JOURNAL OF VIROLOGY, Aug. 2004, p. 8824–8834 Vol. 78, No. 16
0022-538X/04/$08.00+0 DOI: 10.1128/JVI.78.16.8824–8834.2004

Human Respiratory Coronavirus OC43: Genetic Stability and Neuroinvasion

Julien R. St-Jean,[1] Hélène Jacomy,[1] Marc Desforges,[1] Astrid Vabret,[2] François Freymuth,[2] and Pierre J. Talbot[1]*

Laboratory of Neuroimmunovirology, INRS-Institut Armand-Frappier, Laval, H7V 1B7 Quebec, Canada,[1] *and Laboratoire de Virologie Humaine et Moléculaire, Centre Hospitalier Régional et Universitaire de Caen, 14033 Caen, France*[2]

Received 28 January 2004/Accepted 7 April 2004

The complete genome sequences of the human coronavirus OC43 (HCoV-OC43) laboratory strain from the American Type Culture Collection (ATCC), and a HCoV-OC43 clinical isolate, designated Paris, were obtained. Both genomes are 30,713 nucleotides long, excluding the poly(A) tail, and only differ by 6 nucleotides. These six mutations are scattered throughout the genome and give rise to only two amino acid substitutions: one in the spike protein gene (I958F) and the other in the nucleocapsid protein gene (V81A). Furthermore, the two variants were shown to reach the central nervous system (CNS) after intranasal inoculation in BALB/c mice, demonstrating neuroinvasive properties. Even though the ATCC strain could penetrate the CNS more effectively than the Paris 2001 isolate, these results suggest that intrinsic neuroinvasive properties already existed for the HCoV-OC43 ATCC human respiratory isolate from the 1960s before it was propagated in newborn mouse brains. It also demonstrates that the molecular structure of HCoV-OC43 is very stable in the environment (the two variants were isolated ca. 40 years apart) despite virus shedding and chances of persistence in the host. The genomes of the two HCoV-OC43 variants display 71, 53.1, and 51.2% identity with those of mouse hepatitis virus A59, severe acute respiratory syndrome human coronavirus Tor2 strain (SARS-HCoV Tor2), and human coronavirus 229E (HCoV-229E), respectively. HCoV-OC43 also possesses well-conserved motifs with regard to the genome sequence of the SARS-HCoV Tor2, especially in open reading frame 1b. These results suggest that HCoV-OC43 and SARS-HCoV may share several important functional properties and that HCoV-OC43 may be used as a model to study the biology of SARS-HCoV without the need for level three biological facilities.

Human coronaviruses (HCoVs), members of the *Coronaviridae* family, are ubiquitous in the environment and are responsible for up to one-third of common colds (41). In the past few years, we have provided experimental evidence that this virus possesses neurotropic and neuroinvasive properties: it persists in neural cell cultures (7, 8) and human brains (9). Of the two HCoV serotypes available, HCoV-OC43 was selected for further characterization of persistence in the nervous system because of a more efficient infection of primary neural cell cultures (11), as well as a trend toward association with neurological disease (9).

Coronaviruses are enveloped viruses that possess a positive-strand RNA genome of up to 31 kb, which represents the largest known genome among all RNA viruses (35). This genome comprises several genes encoding several structural and nonstructural proteins. Among these proteins, the S protein is biologically very important because it could be implicated in determination of tropism (3) and its modulation (50). Indeed, the S protein could be associated with the capacity of the virus to reach the central nervous system (CNS) and possibly trigger neurological disorders (9, 22). It could also be responsible for conferring the strong degree of host species specificity observed with coronaviruses (28).

* Corresponding author. Mailing address: Laboratory of Neuroimmunovirology, INRS-Institut Armand-Frappier, 531 Boulevard des Prairies, Laval, H7V 1B7 Quebec, Canada. Phone: (450) 686-5515. Fax: (450) 686-5566. E-mail: pierre.talbot@inrs-iaf.uquebec.ca.

Only the 3′ one-third of the HCoV-OC43 genome has been sequenced over the years. Therefore, until now, the complete sequence of the open reading frame 1a (ORF1a) and ORF1b, known as the replicase gene, was still undetermined. This gene is essential for coronavirus survival because it contains several motifs, which could be involved in various important viral functions such as transcription, replication, and pathogenesis (66). The products encoded by these two ORFs are polyprotein precursors, which are processed by two or three different proteinases encoded by ORF1a. These proteinases could include two papain-like proteases (PLP1 and PLP2) and a poliovirus 3C-like protease (3CLpro), which presents the most important cleavage activity. The 3CLpro essential function is reflected by its capacity to cleave at many sites in the replicase polyproteins and to release the key replicative functions, such as the RNA-dependent RNA polymerase (RdRp) and the RNA helicase (67).

The HCoV-OC43 strain belongs to the second genetic group, just as SARS-HCoV apparently does (51). The latter is responsible for the severe acute respiratory syndrome (SARS), which is a life-threatening form of pneumonia (46). Since the outbreak of SARS in the fall of 2002 (60), a lot of work has been done to sequence the entire genome of the virus (34) and to understand the mechanisms underlying virus pathogenesis. As presented here, the whole genome of HCoV-OC43 has now been sequenced and, since this human strain is the most related to SARS-HCoV, it could be used as a model for the study of the SARS-HCoV without the drawbacks of level three bio-

logical confinement. Comparisons with the SARS-HCoV nucleotide and amino acid sequences (34) revealed that the two viruses share extensive homology in some important motifs involved in viral replication and pathogenesis. Indeed, the most significant homology between the genomes of the HCoV-OC43 strain and the one of the SARS-HCoV Tor2 isolate is found in the ORF1b region, which comprises the RdRp and helicase motifs (16). The 3CLpro motif of HCoV-OC43 also displays an important level of identity with the one of SARS-HCoV. This finding is noteworthy since SARS-HCoV 3CLpro thus far represents the most promising target for SARS therapy (58).

We report here the complete genome sequences of the HCoV-OC43 strain from the American Type Culture Collection (ATCC), as well as an HCoV-OC43 respiratory clinical isolate, designated HCoV-OC43 Paris. Both genomes are 30,713 nucleotides (nt) long, share the same genomic organization, and only differ by 6 nt. Differences found in the genome of the HCoV-OC43 Paris isolate, compared to the genome of HCoV-OC43 ATCC, give rise to only two amino acid substitutions, which are located in the S (I958F) and the N (V81A) protein genes. After intranasal inoculation in BALB/c mice, the HCoV-OC43 ATCC strain, as well as the Paris isolate, reached the CNS, where they replicated and disseminated, although mice were apparently more easily infected with the ATCC strain than with the Paris isolate. These results suggest that both viruses possess the ability to reach and infect neural cells in vivo. The fact that a natural OC43 isolate has an intrinsic capacity to invade and replicate within the mouse CNS also suggests that the HCoV-OC43 ATCC strain has not acquired its neuroinvasive properties after propagation in newborn mouse brains. Bioinformatics analyses were also performed on the HCoV-OC43 genome. These analysis showed that this virus strain is closely related to mouse hepatitis virus A59 (MHV-A59) and that it displays significant identity levels with important functional domains of the SARS-HCoV. These data provide evidence that HCoV-OC43 could be used as a model for the study of other group 2 coronaviruses, including SARS-HCoV, and that it will facilitate understanding of the biology of this emerging viral strain.

MATERIALS AND METHODS

Viruses and cell lines. The ATCC HCoV-OC43 strain (ATCC number VR-759), isolated in the 1960s, and the Paris clinical respiratory isolate, isolated in March 2001, were grown on a HRT-18 cell line (human adenocarcinoma rectal) as described previously (37). The clinical sample (HCoV-OC43 Paris) was isolated from the respiratory tract of a 68-year-old immunocompromised male who was not related whatsoever to laboratory work and was not in contact with any laboratory workers who had manipulated the HCoV-OC43 ATCC virus. A reverse transcription-PCR (RT-PCR) was performed to specifically detect the presence of the HCoV-OC43 RNA, and an aliquot of the clinical sample was then used to infect the HRT-18 cell line. The HCoV-OC43 ATCC strain and the Paris isolate were never cultured at the same time, and stringent laboratory precautions were used in order to eliminate possible cross-contamination.

Acute infections of cells. Cells were infected at a multiplicity of infection of 0.02 and 0.2 for the ATCC strain and Paris isolate, respectively. The fifth passage of the ATCC strain and the eighth passage of the Paris isolate were used to perform the infections. Cell lines at 70% confluence were infected with the appropriate virus stock in the presence of TPCK (tolylsulfonyl phenylalanyl chloromethyl ketone)-treated trypsin (10 U/ml; Sigma-Aldrich Canada, Ltd.) and 1% (vol/vol) heat-inactivated fetal calf serum and then incubated at 33°C for 4 days in a 5% (vol/vol) CO_2 humid atmosphere.

Mice and inoculations. In order to determine the susceptibility of mice to an infection by HCoV-OC43 ATCC and HCoV-OC43 Paris variants, MHV-seronegative 14-day-postnatal BALB/c mice (Charles River Laboratories, St-Constant, Quebec, Canada) were inoculated intranasally with 5 μl of a virus stock solution containing 10^6 50% tissue culture infective dose(s) ($TCID_{50}$)/ml. Five mice, inoculated with HCoV-OC43 ATCC or HCoV-OC43 Paris variants, were sacrificed every 2 days postinfection (dpi) and processed for detection of infectious virus particles. Every 2 days, two mice infected by HCoV-OC43 ATCC were processed for immunohistochemical detection of viral antigens.

