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, France²

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 positivestrand 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).

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 RNAdependent 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-

^{*} 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.

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) CO₂ 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 μ l of a virus stock solution containing 10⁶ 50% tissue culture infective dose(s) (TCID₅₀)/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 μ m 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) H₂O₂, 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) CO₂ humid atmosphere, the cells were washed in PBS and fixed with 0.3% (vol/vol) hydrogen peroxide (H₂O₂) 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 in DAB (Sigma) with 0.01% (vol/vol) H₂O₂.

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 μ g 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

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'	

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

^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 $5' \rightarrow 3'$ 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 sequer	ncing of the	HCoV-OC43	genome ^a
----------	--------------	------------	--------------	-----------	---------------------

piner Superior Superior Superior Datame (0) piner Superior Datame (0) NACEUUE GEGATGCTGGCTGGATTGCTGCGC 380-400 JUBI1 CCATGACGTCGTAACGTCTGGCAGCTC 579.8770 JUBI3 GGATAGTCGCAGATGTCGCGCG 380-400 JUBI1 CCATGACGTCGTAACGTCGCAGCAGCTC 579.8770 JUBI3 GGACAATCTCCAGGTGGTTGCTGCAGATG 189-1821 JUBI13 ACCCATGCCATGACGACACGTC 459478 JUBI41 CTTATATAGTAGTGGAGAGTG 225-2272 JUBI13 ACCACACGCATAGACGACACTC 579.8770 JUBI43 CCAGTCATGAGAGTGACGG 321-321 JUBI34 CCAGTCATGAGAGTGATGACG 321-321 JUBI44 TAACGTATGTGGAGGCAG 321-321 JUBI35 CACACAGCATAGACACTTCC 539-519 JUBI45 TAACGTATGTCAGAGCG 321-321 JUBI35 CACACAGCATAGTAGTAGCACATC 539-539 JUBI46 ATACTTATGTAGTGGGCACACTC 329-541 JUBI35 CACACAGCATATACCACACTC 539-531 JUBI46 TAACTTATGTAGTGGC 321-321 JUBI47 GGTGTAGAGTAGTCTAGCGGACACTC 512-5392 JUBI46	Positive-strand	Sequence $(5', 2')$	Logation (nt)	Negative-strand	Sequence (51 × 21)	Location (nt)
RACEJUB2 TGTGATGGTGGATGGTGGCG 380-400 JUBIL CCATGACTTCTGAACGTCTTC 5786770 JUBIS TGTAATGGTGGTGGATGTCGCG 863-82 JUBILS TGTGTGTGGATGTGAAC 448-4186 JUBIS TGTCATGTCTGGGTGGAGACG 223-227 JUBIS ACGGTTTACAACACTTCTGG 448-4186 JUBILS TGTCATGTGGGGAGAGTG 223-227 JUBIS GGTTTAGGACACATCTGTG 438-4186 JUBILS GTTCATGGTGGATGTGGGGGAGAGTG 223-237 JUBIS ACGGGTTGCATATGCAGACATC 339-3473 JUBILS GTTCATGGATGGTGGAGAGTG 233-237 JUBIS ACGGTTGCACATACACACTC 239-2413 JUBILS TGTCATGTGTGGATGGTGGAGAGTG 233-3413 JUBIS ACGCAAAACCATTACTGTG 399-3975 JUBILS TATGTTGGGGGGACACGC 233-5418 JUBIS CAAGCAAAACCATTACTGCG 239-3217 JUBIS TATGTTGGGGGGACACGC 232-332 JUBIS ATGTAAAGCACATTACTACCACCTC 239-321 JUBIS TATGTTGGGGGACACGC 232-332 JUBIS ACGTTGTCGGCGACACGC 232-332 JUBIS TGTACCACAGAGTGGTGGGACACGC 232-3348 <	primer	Sequence $(5' \rightarrow 3')$	Location (nt)		Sequence $(5' \rightarrow 3')$	Location (nt)
JUB31 GTTATATAGATIGATCCTGC 682-602 JUB111 CCATGCCTCTGCATGCATCCC 581-603 JUB131 GATTGATACCCATCAACACCATCAACAC 4918-493 JUB141 GATTGATCTGTGTGTGATGCTGC 681-603 JUB115 ACGCGCATACAACGATCATGCCC 4918-493 JUB141 GATTGATCATGAGAGAGTG 255-2277 JUB151 ACGCGCATAACAATGATGATGCC 395-3131 JUB141 GATGTGAGATGTGC 393-3131 JUB151 ACGCATGACAAGAAGCCC 395-3431 JUB141 ATGCATGTGTCATGTTGTGTGGGG 393-3131 JUB151 ACGCATGACAACACCC 2918-393 JUB161 ATGCATGTGTCTTTTATGGGG 390-3970 JUB131 ACGCATGACCATTATGGCG 191-3011 JUB164 AAGTTATGTATGTTTGTAGAGG 391-3933 JUB163 ATGCATGGCATGTTTCC 191-4913 JUB164 ACGTTGCAGATTTGTATGAGGG 391-3933 JUB61 ATGCATGGAGATTTCC 191-4913 JUB165 GTTGACGATGTATGTTGC 591-5933 JUB61 ATGCAGCAGCATGGAAGCCG 192-4973 JUB170 GCTTGCAGATTTGTATGAGGG 791-5937 JUB181 GTGTAGGAGCATGAAACCGG 592-3972 <tr< td=""><td></td><td></td><td></td><td></td><td></td><td></td></tr<>						
