological role, and possible clinical usefulness of this tandem repeat in CoV-HKU1.

The prevalence of CoV-HKU1 in humans as a cause of respiratory tract infections remains to be determined. HCoV-OC43, HCoV-229E, and probably HCoV-NL63 are endemic in humans. On the other hand, isolation of SARS-CoV-like coronavirus from civet cats and the absence of a resurgent SARS epidemic in 2004 apart from sporadic laboratory-acquired cases imply that SARS-CoV probably originated from animals. For CoV-HKU1, the detection of its existence in the NPAs of two patients almost 1 year apart suggests that it may have been endemic in humans, or alternatively, it may originally have been an animal coronavirus but may have crossed the species barrier in the past few years. In the serological experiments, Western blot analysis revealed that the serum samples of the two healthy blood donors showed some antigen-antibody reaction with the purified N protein of CoV-HKU1 (Fig. 6). It is not known whether these were due to cross-reaction between the N protein of CoV-HKU1 and that of HCoV-OC43, since these two proteins showed 58% amino acid identity, or due to past infections by CoV-HKU1. Further clinical, seroepidemiological, and phylogenetic studies would be required to determine the relative importance of CoV-HKU1 compared to other respiratory tract viruses in causing upper and lower respiratory tract infections, its seroprevalence, and the origin of the virus.

#### ACKNOWLEDGMENTS

This work was partly supported by the Research Grant Council Grant, Research Fund for the Control of Infectious Diseases, "One Mouth, One Mask" Fund, Suen Chi Sun Charitable Foundation, the Public Health Research Grant A195357 from the National Institutes of Allergy and Infectious Diseases, and the William Benter Infectious Disease Fund.