Immunohistochemistry. Mice were perfused by intraventricular injection of 4% (vol/vol) paraformaldehyde, under deep ketamine-xylazine anesthesia, as previously described (22). Brains were dissected and sectioned at a thickness of 40 μm with a Lancer Vibratome. Sections were collected in 0.05 M Tris-buffered saline and then incubated for 2 h at 37°C in a 1/1,000 dilution of an ascites fluid from mouse MAb 1-10C.3, directed against the spike protein of HCoV-OC43 (7). Sections were then rinsed and processed with a Vectastain ABC kit (Vector Laboratories, Burlingame, Calif.). Labeling was revealed with 0.03% (wt/vol) DAB solution (Sigma) and 0.01% (vol/vol) H_2O_2, which yielded a dark brown product.

Infectious virus assays. Brain and lung were dissected, homogenized in 10% (wt/vol) sterile phosphate-buffered saline (PBS), and centrifuged at 4°C for 20 min at 1,000 × *g*, and then supernatants were immediately frozen at −80°C and stored until assayed. The extracts were processed for the presence and quantification of infectious virus by an indirect immunoperoxidase assay, as previously described (22). Briefly, HCoV-OC43-susceptible HRT-18 cells were inoculated with serial logarithmic dilutions of each tissue sample. After 4 days of incubation at 33°C in a 5% (vol/vol) CO_2 humid atmosphere, the cells were washed in PBS and fixed with 0.3% (vol/vol) hydrogen peroxide (H_2O_2) in methanol. After being washed with PBS, they were incubated for 2 h at 37°C in a 1/1,000 dilution of an ascites fluid from mouse MAb 1-10C.3. Afterward, cells were washed in PBS, and horseradish peroxidase-goat anti-mouse immunoglobulins (Dako; Diagnostics Canada, Inc., Mississauga, Ontario, Canada) were added, followed by incubation for 2 h at 37°C. Antibody complexes were detected by incubation in DAB (Sigma) with 0.01% (vol/vol) H_2O_2.

RNA extraction, RT, and PCR. After infection, the cells were washed with PBS, and the total RNA was extracted from the cells by using the GenElute Mammalian Total RNA miniprep kit (Sigma-Aldrich) as recommended by the manufacturer. The RNA was then quantified, and 3 μg was directly used for RT with Moloney murine leukemia virus reverse transcriptase (Invitrogen). For each RT, 500 ng of oligo(dT) primer and 0.5 mM deoxynucleoside triphosphates (Amersham Biosciences) were used, and the reactions lasted between 50 and 60 min at 37°C. Then, 2 μl of the RT cDNA was then used to perform the PCR amplifications. The Expand High-Fidelity *Taq* polymerase (Roche) was used to amplify the HCoV-OC43 genome in six segments, in combination with primers listed in Table 1. All amplifications were performed by using the Cetus DNA thermal cycler (Perkin-Elmer/Applied Biosystems), and an appropriate annealing temperature was used for each specific reaction. Except for the PCR JUB3-12, which required a higher annealing temperature of 65°C, all other annealing temperature used corresponded to the melting temperature of the primers. For each PCR amplification, at least six reactions were performed, pooled together, migrated on a 0.8% (wt/vol) agarose gel (SeaKem), and gel extracted by using the Qiaex II gel extraction kit (Qiagen) prior to sequencing.

RACE and cloning. Rapid amplification of cDNA ends (RACE), cloning, and sequencing were performed for both 5′ and 3′ ends of HCoV-OC43 ATCC strain and the HCoV-OC43 Paris isolate. Primers from the kit used for the RACE are listed in Table 1. An RT reaction of the 5′ end was performed by using the GeneRacer kit (Invitrogen) as recommended by the manufacturer, whereas RT of the 3′ end was performed only by using the GeneRacer oligo(dT) primer provided in the kit (Table 1). In order to amplify both ends, primers from the kit were used in combination with primers specific for the HCoV-OC43 genome. Therefore, the GeneRacer 5′ nested primer was used with JUB2 primer, and GeneRacer 3′ nested primer was combined with JUMO1 primer for the ATCC strain and JUO8 primer for the Paris isolate. Amplicons of the 5′ ends of both viruses and of the 3′ end of the Paris isolate were cloned by using the Zero Blunt TOPO PCR cloning kit for sequencing (Invitrogen), whereas amplicons of the 3′ end of the ATCC strain were cloned by using the TOPO XL PCR cloning kit. The RACE 5′ clones were sequenced by using M13 universal forward and reverse primers and RACEJUB1 and RACEJUB2 primers, and RACE 3′ clones were sequenced by using M13 universal forward and reverse primers and JUO7 primer.

Sequencing. Sequencing reactions were performed by Bio S&T (Montreal, Quebec, Canada) by using the dideoxy method (Sanger) and specific primers, which are listed in Table 2. As described above, PCR products were directly

TABLE 1. Primers used for amplification of the HCoV-OC43 genome

Primer combination and (nt location)	Target region or sequence	Amplicon length (bp)
JUB3–JUB12 (1–20 and 6071–6091)	Leader, 5′UTR, and ORF1a	6,091
JUB5–JUB6 (5319–5339 and 11111–11131)	ORF1a	5,813
JUB7–JUB8 (10901–10921 and 16525–16545)	ORF1a and ORF1b	5,645
JUB9–JUB10 (16309–16329 and 21544–21564)	ORF1b and ns2	5,256
JUNSO1-JUSO2 (21330–21350 and 27754–27774)	ORF1b, ns2, HE, and S genes	6,445
JUMO1-GeneRacer, 3′ nested (27649–27669 and 30742–30764)	S; ns12.9; E, M, and N genes; and 3′UTR	3,116[a]
GeneRacer, oligo(dT)	5′-GCTGTCAACGATACGCTACGTAACGGCATGACAGTG$(T)_{18}$-3′	
GeneRacer, 5′ nested	5′-GGACACTGACATGGACTGAAGGAGTA-3′	
GeneRacer, 3′ nested	5′-CGCTACGTAACGGCATGACAGTG-3′	

[a] This value assumes a poly(A) tail of 28 bp.

sequenced for both genomes, and both strands were sequenced in each case, including RACE clones. For each genome, at least two RACE 5′ and 3′ clones were sequenced for both isolates. Sequences obtained from chromatograms were aligned by using the basic local alignment search tool (BLAST; bl2seq) from the National Center for Biotechnology Information and were analyzed by using the Chromas 2 software.

Bioinformatics analyses. Bioinformatics analyses were performed by Sequence Bioinformatics (Montreal, Quebec, Canada). The BLAST program was used to perform genome versus genome and gene versus genome alignments. RNA folding was analyzed by using MFOLD. PHYLIP was used for phylogenic tree construction. The FASTA-formatted sequences of the complete genomes were aligned with CLUSTAL W (v1.82) by using the default parameters for DNA alignments. The PHYLIP output option of CLUSTAL W was used to produce a multiple alignment file that was used as input for dnaml (v3.6), which produced an unrooted maximum-likelihood phylogenic tree with the default parameters. ORF analysis was performed by using tools from the EMBOSS suite. In the case of SARS-HCoV and HCoV-OC43 ATCC, the extracted ORFs were submitted to HMMPFAM, of the HMMER suite, for motif detection against the PFAM database. The amino acid sequences of the known expressed proteins were also submitted to HMMPFAM and the patmatmotif tool of EMBOSS. This tool performs motif scanning against the PROSITE motif database.

Nucleotide sequence accession number. The GenBank sequence accession numbers for the complete genome of the HCoV-OC43 ATCC strain and the Paris isolate are, respectively AY585228 and AY585229.

RESULTS

Amplification and sequencing of HCoV-OC43 ATCC and Paris genomes. The genomes of the HCoV-OC43 ATCC strain and of the Paris isolate were amplified in six fragments by RT-PCR in order to be sequenced (Fig. 1 and Table 1). The PCR products encompassed the entire genome of the viruses and overlapped each other to make sure the final sequences were complete. Primers used for the amplifications of ORF 1a and ORF1b were created by using the sequence of the bovine coronavirus Quebec strain (BCoV Quebec) (63), which displays 97% identity in this region of the genome and was known to share 92% identity with the 3′ 9 kb of the HCoV-OC43 genome (24, 26, 29, 37, 38, 39). Primers used to amplify the 3′ region were designed on the basis of sequences of HCoV-OC43 available in GenBank. The gene-walking approach was used to sequence the whole genome of the HCoV-OC43 ATCC strain, whereas the Paris isolate was sequenced by using the primers generated during the sequencing of the ATCC strain (Table 2).

Main features of the HCoV-OC43 genome. The genomes of the two variants contain 30,713 nt, excluding the poly(A) tail, and include nine main ORFs flanked by 5′ (nt 1 to 209)- and 3′ (nt 30426 to 30713)-untranslated regions (UTRs) (Fig. 1 and Table 3). The genome of HCoV-OC43 contains multiple secondary ORFs, scattered throughout the genome in all frames and in both orientations (data not shown). By using ShowORF software, it was possible to determine that the translation complex could potentially use several of these ORFs in the 5′→3′ orientation with, for instance, a translation reinitiation mechanism (20, 36). By using better-characterized coronaviruses, putative transcription-regulating sequences (TRSs) of HCoV-OC43 were also identified (4, 44, 58) (Table 3). These sequences are found at the 5′ end of each viral RNA, genomic or subgenomic, and represent signals for the discontinuous transcription of subgenomic mRNA (49). The identified canonical core sequence for HCoV-OC43 was 5′-UCUAAAC-3′, but it was not always perfectly conserved throughout the genome (Table 3).