JUB22 TGATTATACTGETAGTECTTEC 983-1063 JUB112 TGAATTAACCACAACAG 4158-498 JUB35 GACAANCTTACAGGTETTTEC 1884-168 JUB112 ACCTTACAGATTAGCAACAG 4158-418 JUB141 CTTACTAGTAGTAGTGETAGAACAT 1884-168 JUB112 ACCTTACAGATAGTGETAGCACA 4158-418 JUB142 CTTACAGTATTGETAGCACAT 2557-2577 JUB151 ACACACTTCTACAGCACTCC 2193-311 JUB143 ACAGCTTGTACGGAATGGATGGACA 3013-3311 JUB152 ACACACACACACACAGCAGTTCTCTC 2193-2319 JUB163 AATTATTGGGCATGGTACGGAGGC 3111-4331 JUB131 CAAATCTTTCTACCGTCCACCTC 2193-2319 JUB164 AATTATTGGGCATGTACTGGAGGC 3111-4331 JUB131 CAAACCTATTCTACGTCCCCCT 329-3323 JUB164 AATTATTGGGCATGTACTGGAGCC 3111-4331 JUB163 CTCTAAAGTCTCCCCCTC 323-332 JUB164 AATTATGGAGATTGTACCGAGGC 3111-4318 JUB164 CTCTAAGAAGTCTCCCCCTC 323-332 JUB165 AATTATGGGATGTACTGAGGC 3111-4311 JUB164 CTCTAAGAAGTCTCCACCCC 323-332 JUB172 CTCTAG						
JUBH4 AGTTGCTAGGTGTGCAGATG 180.1821 JUBIS AGGGGTTACAMOATAGACTATCTCC 488-4106 JUBH4 GTTCTATATATATAGATGGGAGATC 254-227 JUBIS GCTTTAGACATAGACACC 387-3483 JUBH4 GTTCTAGATTGTCATTGGGAG 253-2577 JUBIS GCTTTAGACATAGACACC 389-3493 JUBH6 TATATCTGTGCATGGTAGTGGTGGT 380-3709 JUBIS CAAATGCTTTACACCATATGGC 239-2419 JUBH6 TATATGTGGGAGATGGTAGTGGT 389-379 JUBIS CAAGCAATGTAGGTAGTGTAGTGGT 239-2419 JUBH6 AAGTTAGTATGTATGTAGGGG 4310-4339 JUBIS2 CCTCAAAGTAGGGATAGTCGT 122-1417 JUBH6 AAGTTAGTAGTATGTATGTAGGGG 539-339 JUB61 ATGCAAGCATGGAAGGCACGAAGGC 111-1131 JUBH5 CCTGCAAGATGTGTGGGAGGACAGACG 531-5333 JUB63 ATGCAGCAGAAAGGACTAGAGC 111-1131 JUBH7 GGTTGTAGAGACGACGACGAAGAAGACGACCAGAGACAAAAGGACTAGAGACATGAGACAACAAGGACGAACAAAAGGACATAGAGCAAGAGACAAGAGACAAGAGACAAGAGACAAGAGACAAGAGACAAGAGACAAGAGACAAGAGACAAGAGACAAGAGACAAGAAG						
JUB141 CITATATAGTAGTGGAGAGTC 2254-2274 JUB15 AACAGCCATAGAATACAGTACTAC 378-386 JUB142 ATAGCTETCAAGATGATGAC 2857-2371 JUB133 ACACAGCATACAGACCTAC 238-3435 JUB143 ATAGCTETCAAGATGATGAC 2871-2371 JUB133 ACACAGCATACAGACCTACCCC 2913-3235 JUB164 AAGCTATACTATCTTC 2930-379 JUB133 ACACCAACAACCATCACATCCCC 1991-2011 JUB164 AAGTTATTGATGCGG 3900-379 JUB131 CAAGCAAAACCACTCACCCC 1191-321 JUB164 AAGTTATTGATGCGGG 3910-339 JUB131 CAAGCAAACCAATCAGACCCCC 1191-321 JUB166 TAGTTACTGATGGTGACAGG 3910-339 JUB20 CAGCCAAACAGTCATTCCT 1192-4174 JUB151 ACTCAGCGTATTATTAAAGCC 5310-5339 JUB62 CATCTAAAAGCAATCAGACC 1092-4174 JUB152 GTGCAGAGTTGGTGACAGC 6310-6321 JUB62 CATCTAGAAGCAACAGCAGC 1092-4174 JUB153 AGCCGAGATTGGTGACAGC 6310-6321 JUB62 CATCTAGAGCATCAGACATCAGACC 1292-3233 JUB153 CATCGACAAAGCAACAGCAGCAGC						
JUBH2 GTCTCGATTTTCATTAGCGG 257-257 JUBI51 GCTTTAGGCATACAGACCC 2395-315 JUBH4 TAACCTTATGGATTGGCAG 2038-311 JUBI52 ACCACAGCTCAGTCAGTCC 2398-315 JUBH54 ACACGTCCTGATGGTAGTGCAG 2398-319 JUBI51 CAAATCTTCTCAGTCAAATCC 2399-329 JUBH54 AATTATTGGATGGCAGG 339-315 JUBI51 CAACGACATCAGTCACTC 2399-329 JUBH54 AATTATTGGATGGCAGG 4319-439 JUBI51 CACGACAAATCATTATCACGACCC 139-149 JUBH54 AAGTTATGGCACAGC 4782-4492 RACE JUBI GAAGCAAATCACTATACGACCC 139-333 JUBH52 GTTCTACCACAGTATTATATAAACCC 531-4533 JUB62 AGGTATGAAAGTCC 4782-492 JUBS2 GTTCTACCACAGAATTGTTATATAAACCC 531-4533 JUB62 AGGCAGTAGGCAATGC 4782-492 JUBS2 GTTCTACCACAAATGCATGGTAGTGC 501-5733 JUB82 AGTCTAACACACATCAGAC 693-9473 JUBS2 GTTCTACACAAAAGGTCCC 591-5703 JUB82 AGTCTAACACAAATGCATGCAGC 921-923 JUBS2 GTTCTACACAAAAGGTCCC 591-5707						
JUB144 TTAACCTTATGGTGGCAG 362-1-3641 JUB153 CAAATTCTTTACGTCCTCACC 2999-2010 JUB161 TATTATGGGCATGGTAGTC 3700-3720 JUB131 CAAGCAAAACCATTATCATCG 1427-1447 JUB164 TATTATGGGCATGGTAGTC 3899-3979 JUB130 CAAGCAAAACCATTATCATCATCG 1427-1447 JUB166 TATTATGGGCATGGTAGTC 3899-3979 JUB161 CAAGCAAAACCATTATCAAATGC 1111-1131 JUB166 TRCTTQGCTATTGCAACATGC 319-333 JUB61 CATGCAAAGCACAAGCACG 1224-1074 JUB151 CATCAAGGCTATTGTAAAAGCC 294-433 JUB62 CTATTAAAAGCAACATGC 1225-1472 JUB152 GTGTAACCATGATAAAAGTCT 717-7197 JUB181 CATGACAACATTAAAAGTCC 291-2933 JUB172 GCTGGCACTTCAAAAAGTGTTC 707-7197 JUB181 CATTACACAACATTAGCAAAAGTCT 291-2913 JUB172 GCTGGCACTTCAAAAGTGGTAGTTAG 9849-890 JUB182 CATTACAAGGCACATCCAC 8945-8950 JUB173 CCTGGCACTTCAAAAGTGGTAGTTAG 9849-890 JUB184 CATTACACTGACGTAGTAACACC 8945-8950 JUB185 CCTGTCAACTTCAA	JUB142	GTTCTGATTTTTCATTAGCGG	2557-2577	JUB151	GCTTTAGGCACATACAGACCC	3395-3415