#### REFERENCES

1. Apweiler, R., T. K. Attwood, A. Bairoch, A. Bateman, E. Birney, M. Biswas, P. Bucher, L. Cerutti, F. Corpet, M. D. Croning, R. Durbin, L. Falquet, W. Fleischmann, J. Gouzy, H. Hermjakob, N. Hulo, I. Jonassen, D. Kahn, A. Knapin, Y. Karavidopoulou, R. Lopez, B. Marx, N. J. Mulder, T. M. Oinn, M. Pagni, F. Servant, C. J. Sigrist, and E. M. Zdobnov. 2001. The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* **29**:37–40.
2. Bateman, A., E. Birney, L. Cerruti, R. Durbin, L. Etwiller, S. R. Eddy, S. Griffiths-Jones, K. L. Howe, M. Marshall, and E. L. Sonnhammer. 2002. The Pfam protein families database. *Nucleic Acids Res.* **30**:276–280.
3. Combet, C., C. Blanchet, C. Geourjon, and G. Deléage. 2000. NPS@: network protein sequence analysis. *Trends Biochem. Sci.* **25**:147–150.
4. Eickmann, M., S. Becker, H. D. Klenk, H. W. Doerr, K. Stadler, S. Censini, S. Guidotti, V. Masignani, M. Scarselli, M. Mora, C. Donati, J. H. Han, H. C. Song, S. Abrignani, A. Covacci, and R. Rappuoli. 2003. Phylogeny of the SARS coronavirus. *Science* **302**:1504–1505.
5. Fischer, F., D. Peng, S. T. Hingley, S. R. Weiss, and P. S. Masters. 1997. The internal open reading frame within the nucleocapsid gene of mouse hepatitis virus encodes a structural protein that is not essential for viral replication. *J. Virol.* **71**:996–1003.
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. Gattiker, A., E. Gasteiger, and A. Bairoch. 2002. ScanProsite: a reference implementation of a PROSITE scanning tool. *Appl. Bioinformatics* **1**:107–108.
8. Goebel, S. J., B. Hsue, T. F. Dombrowski, and P. S. Masters. 2004. Characterization of the RNA components of a putative molecular switch in the 3' untranslated region of the murine coronavirus genome. *J. Virol.* **78**:669–682.
9. 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.
10. Herold, J., S. G. Siddell, and A. E. Gorbalenya. 1999. A human RNA viral cysteine proteinase that depends upon a unique Zn<sup>2+</sup>-binding finger connecting the two domains of a papain-like fold. *J. Biol. Chem.* **274**:14918–14925.
11. Hofmann, K., and W. Stoffel. 1993. TMbase—a database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **374**:166.
12. 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.
13. Kanjanahaluethai, A., D. Jukneliene, and S. C. Baker. 2003. Identification of the murine coronavirus MP1 cleavage site recognized by papain-like proteinase 2. *J. Virol.* **77**:7376–7382.
14. Klaussegger, A., B. Strobl, G. Regl, and A. Kaser. 1999. Identification of a coronavirus hemagglutinin-esterase with a substrate specificity different from those of influenza C virus and bovine coronavirus. *J. Virol.* **73**:3737–3743.
15. Krempl, C., B. Schultze, and G. Herrler. 1995. Analysis of cellular receptors for human coronavirus OC43. *Adv. Exp. Med. Biol.* **380**:371–374.
16. Lai, M. M., and D. Cavanagh. 1997. The molecular biology of coronaviruses. *Adv. Virus Res.* **48**:1–100.
17. Lai, M. M., R. S. Baric, P. R. Brayton, and S. A. Stohman. 1984. Characterization of leader RNA sequences on the virion and mRNAs of mouse hepatitis virus, a cytoplasmic RNA virus. *Proc. Natl. Acad. Sci. USA* **81**:3626–3630.
18. Macfarlane, J. T., A. Colville, A. Guion, R. M. Macfarlane, and D. H. Rose. 1993. Prospective study of aetiology and outcome of adult lower-respiratory-tract infections in the community. *Lancet* **341**:511–514.
19. Makino, S., S. A. Stohman, 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.
20. Marra, M. A., S. J. Jones, C. R. Astell, R. A. Holt, A. Brooks-Wilson, Y. S. Butterfield, J. Khattri, 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. Donovon, 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. Kraiden, M. Petric, D. M. Skowronski, C. Upton, and R. L. Roper. 2003. The genome sequence of the SARS-associated coronavirus. *Science* **300**:1399–1404.
21. Nielsen, H., J. Engelbrecht, S. Brunak, and G. von Heijne. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**:1–6.
22. Nussenzweig, V., and R. S. Nussenzweig. 1985. Circumsporozoite proteins of malaria parasites. *Cell* **42**:401–403.
23. Peiris, J. S., C. M. Chu, V. C. Cheng, K. S. Chan, I. F. Hung, L. L. Poon, K. I. Law, B. S. Tang, T. Y. Hon, C. S. Chan, K. H. Chan, J. S. Ng, B. J. Zheng, W. L. Ng, R. W. Lai, Y. Guan, K. Y. Yuen, and the HKU/UCH SARS Study Group. 2003. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* **361**:1767–1772.
24. Peiris, J. S., K. Y. Yuen, A. D. Osterhaus, and K. Stohr. 2003. The severe acute respiratory syndrome. *N. Engl. J. Med.* **349**:2431–2441.
25. 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, K. Y. Yuen, and the SARS Study Group. 2003. Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* **361**:1319–1325.