Using bioinformatics tools and well-characterized coronaviruses (15, 18, 67), it was also possible to draw a precise map of the main domains contained within the polyprotein 1ab and to determine the location of the putative viral proteolytic cleavage sites (Fig. 2). Most of the main motifs found throughout the genome after bioinformatics analysis corresponded to expected motifs found in other coronaviruses. Indeed, the PLP1 and PLP2 motifs, the membrane-spanning domains (TM), and the 3CLpro motif as well as the RdRp and the RNA helicase motifs, were found in ORF1ab at the expected positions compared to other coronaviruses. Since HCoV-OC43 and BCoV possess a high degree of identity, the positions of the cleavage sites were determined with the BCoV model (15), and 14 sites were identified in the polyprotein 1ab. Among these cleavage sites, three are recognized by PLP1 or PLP2, and the 11 others are recognized by 3CLpro, generating mature products containing key motifs for viral transcription and replication. A putative ribosomal −1 frameshift was also identified upstream

TABLE 2. Primers used for sequencing of the HCoV-OC43 genome[a]

Positive-strand primer	Sequence (5′→3′)	Location (nt)	Negative-strand primer	Sequence (5′→3′)	Location (nt)
JUB3	GATTGTGAGCGATTTGCGTGC	1–20	JUB12	ACATCACCTGTAGCTGTTGGC	6071–6091
RACEJUB2	TGTGATGGTGGATTGTCGCCG	380–400	JUB11	CCAGTAACGTCTGTAACCTTC	57505770
JUB31	GTTATATATGATTGATCCTGC	662–682	JUB111	CCATGCCTCTTGCCATTGAAC	5277–5297
JUB32	TGATTATACTGGTAGTCTTGC	983–1003	JUB112	TGAATTTAACACCATCAACAG	4918–4938
JUB33	GCACAATCTTCAGGTGTTTTG	1488–1508	JUB113	ATCCTCTTGATTATTGCTAAC	4458–4478
JUB14	AGTTGCTAGGTGTGTCAGATG	1801–1821	JUB115	AGGGTTTACAACAACTTCTGC	4086–4106
JUB141	CTTATATAGTAGTGGAGAGTG	2254–2274	JUB15	AACAGCCATAGAATGACTATC	3876–3896
JUB142	GTTCTGATTTTTCATTAGCGG	2557–2577	JUB151	GCTTTAGGCACATACAGACCC	3395–3415
JUB143	ACAGCTCCTGAAGATGATGAC	3093–3113	JUB152	ACCACAGCATAAAATTCCTCC	2915–2935
JUB144	TTAACCTTATGTGATTGGCAG	3621–3641	JUB153	CAAATTCTTCTACGTCCAATC	2399–2419
JUB161	ATGCTATGTTCTTTTATGGTG	3700–3720	JUB13	AATGCTTGACCACTTACTGCC	1991–2011
JUB163	TATTATTGGGCATGGTATGTC	3959–3979	JUB131	CAAGCAAAACCATCTATCATG	1427–1447
JUB164	AAGTTATGTATTGTTAGAGCG	4310–4330	JUB132	CTCCAAGTAGGAAATAATGCC	1049–1069
JUB165	TATTTTGAGTGTACTGGAGGC	4782–4802	RACE JUB1	GGAGCAAATCATATCCACCTC	312–332
JUB166	TGCTTGCCTATTACAACATGC	5128–5148	JUB6	CCTCTAAATGTCTGCTTGTAC	11111–11131
JUB5	CCTGCTAGATTTGTATCGTTG	5319–5339	JUB61	ATAGCAGCCAACAGTGTTTCC	10724–10744
JUB51	ACTCAGCGTATTATTAAAGCC	5913–5933	JUB62	AGGGTCACTGTAAGAACAAGC	10205–10225
JUB52	GTTGACGATGGTGGTGACAGC	6264–6284	JUB63	CTATTAAAAGCAACATCAGAC	9752–9772
JUB53	AATCTACACCACAGAAATTGC	6701–6721	JUB64	AGAAAGTATGGGGTAAACTTG	9453–9473
JUB54	TGGCAGGATTTGATATGTTAG	7036–7056	JUB18	ATCTCTACCACAAAAGGTCCC	9213–9233
JUB17	GGTTTTACCCATTGTTTGCTC	7177–7197	JUB181	TGGTAGGGACATTAAACACTG	8761–8781
JUB171	GCTTGTTCTATGATCGTGATG	7678–7698	JUB182	GCTTTCTTCAATTTATGCTGG	8345–8365
JUB172	CTGTGCTCGTAAAAGTTGTTC	8072–8092	JUB183	CATAGACAAAAATGTATCCAC	7950–7970
JUB173	AATAAGCAGATGGCTAATGTC	8403–8423	JUB184	AAGTATTACCTGGTTTATAAG	7543–7563
JUB55	AAGGTTTTATCCGTCTTCCAG	9043–9063	JUB65	ACCTAACACTCCAATGTAATG	7252–7272
JUB56	CACTTACAATGGCTAGTTATG	9540–9560	JUB66	CTAAAAGTGTTCTTAATCCAC	6926–6946
JUB57	CCGTCTCAACTTCATTCTTGC	9925–9945	JUB67	TATACTAACAAGAGTCATACC	6531–6551
JUB58	CTTGTGGATCTGTTGGTTATG	10378–10398	JUB68	ACATCACCTGTAGCTGTTGGC	6071–6091
JUB7	GGCTTCTACATTTTTGTTTAG	10901–10921	JUB8	TAACTCTGCTTAAAGGCTTCC	16525–16545
JUB71	TGTATTTCACAGATATACCTC	11443–11463	JUB81	GATGTTTGAGAAGAGCAGACC	16117–16137
JUB72	CAGCAGTTAAAACAGCTAGAG	12084–12104	JUB82	CATAATATCCTGTGACAAAGG	15522–15542
JUB73	ATAATGAGGTATCTGCTACTG	12547–12567	JUB83	TATAACTACAGGAACACCACG	15050–15070
JUB74	TGTTAAACCCGATGCTACCAC	13073–13093	JUB84	AAAAGTGCTTCAGATCAACTG	14604–14624
JUB75	AGAGAGATGAAATGCTATGAG	13559–13579	JUB85	ATACTCCAGTGCTTAAAATAC	14158–14178
JUB76	TTGGATTGCGAATTGTATGTG	14066–14086	JUB86	TCCTTGCGTACAATGTGTGGC	13648–13668
JUB77	TTGTATGATTTACGCACTTGC	14465–14485	JUB87	TCCACAAACTTGACAAACATC	13245–13265
JUB78	CATTATCATTTGAGGAGCAGG	14856–14876	JUB88	TCTTGCTAGTGTGTTACAACC	12843–12863
JUB79	GAGGCATGTTGTTCGCAAAGC	15230–15250	JUB89	CTTGGATAGTTTGAATCTGCC	12448–12468
JUB710	ATATAAGTGCCTTTCAACAGG	15636–15656	JUB810	ACTAGAACACGCCTCATCAAG	12048–12068
JUB711	CAAAAGTTTACTGATGAGTCC	16040–16060	JUB811	CTTAAAATTAAGCATAAGGGC	11649–11669
JUB9	TTATTGTGAAGATCATAAGCC	16309–16329	JUB812	AGAACAAATCATGGTTAATGC	11245–11265
JUB91	TCAACACATTGGTATGAAACG	16903–16923	JUB10	AAATGGGTAAGTGGAAAATTG	21544–21564
JUB92	AGTTCATTGTGTTTTAGGGTC	17498–17518	JUB101	AGAGCTAACTTGTCTCGAATC	21064–21084
JUB93	TATTGGTGATTCTGCTGTTAC	18034–18054	JUB102	GCAAACACTCTTACTACCACC	20414–20434
JUB94	TAGAACTGGTTACTATGGTTG	18562–18582	JUB103	GCGGTGGACCTCTTATCATCG	19908–19928
JUB95	TTTTGAGGCACATAAGGACTC	18997–19017	JUB104	GTATTGTAAGACTCTAAGTAC	19369–19389
JUB96	AGTCAAGACTGGTCATTATAC	19498–19518	JUB105	TACAAAAGAGTCTTAACAGAC	18973–18993
JUB97	CGTTCTAATAATGGCGTTTAC	19859–19879	JUB106	CCCATGTAACTAGCACAACAC	18444–18464
JUB98	TAGGCTTGTACCGAAGACAGC	20313–20333	JUB107	TTATATTTGTCATCTACTGCC	17983–18003
JUB99	ATACTCAGTTATGTCAATATC	20736–20756	JUB108	TTATGCCACAAAGGGTTAGCC	17599–17619
JUB100	TCGAGACAAGTTAGCTCTGGG	21067–21087	JUB109	GGTAGTAAACACATACTTACG	17153–17173
JUNSO3	TTAATGATATGGTTTATTCCC	21396–21416	JUB110	CTGCGGAACAAGCGTAGGAGC	16808–16828
JUNSO6	GTGTAGAAGAATTGCATGACG	21807–21827	JUSO2	AAGAACTTTAACAAATGCTAG	27754–27774
JUHEO1	AACAATTCTTGGTTCTTCC	22226–22244	JUSO4	ATCAAGTGACAAATCTGGTGC	27384–27404
JUHEO3	TTTAGGAGTTTTCACTTTACC	22588–22608	JUSO6	ATATGATTACCATTACCACAG	27020–27040
JUHEO5	ACCACTTTGTATTTTTAACGG	22959–22979	JUSO8	TGAATAGCATAAAGGGCATTG	26678–26698
JUHEO7	TCATGGAGATGCTGGTTTTAC	23337–23357	JUSO10	GCTGCCTAGACAACCTAATAC	26307–26327
JUSO24	ATTAATAACCCTGATTTACCC	23482–23502	JUSO12	AATGGCTCAAAATTAGTAAAC	25937–25957
JUSO5	AATGTATAGTGAGTTCCCTGC	23993–24013	JUSO14	TTATAATAAGTCGCATTAACC	25553–25573
JUSO7	CTTTCACACTATTATGTCATG	24378–24398	JUSO16	ACAAGTTAAATAATTAGTACC	25188–25208
JUSO9	TATTCAGGCAGACTCATTTAC	24752–24772	JUSO18	ATAGTTATGCTGGAAAAACAC	24809–24829
JUSO11	AATTGAATGGTTCGTGTGTAG	25123–25143	JUSO20	TAGAAGTGAGAGGTGTAACCC	24436–24456
JUSO13	GTGTTTGTGTTAATTATGACC	25498–25518	JUSO22	TATAGGATGTATTTACAAAAG	24031–24051
JUSO15	TAGGTAGTGGTTACTGTGTGG	25867–25887	JUHEO2	ACATAATAAGTACCCAAACC	23778–23797
JUSO17	GAATGGTGTTACTCTTAGCAC	26234–26254	JUHEO4	CATTATCATACCTAAAAACGC	23408–23428
JUSO19	TGGATGTGCTAAGTCAAAATC	26632–26652	JUHEO6	AGACCATAAATAACACCAGTG	23037–23057
JUSO21	ATTTCTGTGGTAATGGTAATC	27016–27036	JUHEO8	ATGATAAGGCGTAAAATTAAC	22660–22680
JUSO25	TGGCACCAGATTTGTCACTTG	27382–27402	JUNSO2	GAAACAACATTGGTAGGAGGG	22416–22436
JUMO1	GTGATGATTATACTGGATACC	27649–27669	JUNSO4	TTCCTTAATGGACAGTGCTGC	21917–21937
JUMO3	TAGTTGCCATTTGTTTATTGG	28169–28189	JUO2	GCAGCAAGACATCCATTCTG	30495–30514
JUMO5	ATGTGGATTGTGTATTTTGTG	28662–28682	JUO4	TACCAAAACACTGCTGAACAG	29920–29940
JUO8	ATGTCTTTTACTCCTGGTAAGC	29079–29100	JUO6	ATACCATCGTGGCAGCAGTTG	29430–29450
JUO3	TCTACTGGGTCGCTAGTAACC	29512–29532	JUMO2	CCACTTGAGGATGCCATTACC	29136–29156
JUO5	CCCACAGTTCCCCATTCTTGC	29999–30019	JUMO4	TACACAATCCACATAATAATG	28655–28675
JUO7	CTCTCTATCAGAATGGATGTC	30487–30507	JUMO6	TATAAAAATTATTTGCCCCAC	28150–28170