JUBI61 ATGCTATGTTCTTTATGGTG 370-3720 JUBI5 ATAGCATGAAACCATCATCCC 1991-2011 JUBI64 AAGTTATGTGACGTGGTATGTT 3959-3979 JUBI51 CAACCAAACCATCATCATGCC 1949-1069 JUBI65 TATTTGGATGGTAGTAGTGGC 4782-4402 RACE JUBI CGACCAAACCATCATCATCCC 132-312 JUBI65 TATTTGGATGTGGTGGTGGGC 5319-5339 JUB62 ATAGCAACGACGACAGCGGGTAGTGACAGC 5913-5933 JUBS1 ATCCAACGCATAGTGTGGTGGGCAACGC 5913-5933 JUB62 AGGGTACAGTGACGGCAAGTGGGGGAAAGCG 5913-9733 JUBS1 ATCCAACGCAAAGGGTATTGTATGTTGG 5705-766 JUB181 ATCCTAACGCAAAGGGCTACGG 5913-9733 JUB171 GGCAGGGATTTGATATGTGTG 5705-766 JUB181 ATCTCAACGCAAAGGCTACGC 5913-9733 JUB172 CTGGGCAGTGGTAGTGTG 7075-706 JUB181 ATCTCAACGCAAGTGTGTGGC 5913-9733 JUB173 ATAAGCAGGATGGTATGT 8403-8463 JUB181 ACTCTACACAGTAAGGCTAAGTG 5913-9733 JUB173 ATAAGCAGGATGGTAGTGTG 8403-8463 JUB181 ACTCTACACAGTAAGGCTAAGTG 5921-9723 JUB175 <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>						
JUBI63 TATTATTGGCATGCTATGTC 399-3979 JUBI31 CAAGCAAAACCATTATCATG 1427-1471 JUBI64 AAGTTATGTAGCGGAGC 4310-430 JUBIS2 CTCCAAGTGGAAATAGTGC 1149-1069 JUBI65 TATTTGAGCTGTACTGGAGC 4310-430 JUBIS2 CTCCAAGTGGAAATAGTCC 112-312 JUBIS1 CATCCAGCGAATTGTATCCACCTC 313-333 JUBIS2 CTCCAAGTGGCTAAACGC 235-9772 JUBIS3 AACTCAACCCACACGAAATTGC 703-755 JUBIS1 ATCTCAACCCACACACAAATTGC 703-755 JUBI72 CGCAGGTATTGTTATGTATGTTTC 703-7565 JUBI81 ATCTCACCACAAAAGTGCG 935-9772 JUBI73 AATCTACACCACACAGAAATTGC 9043-9063 JUBI81 ACTCACACAAAAATTGCACC 9359-779 JUBI73 CGTGGCTGATAGTTC 8073-8078 JUBI83 CATAGCAAAAATGTATCACCC 9359-779 JUBI73 CGTGGCTGATAGTTC 8073-808 JUBI84 ATCCACACAAAAATGGTATCACC 9359-779 JUBI73 GCTGTCTAATTTCACTGTATGTC 8073-808 JUBI85 CATAGCACAAAATGGTATGTAG 752-752 JUBI73 GCTGTCTACATTTATGTAG 941-94						
JUBI65 TATTTTGAGTGTACTGGAGGC 4782-4902 RACE JUBI GGAGCAAATCATATCCACTC 312-332 JUBI85 CCTGGTAGATTTGAACACTGC 5128-5139 JUBI60 ATTAGCAGCCAACAGACGTGTTCC 10724-1074 JUBIS1 ACTAGCGATTATTAAGCATCG 5913-5339 JUBIS2 ACTAGCGATCAACAGC 10724-1074 JUBIS1 ACTCAGCGATAGATTTGATATGTAAGCC 5911-5923 JUBIS1 GGGCAGATGTGTACCGGGGTAAACTG 10724-1074 JUBIS1 GGCAGCAATTGCAACAGC 5911-5925 JUBIS1 GGTAGGGACATTAACCACG 9835-9873 JUBI71 GGCTTGATCAGATGTGTATG 7087-7095 JUBI81 TGGTAGGACATTACCACGG 8745-8813 JUBI71 GGTTGTCAGATGGTGATTGT 8972-8992 JUBI82 GCATTCACACTGGGTAATTGCG 8753-8723 JUBS5 CACTTACAATGGGTAATTGG 8403-8403 JUBI84 ACATTACCACGGGTAATTACCCG 7854-9701 JUBS5 CACTTACAATGGGTAATTGG 8403-8403 JUBS6 CATCTACAATGGCTAATTGCGTATG 840-8403 JUBS7 GGGTGCTCACGTGACTGGTTAGTGGTATG 1078-10394 JUBS6 CACTCACAAAAGGTGTTCACCG 840-8564 JUBS7 <t< td=""><td>JUB163</td><td>TATTATTGGGCATGGTATGTC</td><td>3959-3979</td><td>JUB131</td><td>CAAGCAAAACCATCTATCATG</td><td>1427-1447</td></t<>	JUB163	TATTATTGGGCATGGTATGTC	3959-3979	JUB131	CAAGCAAAACCATCTATCATG	1427-1447
JUB16 TOCTTGCCTATTACAACATCC 128-5148 JUB6 CCTCTAAATCTTGTACGTTGTC 1111-1113 JUB51 ACTCAGCGTATTATTAAAGCC 591-5933 JUB62 AGGGTCACTGTACGTAAGAACAAGC 1025-1022 JUB51 ACTCAGCATTATTAAAGCC 691-5933 JUB62 AGGGTCACTGTAAGAACAAGC 1025-1022 JUB53 GTTAACCATGCTGTGTCACCACC 624-6244 JUB63 CTATTAAAACCAAAAGTCC 921-9733 JUB53 ATGCCAGGACATTAATTAACACACG 870-8705 JUB17 GCTTGTCTCATGATTGTC 971-9707 JUB17 GCTTGTCCATGTTGTCTC 777-7997 JUB183 CATCTAACACTCGGTTATAGTG 874-8763 JUB53 AACGATGACTAATGTC 804-9403 JUB845 ACGTTAATTCCCC 775-772 JUB55 CACTTAACTGCTTATTGCA 944-9403 JUB85 CATCAACACTCGGTTTATAGTA 752-7722 JUB55 CACTTACACTGTGTGTGGTTATG 949-9403 JUB84 AAAGTTTTACTCCAC 752-7722 JUB75 CACTTGCAGTTTATGTTATG 949-9403 JUB84 AACACTCCTGTGAGCACCACC 661-6901 JUB75 CACTTGACATGTGTTGTGTGTGTTATG 949-9403						