26. Popova, R., and X. Zhang. 2002. The spike but not the hemagglutinin/esterase protein of bovine coronavirus is necessary and sufficient for viral infection. *Virology* **294**:222–236.
27. Riski, H., and T. Hovi. 1980. Coronavirus infections of man associated with diseases other than the common cold. *J. Med. Virol.* **6**:259–265.
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. Ruiz, M., S. Ewig, M. A. Marcos, J. A. Martinez, F. Arancibia, J. Mensa, and A. Torres. 1990. Etiology of community-acquired pneumonia: impact of age, comorbidity, and severity. *Am. J. Respir. Crit. Care Med.* **160**:397–405.
30. Schwarz, B., E. Routledge, and S. G. Siddell. 1990. Murine coronavirus nonstructural protein ns2 is not essential for virus replication in transformed cells. *J. Virol.* **64**:4784–4791.
31. Snijder, E. J., P. J. Bredenbeek, J. C. Dobbe, V. Thiel, J. Ziebuhr, L. L. Poon, Y. Guan, M. Rozanov, W. J. Spaan, and A. E. Gorbalenya. 2003. Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J. Mol. Biol.* **331**:991–1004.
32. Sonnhammer, E. L., G. von Heijne, and A. Krogh. 1998. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**:175–182.
33. Suzuki, H., and F. Taguchi. 1996. Analysis of the receptor-binding site of murine coronavirus spike protein. *J. Virol.* **70**:2632–2636.
34. 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.
35. 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.
36. van den Hoogen, B. G., J. C. de Jong, J. Groen, T. Kuiken, R. de Groot, R. A. Fouchier, and A. D. Osterhaus. 2001. A newly discovered human pneumovirus isolated from young children with respiratory tract disease. *Nat. Med.* **7**:719–724.
37. 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.
38. Vlasak, R., W. Luytjes, J. Leider, W. Spaan, and P. Palese. 1988. The E3 protein of bovine coronavirus is a receptor-destroying enzyme with acetyltransferase activity. *J. Virol.* **62**:4686–4690.
39. Vlasak, R., W. Luytjes, W. Spaan, and P. Palese. 1988. Human and bovine coronaviruses recognize sialic acid-containing receptors similar to those of influenza C viruses. *Proc. Natl. Acad. Sci. USA* **85**:4526–4529.
40. Wenzel, R. P., J. O. Hendley, J. A. Davies, and J. M. Gwaltney, Jr. 1974. Coronavirus infections in military recruits. Three-year study with coronavirus strains OC43 and 229E. *Am. Rev. Respir. Dis.* **109**:621–624.
41. Wessner, D. R., P. C. Shick, J. H. Lu, C. B. Cardellicchio, S. E. Gagneten, N. Beauchemin, K. V. Holmes, and G. S. Dveksler. 1998. Mutational analysis of the virus and monoclonal antibody binding sites in MHVR, the cellular receptor of the murine coronavirus mouse hepatitis virus strain A59. *J. Virol.* **72**:1941–1948.
42. Williams, G. D., R. Y. Chang, and D. A. Brian. 1999. A phylogenetically conserved hairpin-type 3' untranslated region pseudoknot functions in coronavirus RNA replication. *J. Virol.* **73**:8349–8355.
43. Williams, R. K., G. S. Jiang, and K. V. Holmes. 1991. Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins. *Proc. Natl. Acad. Sci. USA* **88**:5533–5536.
44. 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.
45. 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.
46. Woo, P. C., S. S. Chiu, W. H. Seto, and M. Peiris. 1997. Cost-effectiveness of rapid diagnosis of viral respiratory tract infections in pediatric patients. *J. Clin. Microbiol.* **35**:1579–1581.
47. Yang, H. J., S. J. Park, K. I. Im, and T. S. Yong. 2000. Identification of a *Clonorchis sinensis* gene encoding a vitellaria antigenic protein containing repetitive sequences. *Mol. Biochem. Parasitol.* **111**:213–216.
48. Yeh, S. H., H. Y. Wang, C. Y. Tsai, C. L. Kao, J. Y. Yang, H. W. Liu, I. J. Su, S. F. Tsai, D. S. Chen, P. J. Chen, and the National Taiwan University SARS Research Team. 2004. Characterization of severe acute respiratory syndrome coronavirus genomes in Taiwan: molecular epidemiology and genome evolution. *Proc. Natl. Acad. Sci. USA* **101**:2542–2547.
49. Yokomori, K., L. R. Banner, and M. M. Lai. 1991. Heterogeneity of gene expression of the hemagglutinin-esterase (HE) protein of murine coronaviruses. *Virology* **183**:647–657.
50. Yokomori, K., M. Asanaka, S. A. Stohlman, S. Makino, R. A. Shubin, W. Gilmore, L. P. Weiner, F. I. Wang, and M. M. Lai. 1995. Neuropathogenicity of mouse hepatitis virus JHM isolates differing in hemagglutinin-esterase protein expression. *J. Neurovirol.* **1**:330–339.
51. Zhong, N. S., B. J. Zheng, Y. M. Li, L. L. Poon, Z. H. Xie, K. H. Chan, P. H. Li, S. Y. Tan, Q. Chang, J. P. Xie, X. Q. Liu, J. Xu, D. X. Li, K. Y. Yuen, J. S. Peiris, and Y. Guan. 2003. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People's Republic of China, in February, 2003. *Lancet* **362**:1353–1358.