[a] Underlined nucleotides indicate mismatched bases with regard to the genome sequence of the HCoV-OC43 ATCC strain.

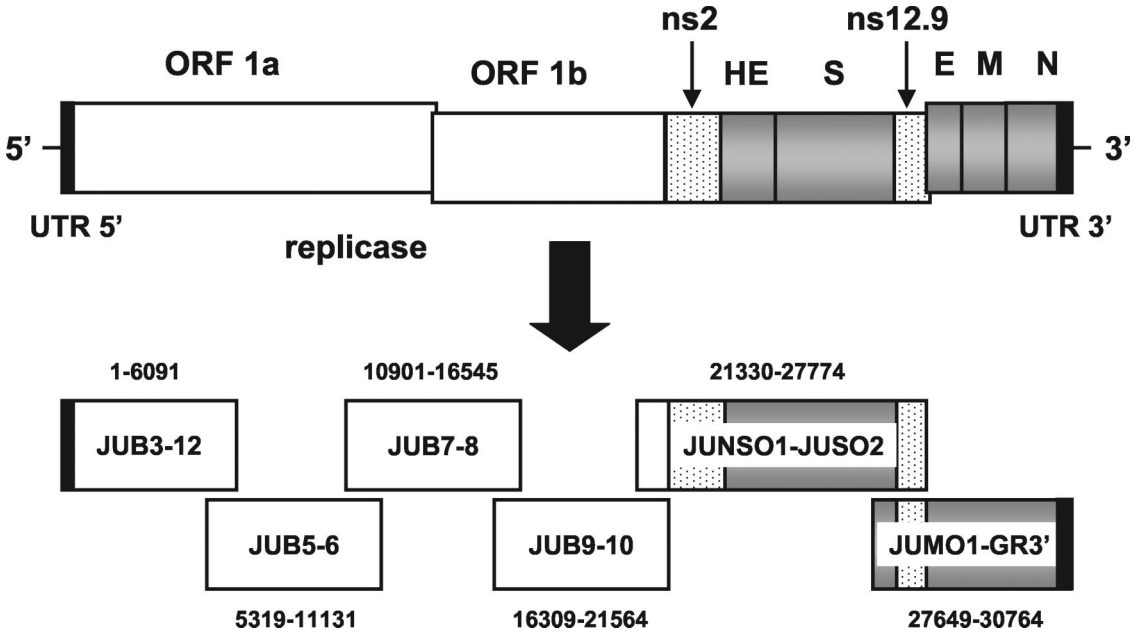

FIG. 1. Schematic representation of the HCoV-OC43 genome and of the amplification strategy used for sequencing. The HCoV-OC43 genome is 30,713 nt long and comprises nine main ORFs: ORF1a, ORF1b, ns2 (the gene encoding the nonstructural protein 2), HE (hemagglutinin-esterase gene), S (spike gene), ns12.9 (the gene encoding a nonstructural protein of 12.9 kDa), E (small envelope gene), M (membrane gene), and N (nucleocapsid gene). The replicase gene includes both ORF1a and ORF1b. The entire genome was amplified in six fragments in order to be sequenced. Each PCR product was named according to the name of the primers used for the amplification, and the location in the genome is indicated above or below each PCR product. Boxes: open, gene encoding the replicase polyprotein; dotted, genes encoding nonstructural proteins; shaded, genes encoding structural proteins; black, UTRs. GR, GeneRacer.

of the intersection of ORF1a and ORF1b. The slippery sequence $_{13334}$UUUAAAC$_{13340}$ was found at the 3′ end of ORF1a and is thought to be involved, in combination with RNA pseudoknot structures, in the frameshift, which would occur at the C_{13340} nt (58).

High degree of identity between the HCoV-OC43 ATCC strain and the Paris isolate. Differences in nucleotides and amino acids between the HCoV-OC43 ATCC strain and the Paris isolate are presented in Table 4. In all, only 6 nt differ between the two variants. These mutations are located in the 5′UTR, in the ns2, S, M, and N genes, and in the 3′UTR. According to the MFOLD software, mutations located in the UTRs would not affect RNA folding (data not shown) and would therefore not have any effect on viral transcription and replication. Mutations located in the ns2 and M genes would not affect virus biology since they do not give rise to any amino

TABLE 3. Organization of the HCoV-OC43 genome

Genome region	Location (nt)	TRS location (nt)	TRS sequence[a]
Leader and 5′UTR	1–209	63–69	**UCUAAAC**. . .139 nt. . .AUG
ORF1a	210–13361		
ORF1b[b]	13361–21496		
Intergenic region	21497–21505		
ns2 gene	21506–22342	21492–21498	**UCUAAAC**UUUAAAAAUG
Intergenic region	22343–22353		
HE gene	22354–23628	22339–22344	**UUAAAC**UCAGUGAAAAUG
Intergenic region	23629–23642		
S gene	23643–27704	23636–23642	**UCUAAAC**AUG
Intergenic region	27705–27791		
ns12.9 gene	27792–28121	27771–27777	**UCUUAAG**GCCACGCCCUAUUAAUG
E gene[c]	28108–28362		
Intergenic region	28363–28376		
M gene	28377–29069	28367–28373	**UCCAAAC**AUUAUG
Intergenic region	29070–29078		
N gene	29079–30425	29065–29070	**UCUAAA**UUUUAAGGAUG
3′UTR	30426–30713		
Poly(A) tail of 28 nt	30714–30741		

[a] Nucleotides in boldface indicate TRS sequences, whereas underlined nucleotides indicate the initiation codon.
[b] Putative ribosomal −1 frameshift between ORF1a and ORF1b.
[c] ORF overlap for the ns12.9 gene and the E gene.

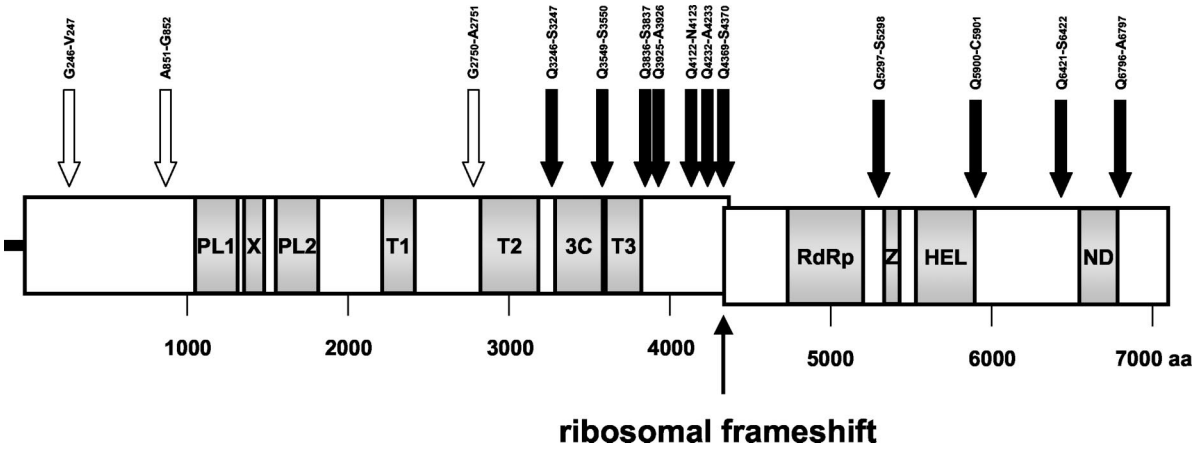

FIG. 2. Schematic representation of the polyprotein 1ab putative proteolytic processing and of the main domains found in ORF1ab. The approximate positions of predicted functional domains and protease cleavage sites in ORF1ab are shown, and amino acids positions are also indicated. The white arrows indicate putative cleavage sites recognized either by the PLP1 or the PLP2, whereas black arrows indicate sites recognized by the main protease, 3CLpro. The 15 putative cleavage products generated by the proteolytic processing are named as follows: leader protein, MHV p65-like protein, nsp1 (PL1, X, PL2, and T1), T2, nsp2 (3CLpro), nsp3 (T3), nsp4, nsp5, nsp6, nsp7, nsp9 (RdRp), nsp10 (HEL), nsp11, nsp12, and nsp13 (15). A putative ribosomal −1 frameshift is indicated between ORF1a and ORF1b. Upstream of the frameshift site, the slippery sequence $_{13334}$UUUAAAC$_{13340}$ is found. PL1 and PL2, accessory protease domains; X, conserved domain of unknown function; T1, T2, and T3, membrane-spanning (hydrophobic) domains; 3C, 3CLpro domain; Z, putative zinc finger; HEL, NTPase RNA helicase domain; ND, domain conserved exclusively in nidoviruses. nsp, Nonstructural protein.

acid substitution. However, two mutations lead to amino acid substitutions. The first is located at nt 26514, in the S2 subunit of the S gene, and gives rise to the I958F (ATCC→Paris) mutation, whereas the second is located at nt 29320, in the N gene, and gives rise to the V81A mutation.