JUBS1 ACTCAGCGTATTATTÄÄÄÄGCC 9913-5933 JUB62 AGGGTCACTGAAGAACAAGC 10262-1025 JUBS2 GTGACGATGGTGGTGAGGACGC 624-6244 JUB84 TCATTAAAAGCAACATGGG 714AACTG 953-973 JUBS3 AATCTACACCACAGAAATTGC 6701-6721 JUB84 TCATTAAAAGCACAGTCGGGTAAACTTG 953-973 JUB17 GCTGTGTTGATAGATGTTAG 708-7698 JUB182 GCTTCTTCACACACAAAATTGCTG 874-9851 JUB17 GCTGTGTCTGTGTCATGAAGTTGTC 891-9902 JUB183 CATCAGACAAAAATTATACCCGG 844-886 JUB55 AAGGTTATATCG 943-9804 JUB184 CATCTACACTCGGTTATAGT 874-7850 JUB57 CGTGTCAACATGGCTAATGTC 943-9804 JUB66 CTAAAAGTGTTAAGCACCG 892-6945 JUB57 CGCTGTCAACATGGCTAATGTC 993-9945 JUB67 TATACAAGCAGCAGCC 612-6945 JUB71 CGTGTCTAACATGGCTAGTCC 1943-14463 JUB84 AAACTGTCTGTAAGCACCC 622-6946 JUB72 CAGCAGTTAAAACAGCTAGGTGGTAGGTAGAGC 11443-14463 JUB84 AAACTGCTGTCAACACCC 622-6945 JUB72 CAGCAGTTAACACTGGTAGGGCGAGGG						
JUBS2 GTTGACGATGGTGGTGACAGC 6264-6284 JUB64 AGAAAGTATGGGTAAACTG 9353-9472 JUBS4 TGGCAGGATTTGATATGTTAG 7036-7056 JUB181 ATCTCACACAAAAGGTCC 9213-9233 JUB17 GCTTTTGATGATCGTGATT 7078-7087 JUB181 ATCTCACACAAAAGGTCCC 9213-9233 JUB17 GCTTOTTCTATGATGCTGATT 7078-7089 JUB182 GCTTUTCAATTATGCGG 8354-8363 JUB55 AAGGTTTACGCTGTTTCCAG 9043-9063 JUB66 ACTAAAAGAGTCTTAATGC 7532-7723 JUB55 CACTTCAACATGCTATATTCT 6038-9560 JUB66 CAAAAGTGTTCAATGC 6732-6551 JUB57 CCGTCTCAACTTCATTGTTG 1078-10989 JUB67 TATAAAAGCAGCCGTGTTCC 6732-6545 JUB77 GGCTTCTACACTTTTGTTGTTG 1078-10989 JUB86 ACTCAACCTGTTAACGCGTGTCC 6712-6691 JUB73 GGCATGCTACACTTC 14901-14921 JUB86 ACTCACCGTGTAACACCG 6712-6691 JUB73 GGCATGCTACACTG 14901-14921 JUB86 CATCACCGTGTGACACCC 6511-6513 JUB73 GGCATGCTACACTG 1539-12807						
JUB53 AATCTACACCACAGAATTGC 6701–6721 JUB64 AGAAAGTTGGGTAAACTTG 9433–9473 JUB17 GGTTGTTACTGTTGTTGGTC 7177-7197 JUB181 TGTACGGCAGATAAACACTG 8718-8781 JUB17 GGTTGTTCATAGTCGTGGTG 7678-7089 JUB182 GCTTTCTCAAATTAGCCGTG 8718-8781 JUB172 CTGTGCGTCGTAAAAGTTGTTC 8072-8092 JUB183 AATCAAAAAGTTAGTCCAC 7950-7703 JUB185 AGGAGCAATAGGCAATGTGCAATG 8072-8092 JUB183 AATCAAAAAGTTATGCCAC 6795-6770 JUB55 CACTTAACAATGGCTAGTTATG 9540-9560 JUB86 CTAAAAGGTCTAACCAC 6792-6740 JUB75 CACTTAACAATGGCTAGTTATG 1907-10921 JUB86 AAATCTCGTGTAGCAAGGC 6711-691 JUB71 TGTAATCACGAATATACCTC 1143-11463 JUB81 GATGTTGGAAGGCAAACC 16117-16137 JUB72 CAGCATATAAACCACCTAGCI CAGGA JUB82 CATAATATCGCGTACAACGC AGGA 1522-1552 JUB73 ATAACTAGGGAATGTGCTAGTGTGT JUB82 CATAATATCACGGAACACCACG 1522-1542 JUB73 TGTAAAACAGCTAGCCAGG JU845 JUB84						
JUBI7 GGTTTTTACCCATTGTTTGCTC 7177-7197 JUBI812 GGTTGGCGACATTAAACACTG 8781-8781 JUBI72 CTGTGCTCTTAAAACTTGTTC 8072-8992 JUBI82 CATGGCAAAAAATGTTATGCCGG 7560-7701 JUBI73 AATAAGCAGATGGCTAATGTC 8072-8992 JUBI83 CATGGCAAAAAATGTTATGCCGC 7561-7563 JUBS6 AACGTATTAACCGGTTTTATG 9540-9560 JUB86 ACTAACAAAAGTGTCATACT 7223-7346 JUBS7 CGTGTCAAACAGCTTCATTGG 9540-9560 JUB86 ACTAACTGCGTCATACTACT 7223-7346 JUBS7 CGGTCTCAAACAGCTAGTTGGTTAG 10378-10392 JUB86 ACATCACTGTGGTGGC 6031-6511 JUB71 TGTATTCCACGATATACCCC 1143-11463 JUB81 CATGATCGTGACAAGGC 1632-1632 JUB73 ATAAGCAGTATCTGCTGCTACG 12547-12567 JUB83 AAAGTGTCTCCAAACGC 1505-1630 JUB74 TGTAAACAGCTAACTGCAAGG 1308-1144 JUB84 AAAGTGTCTCCAAACGC 1505-1630 JUB75 TGTGAATTCAGTGCAAGG 1308-1432 JUB87 TGTAAGTAGCACAACCC 13245-1325 JUB75 ACGCAGTTTCACAAGGC	JUB53	AATCTACACCACAGAAATTGC	6701-6721			9453-9473
JUB171 GCTTGTTCTATCATCCTGATG 7678–7698 JUB182 GCTTTCTCAATTTATGCTGC 8345–836 JUB173 AATAAGCAGATGGCTAATGTC 8403–8423 JUB184 AAGTTTATCCGCTTTATAGG 7534–7563 JUB55 AAGGTTTATCCGTCTTCCAG 943–9063 JUB185 CATAGAACAAAAATGGTCCAATGTAGG 7534–7563 JUB55 CACTTACAATGGCTAGTTATG 954–9560 JUB66 CTAAACAGTCCAATGTATCCC 651–651 JUB7 CCGTCTCAACTTTGTTTGCT 1008–1002 JUB87 ACTTAACAGTGTAAGGTCATCC 651–651 JUB7 CGGTTCACACTTAACAGTAGGA 1208–114 JUB88 AAATGTTGAGGGTCACGCACC 1611–16137 JUB72 CAGAGTTAAAACAGCTAGGA 1208+12104 JUB88 AAATGTGCAAGGACACACGC 15522–1532 JUB73 ATAATGAGGTAATGGTATGGA 1307–13093 JUB84 AAAAGTGCTACAACG 1588–1478 JUB74 TGTAATCAGGAAATGGATGGGA 1208–1450 JUB85 TATACCACATAGGTGTACACC 1348–1478 JUB75 CAGAGTATGGAAATGGTATGGG 1406–14048 JUB85 TCTGGATCACACG 1348–1478 JUB75 CAGAGTATGGCAATGGGTATGGGA <t< td=""><td></td><td></td><td></td><td></td><td></td><td></td></t<>						