Neuroinvasion in BALB/c mice. After inhalation of virus, mice were processed for histochemical labeling of HCoV-OC43 ATCC antigens. Cells positive for viral antigens were first observed ca. 3 dpi in the olfactory bulb as patches of labeled neurons (Fig. 3A). No cells positive for viral antigens could be seen in other part of the brain, even near perivascular blood cells. At 7 dpi, viral antigens were detected in all brain regions, indicating a rapid dissemination throughout the CNS (Fig. 3B). Five mice of each group, infected with HCoV-OC43 ATCC and Paris, were sacrificed every 48 h, and virus titers were measured in the CNS and lung. Even though mice inhaled a viral suspension, virus was rarely found in the lung (limit of detection was $10^{1.5}$ $TCID_{50}$/g due to a lung extract toxicity on HRT-18 cells) and only when brain titers reached at least 10^4 $TCID_{50}$/g (data not shown). HCoV-OC43 ATCC infectious virus could be detected in mouse CNS, as early as 2 dpi. Virus titers were maximal ca. 4 dpi and remained high throughout the experiments (Fig. 3C). When virus reached the brain, replication of HCoV-OC43 ATCC led to a fatal encephalitis. Infectious HCoV-OC43 Paris could only be detected in mice starting at 6 dpi (Fig. 3C), and a lower number of mice were productively infected by HCoV-OC43 Paris than by HCoV-OC43 ATCC. Nevertheless, when infectious virus reached the brain, infectious virus titers were comparable for the two HCoV-OC43 variants, suggesting that the ATCC and Paris variants both exhibit neuroinvasive and neurotropic properties.

TABLE 4. Sequence differences between the reference strain HCoV-OC43 ATCC and the Paris isolate

Mutation location (nt)	Region of mutation	Consequence of mutation (ATCC→Paris)
31	5′UTR	C→T
22243	ns2 gene	TTC (Phe246)→TTT (Phe246) (no amino acid change)
26514	S gene	ATC (Ile958)→TTC (Phe958)
28808	M gene	ACT (Thr144)→ACC (Thr144) (no amino acid change)
29320	N gene	GTA (Val81)→GCA (Ala81)
30632	3′UTR	C→A

Comparison of HCoV-OC43 with other coronaviruses. HCoV-OC43 is part of the second genetic group of coronaviruses (30) and displays higher identity levels with virus strains that belong to this group, including SARS-HCoV. The coronavirus strains that present the highest degree of identity with HCoV-OC43 are BCoV and MHV-A59, with 95 and 71% identities, respectively. HCoV-OC43 and BCoV are very related at the nucleotide level, and most of the differences between the two genomes are found in the S1 subunit of the S gene, suggesting that the two virus strains possess similar biological properties but display a different cellular tropism. SARS-HCoV and HCoV-229E display 53.1 and 51.2% identity with regard to the HCoV-OC43 strain. Although SARS-HCoV is apparently part of group 2 (51), the overall identity level with the OC43 strain is not striking, but the two strains present a very high degree of amino acid identity in some important functional domains, such as the RdRp, the RNA helicase, and 3CLpro.

A phylogenic unrooted tree regrouping seven coronavirus strains from the three genetic groups was obtained by using the complete genome sequences of all strains (Fig. 4). This tree is the first one that includes the complete genome of the HCoV-OC43 strain. It shows that HCoV-OC43 and BCoV are evolutionary very related and that they form a clade with MHV-

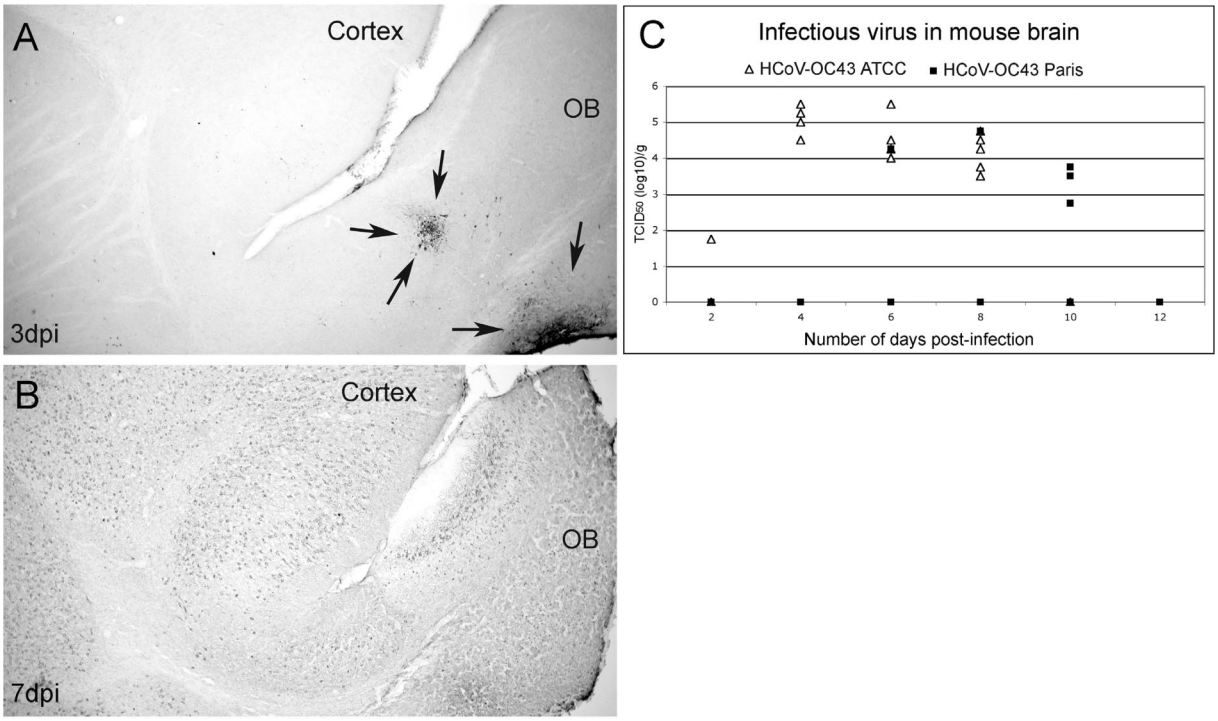

FIG. 3. Neuroinvasive properties of HCoV-OC43 ATCC and HCoV-OC43 Paris variant in BALB/c mice after intranasal inoculation. (A) At 3 dpi, cells positive for viral antigens (arrows) were first observed in the olfactory bulb (OB). No infected cells could be detected in the cortex or other brain structures, illustrating transneuronal spreading of the virus. (B) At 7 dpi, the virus has disseminated to the entire CNS, as illustrated by the presence of immunopositive cells throughout the brain. Magnification (A and B), ×32. (C) Quantification of infectious virus in the brain of each mouse at different times postinfection. Virus titers are presented as logarithmic value of $TCID_{50}$ per gram of tissue (the limit of detection was $10^{0.5}$ $TCID_{50}$/g). Infection by HCoV-OC43 ATCC was detected in one mouse as early as 2 dpi, and gradually more mice became positive. HCoV-OC43 ATCC infectious particles were found between 2 to 8 dpi in mouse brain and led to fatal encephalitis before the end of the experimentation. Antigens of the HCoV-OC43 Paris variant were first revealed in mouse brain at 6 dpi. Infectious particles were detected in some of the brains up to 10 dpi. HCoV-OC43 Paris infectious titers in susceptible animals were similar to those found after HCoV-OC43 ATCC infection, and mice positive for either variant presented all pathological and clinical signs of encephalitis.

A59. Although SARS-HCoV is apparently part of group 2 (51), the analysis shows that it is more divergent from strains of the previous clade and that infectious bronchitis virus (IBV) and SARS-HCoV display the highest divergence among the strains analyzed. Group 1 coronaviruses are also grouped in such a clade.

A BLAST analysis with coronaviruses from all three genetic groups showed that different degrees of identity exist between several regions of different virus strains but that the most conserved region among all coronaviruses is located within ORF1b (data not shown). More stringent BLAST analysis was carried out on the genome sequences of HCoV-229E, MHV-A59, and SARS-HCoV with the HCoV-OC43 genome as a reference (data not shown). Among all genes analyzed, the most significant identity levels were found in ORF1ab, as well as in the S2 subunit of the S gene. Significant identity levels were observed with MHV-A59 and SARS-HCoV. With regards to the ORF1ab, the identity levels were usually lower in ORF1a than in ORF1b and were more significant in the case of MHV-A59 than for SARS-HCoV. Moreover, identity was more significant at the amino acid level. Several domains which are essential for viral replication, such as the 3CLpro, RdRp, and helicase domains, are very interesting because of their functional importance. The S2 subunit of the S gene also dis-

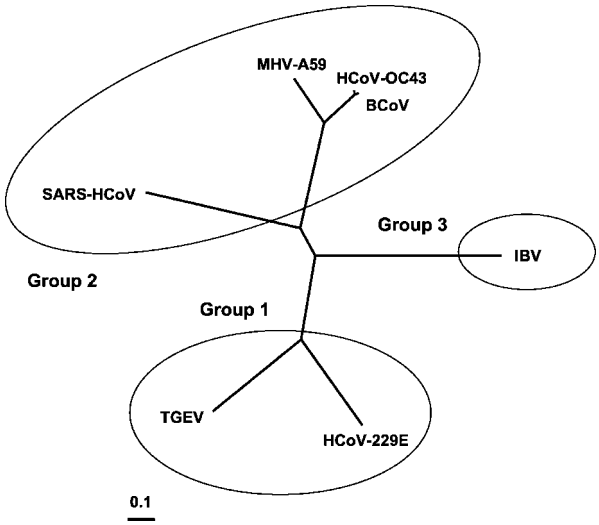

FIG. 4. Phylogenic unrooted tree regrouping seven coronavirus complete genomes from the three genetic groups. Circles regroup members of each three genetic groups. The 0.1 sliding bar represents the genetic distance between the species (i.e., nucleotide substitution units per studied site). Strains: MHV-A59 (NC_001846); BCoV, bovine coronavirus Quebec strain (AF220295); SARS-HCoV, SARS-HCoV Tor2 strain (AY274119); IBV, IBV Beaudette strain (NC_001451); TGEV (NC_002306); HCoV-229E (NC_002645).

FIG. 5. Multiple alignments of amino acids of the main proteases of coronaviruses from all three genetic groups. Positions with absolute conservation are shadowed, whereas residues of the putative catalytic dyad, His^{41} and Cys^{145}, are boxed. Conservation level among group 2 coronaviruses was ca. 46.2%, whereas all strains displayed 26% identity. Strains: OC43, HCoV-OC43 (group 2); BCoV, BCoV Quebec group 2; MHV-A59, MHV group 2; SARS Tor2, SARS-HCoV Tor2 group 2; 229E, HCoV-229E group 1; IBV, IBV group 3.

played high identity levels with its counterparts from other coronaviruses. For instance, the MHV-A59 S2 subunit displayed 76% identity and 88% similarity, whereas the S1 subunit presented only 53% identity and 65% similarity. This result is logical since it has been shown that the membrane fusion function resides within the S2 subunit (54, 62), whereas the S1 subunit is involved in receptor binding (56) and determination of tropism (3), which is different from one virus to another.

Since SARS-HCoV is now considered as a serious pathogen that has recently emerged and that we believe HCoV-OC43 could represent an excellent model for the study of this virus, it was of interest to analyze some functionally important motifs that display significant identity levels with the HCoV-OC43 genome. The most striking identities between the two strains were found mainly in ORF1b, albeit the 3CLpro motif, in ORF1a, also presented a significant identity level. The cleavage product containing the 3CLpro motif displayed 48% identity and 64% similarity with the corresponding region of HCoV-OC43. Of the three viral proteases that play a role in the processing of the polyprotein 1ab, 3CLpro is the main protease (67). This domain of the viral genome is essential for replication since it cleaves the HCoV-OC43 polyprotein 1ab at 11 sites and allows the release of important functional domains (Fig. 2) (32). Like other coronavirus 3CLpros, HCoV-OC43 3CLpro acts via a catalytic dyad, which is composed of a His^{41} and a Cys^{145} (6). The HCoV-OC43 3CLpro is 303 amino acids long and displays an outstanding conservation among coronaviruses from the three genetic groups (Fig. 5). Seventy-nine residues are strictly conserved among sequences from six different coronaviruses, displaying 26% identity among all 3CLpro sequences analyzed, whereas group 2 coronaviruses display 46.2% identity for the same motif between each other.

DISCUSSION

HCoV-OC43 belongs to the second genetic group of coronaviruses and represents the HCoV that is most related to SARS-HCoV. Here, we present the first report of a complete sequence of the HCoV-OC43 genome, including the complete sequence of a clinical respiratory isolate of the OC43 serotype. The two genomes are 30,713 nt long and only differ by 6 nt, including two amino acid substitutions located in the S (I958F)- and N (V81A)-protein genes. The genomes of the two virus variants display 71, 53.1, and 51.2% identity with the genomes of MHV-A59, SARS-HCoV Tor2, and HCoV-229E, respectively. Using bioinformatics tools and well-characterized coronaviruses, further characterization of the HCoV-OC43 genome was performed, and these analyses revealed that HCoV-OC43 is closely related to BCoV and MHV and that it displays significant amino acid identity levels with important functional domains of the SARS-HCoV. Like the ATCC strain that was isolated in the 1960s, HCoV-OC43 Paris, isolated in 2001, exhibited neuroinvasive properties in BALB/c mice. Although

mice were more easily infected with the ATCC strain than with the Paris isolate, these results suggest that both viruses possess the intrinsic ability to infect neural cells and to reach the CNS from the periphery.

Recently, L. Vijgen and coworkers have submitted a complete sequence of the HCoV-OC43 genome to GenBank (NC_005147). The virus strain used for this sequencing is described as corresponding to the virus strain that was used in our laboratory (VR-759). However, comparison of our sequence with theirs show that they differed at 33 positions, 29 mutations being located in the S gene, including two mutations in the S2 subunit. Of the four other differences, one is located at the beginning of the genome sequence, where a guanine is added with respect to our sequence, whereas the other three are scattered throughout ORF1a. Despite these differences, the availability of the complete genome sequence from a clinical isolate reinforces the validity of our sequence, since the HCoV-OC43 ATCC and Paris sequences only differ by 6 nt. Therefore, this observation suggests that the viral strain used by Vijgen and collaborators could have been adapted in cell culture, given the differences observed in the S gene, which is known to be associated with viral adaptation (27). No differences were noticed among ORF1b sequences between HCoV-OC43 ATCC, Paris, and the one from Vijgen and coworkers. This observation suggests that this region of the genome needs a high rate of conservation in order to remain functional and that genes located downstream of the replicase gene are more permissive to sequence modifications.

Using a recent HCoV-OC43 clinical respiratory isolate, we showed here that HCoV-OC43 apparently remains genetically stable in the environment. Indeed, despite virus shedding and chances of persistence in the host, the HCoV-OC43 Paris isolate displays differences at only six positions with regard to the ATCC strain sequence, despite about 40 years have elapsed between the two isolations. Since the viral persistence could be associated with molecular adaptation (7, 8), the low rate of mutation observed here could be explained by the fact that the HCoV-OC43 Paris isolate has never or rarely persisted before. However, it is too soon to speculate about such an issue given that the exact origin of the virus before its isolation remains undetermined. It is also worth noting that viral persistence does not necessarily require an adaptation to the environment (2) and that, despite the high rate of mutation of the coronavirus RdRp (1), 95% of the mutations engendered by RNA virus polymerases are deleterious and therefore not conserved (42).

Our observation that inhalation of HCoV-OC43 led to a generalized infection of the whole CNS in mice demonstrates neuroinvasiveness. This result confirms that HCoVs have neuroinvasive properties in mice, which was first shown in newborn mice (10, 22) and which is consistent with their detection in human brain (9, 12, 40, 53). After inhalation, the first infected cells were detected in the olfactory bulb, illustrating that virus directly reached the brain by a transneuronal route, as already demonstrated for MHV (10, 31, 47). The HCoV-OC43 Paris isolate, which was never propagated in mouse brain or other neurological tissue, also exhibited neuroinvasive properties in mice. Replication within the CNS was similar for the two variants, but fewer mice were infected by the HCoV-OC43 Paris isolate than by the ATCC strain. These data suggest that only one mutation in the S gene, giving rise to one amino acid modification, could partially modulate the neuroinvasiveness of one variant over the other. Indeed, a single amino acid change has already been demonstrated to influence MHV ability to spread within the CNS (43, 59).

Although the degree of sequence conservation between the genomes of the HCoV-OC43 ATCC and Paris variants is very high, their phenotypes seem to differ slightly in mice, since the ATCC strain reached the CNS more easily. As we have demonstrated in vitro with primary hippocampus and cortical cell cultures, both HCoV-OC43 ATCC and Paris variants were able to replicate in rodent neurons, although the HCoV-OC43 ATCC strain yielded more infectious virus particles than the HCoV-OC43 Paris isolate. However, the two viral variants exhibited different biological properties, such as plaque formation and cytopathic effects on different cell lines (H. Jacomy and P. J. Talbot, unpublished data).

Although both mutations preserve some but not all properties of the parental residues, the I958F mutation leads to a substitute phenylalanine that does not display the same steric hindrance than the isoleucine, which could potentially affect protein folding and function. Moreover, the I958F mutation is located in the S2 subunit of the S gene and would probably be positioned in the putative fusion peptide domain (23), conferring a lot of impact to this mutation at the biological level. On its own, this mutation could therefore have the capacity to influence the phenotype of the HCoV-OC43 Paris isolate because it may interfere with the fusion process in a positive or a negative manner (43). Given the known involvement of the S protein in viral biology and pathogenesis (7, 8, 15, 48), this mutation is more likely to influence the phenotype of the Paris isolate. It has been reported that the N protein may be involved in viral RNA synthesis (30) and that it could colocalize with nucleolar antigens and delay the cell cycle (14). However, the fact that the V81A mutation within the N gene is positioned in domain I of the protein should not influence the RNA binding properties of N, since this functional feature of the protein lies in domain II (45). Therefore, even though the role of both mutations needs to be investigated, we feel that the S mutation is more likely to influence the virus phenotype.