JUB12 CTGTGCTCGTAAAAGTTGTTC 8072-8092 JUB133 CATAAGCACAAAATGTATCCC 7590-7970 JUB55 AAGAGTGTTATATGC 9043-9063 JUB64 ACCTAACACCTCCATTATATG 7522-722 JUB55 CACTTACAATGGCTÄGTTATAG 9943-9063 JUB66 ACCTAACACCTCAATGCTATAG 7522-723 JUB57 CCGTCTCAACTTCGTTATG 9925-9945 JUB66 ACCTAACACAGCATAACC 6531-651 JUB78 GCGTTCTACACTTTATG 1907-1021 JUB88 CATAACAGCGTACAGC 1617-16491 JUB71 GGCTTCTACACATTTGTAG 10901-10921 JUB8 TAACTCACCACTAACGGC 1517-1532 JUB73 ATAATGAGGTATCTGCTACTG 1244-12464 1448-1464 JUB83 CATACTCAGAAGAGCACCCG 15051-1507 JUB74 ATAATGAGGTATGGTATGGT 1304-14634 JUB84 AAAAGTGCTTACAGCAACTCG 1446-14485 JUB75 AGAGAATGAAATGGTATGGT 1406-14485 JUB86 TCCTCAGGTACTAACACCG 1348+-1368 JUB76 TTGGATTGGCAATGGTATGGTAGGAG 1445-14485 JUB86 TCCTTCAGACATGGTAAGATCAAGAGAAAGAAACACC 1234-5125 JUB70 CAT						
JUBS5 AAGGTTTTATCCCATCTTATG 9043-9063 JUBS6 CCTAAACATCCATCTTATATG 9525-924 JUBS6 CCATTACAACATGGCTATATTATG 9925-9945 JUB67 TATACTAACAAGAGTACATACC 6621-6946 JUBS7 CCGTTCTAACTTCTTATG 1978-10398 JUB67 TATACTAACAAGAGTACCC 6621-6946 JUBS7 GCGTTCTACATTTTGTTATG 1078-114-11463 JUB86 ACATTACCACGTATAGGCG 6011-6091 JUB71 GGCTTCTACAGATATACTCT 11441-11463 JUB81 TAAACTAGCGAAAGGCTACCAC 11522-1551 JUB72 CAGCAGTTAAAACAGCTAGGA 12084-12104 JUB83 TATAACTACCGGTACAAGG 1552-1551 JUB73 AGAGAGTATCTGCTACGC 13073-13093 JUB84 AAAACTACCCAATGGTACACACT 1448-1418 JUB75 TAGGATAGGCAATGGAGTAGG 1359-13579 JUB85 TCCTCAGATCCAACATGT 1364-13505 13084-1418 JUB75 TGGATGCGAATGGTATGGC 1363-13579 JUB85 TCTGCAATGGTACACCC 1248-1268 JUB77 TGGATGCGAATGGTATGGC 1540-1485 JUB87 TCCCAAAACTGCACATGGGCC 1248-1268 JUB76 CAGGATGTGGCACTGC 1486-14485	JUB172	CTGTGCTCGTAAAAGTTGTTC	8072-8092	JUB183	CATAGACAAAAATGTATCCAC	7950-7970
JUB56 CACTTACAATGGCTÃGTTATG 9540-9560 JUB66 CTTAACAAAGGTCTATACC 6926-9945 JUB58 CTGTGGAACTTCATTTGTG 10378-10398 JUB68 ACATCACCGTGTAACGC 6071-651 JUB71 GGCTTCTAAAGCATCATTTTGTA 10071-10921 JUB81 GATCTTCAAAGAGGCAAGC 16117-16137 JUB72 CAGCAGTTAAAACACGCATGAG 12948-12104 JUB83 TAAACTACCAGGAACACCACG 16117-16137 JUB73 ATAATGAGGTATCTGCTACGA 12947-12567 JUB83 TATAACTACAGGAACACCACG 15050-1507 JUB75 AGAGAATGAAATGCTATGAG 12539-13579 JUB85 ATACTCACGAGTGTAAAATGC 14188-14178 JUB76 TGGAATGCAATGCATGTGTG 14850-14876 JUB88 TCCTGCATAGTGTACACG 1294-1204 JUB77 TGGAATGCATGTAAGG 15539-13579 JUB88 TCTGCATAGTGTGTAAAATGC 14188-14178 JUB78 TCATATCATTTGCAGGAGCAGG 1488-14876 JUB88 TCTGCATAGTGTGTACAACC 1294-12648 JUB70 ATATACATTTACTGTGAGGCAAAGC 12930-12520 JUB810 ACTAGCAGTTAGAAGC 1294-1265 JUB71 ATATACATTTACTGAGGAGCA						10 10 10 00
JUB85 CTTGTGGATCTGTTGTTGTTGT 10078-10398 JUB86 ACACACCTGTAAGCTGTTGGC 6071-6091 JUB71 TGTATTTCACAGATATACCTC 11443-11463 JUB81 GATGTTTGAGAAGAGCAAGC 16117-16137 JUB73 ATAATGAGGTATCTGCTACGG 12547-12567 JUB83 TATAACTACAGGGAACACCACG 15050-1570 JUB74 TGTTAAACCCGATGTATGAG 13559-13579 JUB85 ATAACTCCGAGTGTAAAATGC 14158-14178 JUB75 AGAGAGATGGAAATGCTATGAG 13559-13579 JUB85 ATACTCCACATGTGGC 13468-13668 JUB76 TGGATTGGATTATGG 14465-14485 JUB87 TCCGTGCACACACTGGC 13245-13265 JUB70 GAGGCATGTTGTTGCCAAAGC 15230-15250 JUB88 CTTGGATAGTGTGGCAAAACCC 12448-12668 JUB71 CAAACACATTGGTTATGAGGC 16404-16000 JUB811 CTAAGACATTAGCGAAAACCC 12448-1268 JUB71 CAAACACATTGGTTAGAGGC 16404-16000 JUB811 CTAGGAAGGAAAATGC 1244-1264 JUB91 CAACACATTGGTGTATGAAGCC 16093-16323 JUB10 AAACACACTGTGTGATCAGC 1244-1264 JUB92 TATTGGAGAGGAG						
JUB7 GGCTTCTACATTTTGTTTAG 10901-10921 JUB8 TAATCTGCTCTTAAAGGCTTCC 16525-16545 JUB72 CAGCAGTTAAAACAGCTGAGG 1284-12104 JUB82 CATAATTAGCAGAAACAGC 1552-15542 JUB73 ATAATGAGGTACTGCATCAC 13073-13093 JUB84 AAAGTGGTTCAAACAGC 1509-1570 JUB76 TTGAAGGAAAATGCATAGAG 13559-1357 JUB85 ATAACTCAGGTCAAAACCC 16404-1462 JUB76 TTGGATGGCAAATGTATGAG 13559-1357 JUB85 ATAACTCCAGTGCTAAAAACCC 13648-1368 JUB76 TTGGATGGCAAATGTATAGG 15309-1570 JUB85 ATACTCCAGTGCTAAAAACCC 12848-1268 JUB76 TTGGATGGCAAATTGTATGG 14066-14086 JUB86 TCCCACAAATCATGGTGTGAAACC 12848-1268 JUB70 ATATAAGTGCCTTTCAACAGG 1530-1550 JUB810 ACTGAATGTGGTAACAGGC 1248-1248 JUB91 TATATGAGAACCAGCC 1309-1302 JUB810 ACTGGAAACTGGGTAAGTGGAAAGTG 1248-1248 JUB90 TTATGTGÄAGAGGAGCC 1309-1302 JUB810 ACTGAGAAAGCAGCCCCTATCAAGG 1248-1244 JUB91 TCAACAACTGGTATAGGAAGC </td <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>						
JUB71 TGTATTCACAGATATACCTC 11443-11463 JUB81 GATGTTTGAGAAGAGCAGAGC 16117-16137 JUB73 ATAATGAGGTATCTGCTAGAG 1284-12104 JUB83 TATAACTACAGGAAACACCACG 1552-1552 JUB73 ATAATGAGGTATCTGCTACCAC 1373-1303 JUB84 AAAAGTGCTTCAGACTACACG 1550-1537 JUB75 AGAGAGATGAAATGCTATGGG 13559-13579 JUB85 ATACTCCAGTGCTAAATCAC 1418-14178 JUB76 TTGGATGGCAATTGTGTG 14466-14485 JUB87 TCCCTGCGTACATGTGTGTGGCG 1324-1235 JUB79 GAGGCATGTTGTTCGCAAAGG 15230-15250 JUB89 TCTGGATAGTTGAAACCCC 1248-12863 JUB71 CAAAAGTTTACTGATGAGAGC 1530-1555 JUB810 ACTAGAAACCACCCCTCACAG 1248-12863 JUB711 CAAAAGTTTACTGATGAGACC 1609-16923 JUB811 ACTAAAATCATGGTATAAGCC 1208-1225 JUB93 TATGGTAGTGTGTGTAGACC 1609-16923 JUB101 AGACAAATCATGGTATAAGCC 1208-1225 JUB93 TATGGTAGTTGTGGTGTGACAC 1890-19017 JUB101 AGACCAAACCCTTAACGACC 1204-2124 JUB93 TATGGTAGTTGG						
JUB73 ATAAATGAGGTATCTGCTACCAC 1257-12567 JUB83 TATAACTAÃAGGAACACCACG 15050-1570 JUB75 AGAGAGATGAAATGCTATCAG 13073-1309 JUB84 AAAAGTGCTTCAGCACACAC 1460-14624 JUB75 AGAGAGATGAAATGCTATCAG 13559-13579 JUB85 ATACTCCCAGTGCTCAAAATAC 1446-14624 JUB77 TTGGATGATTTACGCACTTGC 14465-14485 JUB87 TCCTACACAACGTGTGTGACACACC 12345-13265 JUB79 GAGGCATGTTGTTCCCCAAAGC 1530-15550 JUB88 TCTTGCATTGAACTTGCC 1248-12485 JUB71 CAAAAGTTTACTGGTCC 16400-16600 JUB810 ACTAAACACCCCTCATCCAAC 1248-1248 JUB91 TCAACACACTGTGTTGTACAAGCC 16309-16323 JUB810 ACTAACACACTGTGTATGC 1248-1248 JUB91 TCAACACACTGGTATCAACG 1690-1600 JUB811 CTAAACTGGTAATGCC 1248-1268 JUB93 TATTGGTATTGACTAACG 1690-1623 JUB810 ACACACCCTCTAACAACC 1248-1264 JUB93 TATTGGTGTTACAACG 1690-1690 JUB10 ACACACCCTCTTATCAC 1264-2154 JUB80 TAGTGGTAATTGGTTAC						
JUB74 TGTTAAACCCGATGCTACCAC 13073-13093 JUB85 AAAAGTGCTTCAGATCACATG 14604-1462 JUB76 TTGGATTGCGAATTGTATGAG 14066-14086 JUB85 ATACTCCAGTGCTTAAAATC 14158-14178 JUB76 TTGGATTGCGAATTGTATGG 14066-14086 JUB87 TCCACAAACTTGCACAAACTG CCCACACACTGG 1445-14485 JUB78 CATTATCATTTGGGGAGCAGG 14856-14876 JUB88 TCTGCTAGTGTGTGCACAACC 1284-12868 JUB710 ATATAAGTGCTTTCAACAGG 1520-1520 JUB89 CTTGGATAGTTGAATTCC 1248-12868 JUB91 TCAACACTTGGAAGCC 16400-1660 JUB810 ACTAGAACATCATCAGGTATAGTG 1248-1266 JUB90 TCAACACATTGGATGCA 16401-4622 JUB812 AGACAAATTCATCGGTAATGC 1248-1266 JUB91 TCAACACATTGGTATTACAGC 16903-1622 JUB101 AGAGCAAACTCCTCAAACG 1248-1264 JUB92 AGTCACATAGGGTAGTATGGT 16903-1622 JUB101 AGAGCAAACTCCTCAACGC 1242-1124 JUB93 TATGGGTATCATAGGGT 16903-1622 JUB101 AGACACATTGTCACACAC 1214-2154 JUB94 TAGACTGGTATATAG						
$ \begin{array}{llllllllllllllllllllllllllllllllllll$						
JUB77 TTGTATGATTTACGCACTGC 14465–14485 JUB87 TCCACAAACTC 13245–13265 JUB79 GAGGCATGTTGTTCGCAAAGC 1530–15250 JUB89 TCTGGTATGTTGTACACC 12448–12468 JUB71 CAAAAGTTCAACAGC 1530–15556 JUB810 ACTGAAACTGCCTTACACAGC 12448–12468 JUB71 CAAAAGTTCACTGATGAGTCC 16400–16060 JUB810 ACTGAAACTGGTTAACGC 11248–1268 JUB91 TCAATCACATTGGTATGAAACC 16903–16923 JUB10 AAGACAAATCATGGTAAAGGCC 1248–12564 JUB92 AGTTCATTGTGATTCAGAACG 18034–18054 JUB10 AGACCAACTCTTACTGTGAAACTG 2104–2164 JUB93 TATGGATCAGGTTACTATGGTTG 18054–18054 JUB102 GCAACAACTCTACTACCC 2104–2163 JUB94 TAGAACTGGTTACTATGGTGT 18052–1852 JUB103 GCGGTGGACCTCTTATACCAC 2106–21084 JUB94 TAGACACGGCTTACTATGGTG 18052–1852 JUB104 GTTGTAACAGACC 1948–19518 JUB96 AGTCAAGACTGGTTACTATGGTG 18959–19017 JUB104 GTAGTAAACACCTACACAC 1844–1844 JUB98 TAGGCAGATAAGGGTTAAC	JUB75	AGAGAGATGAAATGCTATGAG	13559-13579	JUB85	ATACTCCAGTGCTTAAAATAC	14158-14178
JUB78 CATTATCATTIGAGGAGCAGG 14856-14876 JUB89 TCTTGCTAGTGTGTTACAACCC 12448-12863 JUB710 ATATAAGTGCCTTTCGAACAGG 15636-15566 JUB810 ACTAGAACACGCTCATCAAG 12448-12468 JUB711 CAAAAGTTTACTGATGAGTCC 1640-16060 JUB811 ACTAGAACACGCTCATCAAG 12448-12468 JUB91 TCAACACATTGGTTTTACAGAGC 16309-16329 JUB812 AGAACAAATCATGGGGAAATGC 1245-11256 JUB93 TATTGGTGATTTAGGTGT 17498-17518 JUB101 AAATGGGTAAGTGGAACT 1245-11264 JUB93 TATGGTGATTTTAGGTGT 18304-18054 JUB103 GCGGTGGACCTCTAATCATC 19498-1938 JUB95 TTTTGAGGCACATAAGGACTC 18397-19017 JUB103 GCAGGAGACCTCTAACGACA 19408-1938 JUB96 CTTCAAATAAGGCGTTTAC 19498-19518 JUB105 TACAAAAGACTCTAACGACA 18441-18464 JUB97 CGTCAAAGAGGTGACATAATC 20736-20756 JUB100 TTAGCACAAAGCAGGTAGCC 1789-17173 JUB98 TTACTGAGTATGTGCTCTGG 21067-21087 JUB100 TTAGCACAAAGCTAGCAGCA 1784-21714 JUB99						