Comparison with better-characterized coronaviruses (23, 59) suggests that the I958F mutation is located in the putative S fusion peptide and could therefore affect viral fusogenic properties and phenotype. Although no fusion peptides have formally been identified in any coronavirus S protein, predictions have located such fusion sequences near the N terminus of the heptad repeat 1 (HR1) for MHV (33). Studies with the MHV-A59 S protein also showed that mutations introduced in the HR1 region severely affected cell-cell fusion ability (33). Moreover, it has already been reported that a single mutation introduced in HR1 could influence the degree of MHV virulence (59). Depending on the effect of the mutation on cleavage ability, the phenotype of the resulting virus could also be affected. Although the cleavage of the S protein is not absolutely required for fusion (23, 52, 55), it has been shown to enhance fusogenicity (55). Thus, inhibition of S-protein cleavage would be associated with a more stable interaction between S1 and S2 and would correlate with a loss of fusogenicity (25). So, as observed by Tsai et al. (59) for the MHV-JHM strain, the I958F mutation in the S gene of the HCoV-OC43

Paris isolate could either alter the conformation of the S protein or have an incidence on its cleavage, impairing the ability of the virus to spread within the CNS.

An animal model for the HCoV-OC43 ATCC strain has recently been developed and optimized in our laboratory (22). Moreover, HCoV-OC43 may also be used as a model for the study of SARS-HCoV, not only because of the identity level the two virus strains display but because HCoV-OC43 can also be studied without the requirement of a level three, aerosol-aware, biological confinement. Indeed, we have now demonstrated that the two virus strains present a high level of conservation for some essential functional domains, especially within 3CLpro, the RdRp, and the RNA helicase. This result is consistent with the possible sharing of several important properties by these two viruses. All of these motifs represent potential candidates for therapy of coronavirus-mediated diseases because they are specific targets and because of the specificity they exhibit toward their substrate. Indeed, substrate specificities of all coronavirus proteases, and mainly 3CLpro, are conserved among the three established groups (19), and this is also true for SARS-HCoV. The picornavirus RdRp (21) and viral proteases (17) have notably been designated as such targets for antiviral therapy. At present, the SARS-HCoV 3CLpro enzyme represents the most promising target for SARS therapy (58). The availability of 3CLpro crystal structures should provide a valuable tool for rapid identification of potential drugs against SARS. Thus far, 3CLpro crystal structures have been obtained for transmissible gastroenteritis virus (TGEV) (5), HCoV-229E (6) and, more recently, for SARS-HCoV (61). A putative in vitro inhibitor has also been identified for TGEV (5) and SARS-HCoV (61). This inhibitor, hexapeptidyl chloromethyl ketone, was shown to bind the 3CLpro enzyme very efficiently in vitro and, although it provides an excellent structural basis for drug design, in vivo experiments need to be performed on this issue.

Now that the complete genome sequence of HCoV-OC43 has been deciphered, it will provide a very useful tool for the study of coronaviruses from all genetic groups and particularly for those of group 2, including SARS-HCoV. Indeed, the genome sequence will allow comparative studies with other coronavirus strains and RNA viruses and will also allow optimization of prediction models. This sequence will also allow the assembly of an infectious cDNA clone of HCoV-OC43, which is currently under way. Thus far, cDNA clones have been assembled for several coronavirus strains by using different approaches. Among these clones, those of TGEV (3, 64), HCoV-229E (57), IBV (13), MHV-A59 (66), and even SARS-HCoV (65) are now available. The HCoV-OC43 clone will provide an invaluable tool to further understand the underlying mechanisms for replication and pathogenesis of HCoVs.

ACKNOWLEDGMENTS

We are grateful to Francine Lambert for technical assistance with cell and virus culture. We also thank D. W. Yoo, who kindly provided the ORF1ab sequence from the BCoV, which allowed us to sequence the HCoV-OC43 ORF1ab.

This study was supported by grant MT-9203 from the Canadian Institutes for Health Research (Institute of Infection and Immunity). P.J.T. is the holder of a Tier-I (Senior) Canada Research Chair. J.S.J. acknowledges a studentship from the Fonds Québécois de Recherche sur la Nature et les Technologies.

REFERENCES

1. **Adami, C., J. Pooley, J. Glomb, E. Stecker, F. Fazal, J. O. Fleming, and S. C. Baker.** 1995. Evolution of mouse hepatitis virus (MHV) during chronic infection: quasispecies nature of the persisting MHV RNA. Virology **209:** 337–346.
2. **Addie, D. D., I. A. T. Schaap, L. Nicolson, and O. Jarrett.** 2003. Persistence and transmission of natural type I feline coronavirus infection. J. Gen. Virol. **84:**2735–2744.
3. **Almazán, F., J. M. González, Z. Pénzes, A. Izeta, E. Calvo, J. Plana-Durán, and L. Enjuanes.** 2000. Engineering the largest RNA virus genome as an infectious bacterial artificial chromosome. Proc. Natl. Acad. Sci. USA **97:** 5516–5521.
4. **Alonso, S., A. Izeta, I. Sola, and L. Enjuanes.** 2002. Transcription regulatory sequences and mRNA expression levels in the coronavirus transmissible gastroenteritis virus. J. Virol. **76:**1293–1308.
5. **Anand, K., G. J. Palm, J. R. Mesters, S. G. Siddell, J. Ziebuhr, and R. Hilgenfeld.** 2002. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α-helical domain. EMBO J. **21:**3213–3224.
6. **Anand, K., J. Ziebuhr, P. Wadhwani, J. R. Mesters, and R. Hilgenfeld.** 2003. Coronavirus main proteinase ($3CL^{pro}$) structure: basis for design of anti-SARS drugs. Science **300:**1763–1767.
7. **Arbour, N., G. Côoté, C. Lachance, M. Tardieu, N. R. Cashman, and P. J. Talbot.** 1999. Acute and persistent infection of human neural cell lines by human coronavirus OC43. J. Virol. **73:**3338–3350.
8. **Arbour, N., S. Ekandé, G. Côoté, C. Lachance, F. Chagnon, M. Tardieu, N. R. Cashman, and P. J. Talbot.** 1999. Persistent infection of human oligodendrocytic and neuroglial cell lines by human coronavirus 229E. J. Virol. **73:**3326–3337.
9. **Arbour, N., R. Day, J. Newcombe, and P. J. Talbot.** 2000. Neuroinvasion by human respiratory coronaviruses. J. Virol. **74:**8913–8921.
10. **Barthold, S. W., M. S. de Souza, and A. L. Smith.** 1990. Susceptibility of laboratory mice to intranasal and contact infection with coronaviruses of other species. Lab. Anim. Sci. **40:**481–485.
11. **Bonavia, A., N. Arbour, V. W. Yong, and P. J. Talbot.** 1997. Infection of primary cultures of human neural cells by human coronaviruses 229E and OC43. J. Virol. **71:**800–806.
12. **Burks, J. S., B. L. DeVald, L. D. Jankovsky, and J. C. Gerdes.** 1980. Two coronaviruses isolated from central nervous system tissue of multiple sclerosis patients. Science **209:**933–934.
13. **Casais, R., V. Thiel, S. G. Siddell, D. Cavanagh, and P. Britton.** 2001. Reverse genetics system for the avian coronavirus infectious bronchitis virus. J. Virol. **75:**12359–12369.
14. **Chen, H., W. Torsten, P. Britton, G. Brooks, and J. A. Hiscox.** 2002. Interaction of the coronavirus nucleoprotein with nucleolar antigens and the host cell. J. Virol. **76:**5233–5250.
15. **Chouljenko, V. N., X. Q. Lin, J. Storz, K. Kousoulas, and A. E. Gorbalenya.** 2001. Comparison of genomic and predicted amino acid sequences of respiratory and enteric bovine coronaviruses isolated from the same animal with fatal shipping pneumonia. J. Gen. Virol. **82:**2927–2933.
16. **Denison, M. R., W. J. Spaan, Y. van der Meer, C. A. Gibson, A. C. Sims, B. Prentice, and X. T. Lu.** 1999. The putative helicase of the coronavirus mouse hepatitis virus is processed from the replicase gene polyprotein and localizes in complexes that are active in viral RNA synthesis. J. Virol. **73:**6862–6871.
17. **Dragovich, P. S., T. J. Prins, R. Zhou, T. O. Johnson, Y. Hua, H. T. Luu, S. K. Sakata, E. L. Brown, F. C. Maldonado, T. Tuntland, C. A. Lee, S. A. Fuhrman, L. S. Zalman, A. K. Patick, D. A. Matthews, E. Y. Wu, M. Guo, B. C. Borer, N. K. Nayyar, T. Moran, L. Chen, P. A. Rejto, P. W. Rose, M. C. Guzman, E. Z. Dovalsantos, S. Lee, K. McGee, M. Mohajeri, A. Liese, J. Tao, M. B. Kosa, B. Liu, M. R. Batugo, J. P. Gleeson, Z. P. Wu, J. Liu, J. W. Meador III, and R. A. Ferre.** 2003. Structure-based design, synthesis and biological evaluation of irreversible human rhinovirus 3C protease inhibitors. 8. Pharmacological optimization of orally bioavailable 2-pyridone-containing peptidomimetics. J. Med. Chem. **46:**4572–4585.
18. **Gorbalenya, A. E.** 2001. Big nidovirus genome. When count and order of domains matter. Adv. Exp. Med. Biol. **494:**1–17.
19. **Hegyi, A., and J. Ziebuhr.** 2002. Conservation of substrate specificities among coronavirus main proteases. J. Gen. Virol. **83:**595–599.
20. **Horvath, C. M., M. A. Williams, and R. A. Lamb.** 1990. Eukaryotic coupled translation of tandem cistrons: identification of the influenza B virus BM2 polypeptide. EMBO J. **9:**2639–2647.
21. **Hung, M., C. S. Gibbs, and M. Tsiang.** 2002. Biochemical characterization of rhinovirus RNA-dependent RNA polymerase. Antivir. Res. **56:**99–114.
22. **Jacomy, H., and P. J. Talbot.** 2003. Vacuolating encephalitis in mice infected by human coronavirus OC43. Virology **315:**20–33.
23. **Jan Bosch, B., R. van der Zee, C. A. M. de Haan, and P. J. M. Rottier.** 2003. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. J. Virol. **77:**8801–8811.
24. **Kamahora, T., L. H. Soe, and M. M. C. Lai.** 1989. Sequence analysis of