JUB710 ATATAAGTGCCTTTCAACAGG 15636–15656 JUB811 ATAGAACACGCCTCATCAAG 12048–12068 JUB91 TCAACAACATTGGTGATGAAGCC 16009–1629 JUB811 CTAAAAATTAAGCATAAGGC 11245–11265 JUB91 TCAACACATTGGTATGAAACG 16903–16223 JUB812 AGAACAAATCATGGGTAAATGC 11245–11265 JUB93 TATTGGTGATCATTGGTGTTA 16903–16923 JUB10 AGAGCTAACTTGGAAATTGC 11245–11265 JUB93 TATGGGGATCTTGTGCTGTTAC 18034–18054 JUB101 AGAGCAACTCTTACTCGCATT 1206–12108 JUB95 TTTGAGGCACATAAGGACTC 18997–19017 JUB103 GCGGTGGACCTTAACTAACGAC 18997–19393 JUB96 CTGTCAAAAATGGCGTTTAC 1948–91518 JUB105 TACAAAGAGTCTAACAACC 1897–19373 JUB97 CGTTCTAAATATGGCGTTTAC 1948–19518 JUB106 CCCATGAACACACACAC 1844–18464 JUB98 TAGGCACAAGTAAGCC 2033-20333 JUB107 TTATATTGCACACAACAC 1844–18464 JUB90 TGGAGACAGTTAGTGAGC 21807–21827 JUB108 TGTATAACACACTACGC 1793-17173 JUNS06 GTGTGAAAG						
JUB911 CAAAAGTTIACTGATGAGTCC 16040-16060 JUB81 CTTAÄTATTÄAGCATAAGGCC 1149-11669 JUB91 TTATTGTGAAGATCATAAGCC 16303-16923 JUB812 AGAACAAATCATGGTATAAGCC 11245-11265 JUB92 AGTTCATTGTGTTTTAGGGTC 17498-17518 JUB10 AAATGGGTAAGTGGAAATC 2154+21564 JUB94 TAGACCGGTACTAAGGACT 18034-18054 JUB10 AGAACAACTCGTTATCACC 2041-20434 JUB94 TAGACTGGTATCAAGGACT 18034-18054 JUB103 GCGGTGGACCTCTAATCACC 2041-20434 JUB96 AGTCAAGACTGGTCATTATAC 19498-19518 JUB104 GTATGTAACGACCACACAC 19369-19389 JUB97 CGTCTGAACAGACGC 20313-20333 JUB107 TTATATTGTCAACAACCACACACC 1844-18464 JUB99 ATACCAGATTAGGCTTATCC 2136-21416 JUB109 TTAATGATAAGGACT 17153-17173 JUNS03 TTAAGAATGCGATTATCC 21396-21416 JUB109 GTAGTAAACACATCTACCACA 27754-27774 JUHE01 AACAATTCGATGCAGCG 21396-21416 JUB109 GTAGTAACACAATTACCACACC 27754-27774 JUHE02 ACAATTGATGC		GAG <u>G</u> CATGTTGTTCGCAAAGC			CTTGGATAGTTTGAATCTGCC	
$ JUB9 TATTGTGĀGĀGATCATAAGCC 16309-16229 JUB12 AGAACAAATCATGGTTAATGC 11245-11265 \\ JUB92 AGTCATTGTGTTTTAAGGGTC 17498-17518 JUB10 AAATGGGTAAGTGGAAAATC 21064-21084 \\ JUB93 TATTGGTGATTCTGCTGTTAC 18034-18054 JUB101 AGAGCTAACTTGTCTCACACT 21064-21084 \\ JUB93 TATTGGTGATTCTATGGTTA TAGGTTG 18452-18582 JUB102 GCAAACACTCTTATCATCG 19908-19328 \\ JUB95 TTTGAGGCACATAAGGACTC 18997-19017 JUB104 GTATTGTAAGACTCTAACGAC 18973-18993 \\ JUB97 CGTTCTAATAATGGCGTTAAC 19489-19518 JUB105 TACAAAAGAGTCTTAACGAC 18973-18993 \\ JUB97 CGTTCTAATAATGGCGTTTAC 19489-19518 JUB106 CCCATGTAACTAGCACACACC 18444-18464 \\ JUB99 ATACCCAGATGGTTAGCTCGG 21067-21087 JUB106 CCCATGTAACTAGCAACACC 17599-17619 \\ TCAGACAAGTTAGCTCTGGC 210756 JUB108 TTATGCCAAAGGGTTAGCC 17599-17619 \\ JUB109 TCAGAGAAGTTGGCTTGTCC 21205-20756 JUB108 TTATGCCAAAAGGGTTAGCC 17599-17619 \\ JUB100 TCGAGACAAGTTGCTCTGC 21087-21087 JUB109 GGTAGTAAACACATACTTACG 717.51717 JUNS03 TTAATGATATGGTTTATCC 21396-21416 JUB10 CTGCGGAACAAGCGTAGGAGC 16808-16828 JUNS06 GTGTGAGAAGATTGCATGCG 21087-21827 JUS02 AAGAACTTTAACAAATCTGCG 27384-27404 JUHEO1 AACAATTCTTGGTTTTCC 22258-22244 JUS04 ATCAAGTGACAAACTCAACCACG 27020-27040 JUHEO3 TTTAGGAGTTTTATCCC 2388-22608 JUS06 ATAGGACAAAGCGAAAGCGTAAGCC 2738-27404 JUHEO1 TCAGGAGAGTGTCGTCTCC 2238-23502 JUS014 GTAAACACACTACTTACCACAG 27020-27040 JUHEO5 ACCACTTTGATTATGCCC 2399-24013 JUS014 TTATAGCATAAGCACAAAGGCCATTACC 2553-2557 JUS05 AATGTATAAGCCTGGTTTAC 24752-24772 JUS014 ATAGCATAAAACCACC 24809-24829 JUS014 ATATAACCCCGATTAACC 2549-25417 JUS014 TATAATAGCAGAAAGGCCCATAACC 24480-24450 JUS015 TAGGTGTACTGTGTGTG 2438-23502 JUS014 TTATAGTAAACACCC 24480-24450 JUS015 TAGGTGTACTGTGTGGG 25467-25887 JUHEO2 ACAAATATAGCACAAAGCC 24480-24450 JUS015 TAGGTGTACTGTGTGTGG 25487-25887 JUHEO2 ACAAATAAAACACC 24809-24829 JUS016 ATTCAGGGTTACCTGTG 2438-25487 JUHEO2 ACATAATAAGCACCAAACC 24480-24450 JUS015 TAGGTGTAGTGTGTGTGTGG 25667-25887 JUHEO2 ACATAATAAACACC 2560-25805 JUS010 ATGGTAAAATAACACC 25480-24450 JUS013 TAGTGAAAGTAACCAAACC 2440-24450 JUS02 TAGAGAGGGTGTAAAATAACACC 24480-24450 $						
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	JUB9	TTATTGTGĀAGA <u>T</u> CATAAGCC		JUB812		11245-11265