nucleocapsid gene and leader RNA of human coronavirus OC43. Virus Res. **12:**1–9.
25. **Krueger, D. K., S. M. Kelly, D. N. Lewicki, R. Ruffolo, and T. M. Gallagher.** 2001. Variations in disparate regions of the murine coronavirus spike protein impact the initiation of membrane fusion. J. Virol. **75:**2792–2802.
26. **Kündel, F., and G. Herrler.** 1993. Structural and functional analysis of the surface protein of human coronavirus OC43. Virology **195:**195–202.
27. **Kündel, F., and G. Herrler.** 1996. Structural anf functional analysis of the S proteins of two human coronavirus OC43 strains adapted to growth in different cells. Arch. Virol. **141:**1123–1131.
28. **Kuo, L., G.-J. Godeke, M. J. B. Raamsman, P. S. Masters, and P. J. M. Rottier.** 2000. Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. J. Virol. **73:**1393–1406.
29. **Labonté, P., S. Mounir, and P. J. Talbot.** 1995. Sequence and expression of the ns2 protein gene of human coronavirus OC43. J. Gen. Virol. **76:**431–435.
30. **Lai, M. M. C., and D. Cavanagh.** 1997. The molecular biology of coronaviruses. Adv. Virus Res. **48:**1–97.
31. **Lavi, E., P. S. Fishman, M. K. Highkin, and S. R. Weiss.** 1988. Limbic encephalitis after inhalation of a murine coronavirus. Lab. Investig. **58:**31–36.
32. **Lu, Y., X. Lu, and M. R. Denison.** 1995. Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59. J. Virol. **69:**3554–3559.
33. **Luo, Z., and S. R. Weiss.** 1998. Roles in cell-to-cell fusion of two conserved hydrophobic regions in the murine coronavirus spike protein. Virology **244:**483–494.
34. **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.
35. **Masters, P. S.** 1999. Reverse genetics of the largest RNA viruses. Adv. Virus Res. **53:**245–264.
36. **Meyers, G.** 2003. Translation of the minor capsid protein of a calicivirus is initiated by a novel termination-dependent reinitiation mechanism. J. Biol. Chem. **278:**34051–34060.
37. **Mounir, S., and P. J. Talbot.** 1992. Sequence analysis of the membrane protein gene of human coronavirus OC43 and evidence for O glycosylation. J. Gen. Virol. **73:**2731–2736.
38. **Mounir, S., and P. J. Talbot.** 1993. Molecular characterization of the S protein gene of human coronavirus OC43. J. Gen. Virol. **74:**1981–1987.
39. **Mounir, S., and P. J. Talbot.** 1993. Human coronavirus OC43 RNA 4 lacks two open reading frames located downstream of the S gene of bovine coronavirus. Virology **192:**355–360.
40. **Murray, R. S., B. Brown, D. Brian, and G. F. Cabirac.** 1992. Detection of coronavirus RNA and antigen in multiple sclerosis brain. Ann. Neurol. **31:**525–533.
41. **Myint, S. H.** 1994. Human coronaviruses: a brief review. Rev. Med. Virol. **4:**35–46.
42. **Novella, I., S.** 2003. Contributions of vesicular stomatitis virus to the understanding of RNA virus evolution. Curr. Opin. Microbiol. **6:**399–405.
43. **Ontiveros, E., T. S. Kim, T. M. Gallagher, and S. Perlman.** 2003. Enhanced virulence mediated by the murine coronavirus, mouse hepatitis virus strain JHM, is associated with a glycine at residue 310 of the spike glycoprotein. J. Virol. **77:**10260–10269.
44. **Ozdarendeli, A., S. Ku, S. Rochat, G. D. Williams, S. D. Senanayake, and D. A. Brian.** 2001. Downstream sequences influence the choice between a naturally occurring noncanonical and closely positioned upstream canonical heptameric fusion motif during bovine coronavirus subgenomic mRNA synthesis. J. Virol. **75:**7362–7374.
45. **Parker, M. M., and P. S. Masters.** 1990. Sequence comparison of the N genes of five strains of the coronavirus mouse hepatitis virus suggests a three domain structure for the nucleocapsid protein. Virology **179:**463–468.
46. **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, and K. Y. Yuen.** 2003. Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. Lancet **361:**1767–1772.
47. **Perlman, S., G. Evans, and A. Afifi.** 1990. Effect of olfactory bulb ablation on spread of a neurotropic coronavirus into the mouse brain. J. Exp. Med. **172:**1127–1132.
48. **Rowe, C. L., S. C. Baker, M. J. Nathan, and J. O. Fleming.** 1997. Evolution of mouse hepatitis virus: detection and characterization of spike deletion variants during persistent infection. J. Virol. **71:**2959–2969.
49. **Sawicki, S. G., and D. L. Sawicki.** 1998. A new model for coronavirus transcription. Adv. Exp. Med. Biol. **440:**215–219.
50. **Schickli, J. H., B. D. Zelus, D. E. Wentworth, S. G. Sawicki, and K. V. Holmes.** 1997. The murine coronavirus mouse hepatitis virus strain A59 from persistently infected murine cells exhibits an extended host range. J. Virol. **71:**9499–9507.
51. **Snijder, E. J., P. J. Bredenbeek, J. C. Dobbe, V. Thiel, J. Ziebuhr, L. L. Poon, Y. Guan, M. Rozanov, W. J. M. 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.
52. **Stauber, R., M. Pfleiderera, and S. Siddell.** 1993. Proteolytic cleavage of the murine coronavirus surface glycoprotein is not required for fusion activity. J. Gen. Virol. **74:**183–191.
53. **Stewart, J. N., S. Mounir, and P. J. Talbot.** 1992. Human coronavirus gene expression in the brains of multiple sclerosis patients. Virology **191:**502–505.
54. **Sturman, R. M., C. S. Ricard, and K. V. Holmes.** 1990. Conformational change of the coronavirus peplomer glycoprotein at pH 8.0 and 37°C correlates with virus aggregation and virus-induced cell fusion. J. Virol. **64:**3042–3050.
55. **Taguchi, F.** 1993. Fusion formation by the uncleaved spike protein of murine coronavirus JHMV variant cl-2. J. Virol. **67:**1195–1202.
56. **Taguchi, F.** 1995. The S2 subunit of the murine coronavirus spike protein is not involved in receptor binding. J. Virol. **69:**7260–7263.
57. **Thiel, V., J. Herold, B. Schelle, and S. G. Siddell.** 2001. Infectious RNA transcribed in vitro from a cDNA copy of the human coronavirus genome cloned in vaccinia virus. J. Gen. Virol. **82:**1273–1281.
58. **Thiel, V., K. A. Ivanov, Á. Putics, T. Hertzig, B. Schelle, S. Bayer, B. Weissbrich, E. J. Snijder, H. Rabenau, H. W. Doerr, A. E. Gorbalenya, and J. Ziebuhr.** 2003. Mechanisms and enzymes involved in SARS coronavirus genome expression. J. Gen. Virol. **84:**2305–2315.
59. **Tsai, J. C., L. de Groot, J. D. Pinon, K. T. Iacono, J. J. Phillips, S. H. Seo, E. Lavi, and S. R. Weiss.** 2003. Amino acid substitutions within the heptad repeat domain 1 of murine coronavirus spike protein restrict viral antigen spread in the central nervous system. Virology **312:**369–380.
60. **Tsang, K. W., P. L. Ho, G. C. Ooi, W. K. Yee, T. Wang, M. Chan-Yeung, W. K. Lam, W. H. Seto, L. Y. Yam, T. M. Cheung, P. C. Wong, B. Lam, M. S. Ip, J. Chan, K. Y. Yuen, and K. N. Lai.** 2003. A cluster of cases of severe acute respiratory syndrome in Hong Kong. N. Engl. J. Med. **348:**1977–1985.
61. **Yang, H., M. Yang, Y. Ding, Y. Liu, Z. Lou, Z. Zhou, L. Sun, L. Mo, S. Ye, H. Pang, G. F. Gao, K. Anand, M. Bartlam, R. Hilgenfeld, and Z. Rao.** 2003. The crystal structures of severe acute respiratory syndrome virus main protease and its complex with an inhibitor. Proc. Natl. Acad. Sci. USA **100:**13190–13195.
62. **Yoo, D. W., M. D. Parker, and L. A. Babiuk.** 1991. The S2 subunit of the spike glycoprotein of bovine coronavirus mediates membrane fusion in insect cells. Virology **180:**291–304.
63. **Yoo, D., and Y. Pei.** 2001. 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.
64. **Yount, B., K. M. Curtis, and R. S. Baric.** 2000. Strategy for systematic assembly of large RNA and DNA genomes: transmissible gastroenteritis virus model. J. Virol. **74:**10600–10611.
65. **Yount, B., K. M. Curtis, E. A. Fritz, L. E. Hensley, P. B. Jahrling, E. Prentice, M. R. Denison, T. W. Geisbert, and R. S. Baric.** 2003. Reverse genetics with a full-length infectious cDNA of severe acute respiratory syndrome coronavirus. Proc. Natl. Acad. Sci. USA **100:**12995–13000.
66. **Yount, B., M. R. Denison, S. R. Weiss, and R. S. Baric.** 2002. Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59. J. Virol. **76:**11065–11078.
67. **Ziebuhr, J., E. J. Snijder, and A. E. Gorbalenya.** 2000. Virus-encoded proteinases and proteolytic processing in the *Nidovirales*. J. Gen. Virol. **81:**853–879.