JUB93 TATTGGTGATTCTGCTGTTAC 18034-18054 JUB102 GCAAAACACTCTTACCACC 20414-20434 JUB94 TAGAACTGGTTACTATGGTTG 18562-18582 JUB103 GCGGTGGACCTCTTACTACCACC 19309-19383 JUB95 AGTCAAGACTGGTCATTATAC 19498-19518 JUB105 TACAAAAAGAGTCTTAACCACAC 1844-1844 JUB98 TAGGCTTGTACCGAAGACAGC 20313-20333 JUB107 TTATATTGCACACACAC 1844-1846 JUB99 ATACTCAGTTAGCTCGGG 20167-21087 JUB108 GCTGCACACAGAGC 17933-17051 JUB100 TCGAGACAAGTTAGCTCGGG 20167-21087 JUB109 GGTACAACACACACAC 1844-18464 JUB99 ATACTCAGTTAGCTCGGG 20167-21087 JUB109 GGTACAACAACGGTAGGACC 17939-17619 JUB100 TCGAGACAATTGCTGGGC 21396-21416 JUB100 CTGCGGAACAAATCGCTGG 16808-16828 JUNS03 TTAATGATAGGTTGTTCC 2226-22244 JUS02 AGAACATTACACAAAGGGTAGC 26678-26698 JUHE01 AACAATTCTGGTTTAC 2337-23357 JUS010 GCTGCTAGACAAACACTACAC 26307-26527 JUS024 ATTAATAGCTGAGTTAC						
JUB94 TAGAACTGGTTACTATGGTTG 18562-18582 JUB103 GCGGTGGACCTCTTATCACTG 19908-19928 JUB95 TTTTGAGGCACATAAGGACTC 18997-19017 JUB104 GTATTGTAAGACTCTAAGTAC 19369-19389 JUB96 AGTCAAGACTGGTCATTATC 19498-19518 JUB105 TACAAAAGGACTCTTAACCAGCA 18473-18993 JUB97 CGTTCTAACTAGCGACACAC 2013-2033 JUB106 CCCATGTAACTAGCCAACAC 1844+18464 JUB99 ATACTCAGGTTAGTCAATATC 20736-20756 JUB108 TTATGCCACAAAGGTTAGCC 17983-18003 JUNS03 TTAATGATAGGTTTATTCC 2136-21416 JUB109 GGTAGTAAACACATACTAGCG 17153-17173 JUHE03 TTAAGGATATGGTTTTACC 2136-21416 JUB109 GTAGTAAACACATACTGGTGG 27784-27744 JUHE03 TTAAGAGATGCTGGTTTTAC 22356-22144 JUS02 AAGAACACATACCAGG 27784-27744 JUHE03 TTAGGAGATGCTGGTGTTTAC 22337-23357 JUS010 GCTGCTAGACCATAACG 27784-27740 JUHE04 ATTAATAGGATTTACCACTGC 2393-24013 JUS014 TTATAATAGGTCATAAC 25337-23557 JUS024 A						
JUB96 AGTCAAGACTGGTCATTATAC 19498-19518 JUB105 TACAAAAGAGTCTTAACCAGAC 18973-18993 JUB97 CGTTCTAATAATGGCGTTTAC 19859-19879 JUB106 CCCATGTAACTAGCACAACAC 18444-18464 JUB98 TAGGCTGTACCGAAGACAGC 20313-20333 JUB107 TTATATTTGTCATCATCCC 178441-18464 JUB99 ATACTCAGTTATGTCAATATC 20736-20756 JUB109 GGTAGTAACACACATAGCC 17899-17619 JUNS03 TTAATGGTTTATGCATGACG 21067-21087 JUB109 GGTAGTAACACACATAGCG 17599-17619 JUHE01 AACAAATCTTGGTTCTCC 22226-22274 JUS02 AAGAACTTAACAAATGGTCG 27754-27744 JUHE03 TTTAGGAGTTTGACTTACC 22382-22079 JUS04 ATCAAGGACATAAGGGCACTG 26678-26698 JUS04 ACCACTTAGCGAGTTTACC 23337-23357 JUS010 GCTGCCAAAAAGGGCACTAAAAC 2607-26327 JUS05 AATGATAGTGAGTGGTTTACC 23937-23357 JUS010 GCTGCCAAAAATGGAAAACAC 25537-25577 JUS05 AATGATAAGGGAGACTCATTAGC 24378-24398 JUS016 ACAAGTAAATAGGAGTAAAC 25537-25577 JUS05	JUB94	TAGAACTGGTTACTATGGTTG	18562-18582	JUB103	GCGGTGGACCTCTTATCATCG	19908-19928
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$						
JUB99 ATACTCAGTTATGTCAATATC 20736-20756 JUB108 TTATGCCACAAAGGGTTAGCC 17599-17619 JUB100 TCGAGACAAGTTAGCTCTGGG 21067-21087 JUB109 GGTAGTAAAACACATACTTACG 17153-17173 JUNS06 GTGTAGAAGAATTGCATGACG 21396-21416 JUB100 CTGCGGAACAAGCGAAGGGAGC 27754-27774 JUHE01 AACAATTCTGGTTCTCC 22226-22244 JUS04 ATCAAGTGACACAATCGGTGC 27784-27704 JUHE05 ACCACTTTGTATTTTAC 22382-22608 JUS06 ATCAGGTACACACACACACCACACAC 27764-27704 JUHE07 TCATGGAGATGCTGGTTTTAC 22382-22608 JUS064 ATCAAGTACACACACACACACACCACACAC 27020-27040 JUHE07 TCATGGAGATGCTGGTTTTAC 2337-23357 JUS010 GCTGCCTAGACAAATCACACAC 26307-26327 JUS05 AATGTATAGTGAGGTTCCCTGC 23993-24013 JUS014 TTATAAATAATAATAACC 25937-25957 JUS07 CTTCAAGCAGACTCATTTAC 24378-24398 JUS016 ACAAGTTAAAATAATTAGTAAC 25183-25278 JUS013 GTGTTGTGTGTGTGTGAG 25123-25143 JUS020 TAGGAAGTGAGAGCC 24489-24485 J						
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $						
JUNSO3TTAATGATATGGTTTATTCCC21396-21416JUB110CTGCGGAACAAGCGTAGGAGC16808-16828JUNSO6GTGTAGAAGAATTGCATGACG21807-21827JUSO2AAGAACTTTAACAAATGCTAG27754-27774JUHEO1AACAATTCTTGTGTTCTCC22258-22608JUSO6ATCAAGTGACAAATCTCGGTGC27384-27404JUHEO3TTTAGGAGATGCTGGTTTACC22358-22608JUSO6ATATGATTACCATTACCACAG27020-27040JUHEO5ACCACTTTGTATTTTTAACGG22959-22979JUSO8TGAATAGCATAAAGGGCATTG26678-26698JUBO24ATTAATAACCCTGATTTACC23337-23357JUSO10GCTGCCTAAAAATGGACATAACC25937-25957JUSO5AATGTATAGTGAGTTCCCTGC23993-24013JUSO12AATGGCTCAAAATTAGTAACC25937-25957JUSO5AATGTATAGTGAGTTCCTGC2393-24138JUSO14TTATAATAAGTGCCATTAACC25188-2288JUSO7CTTTCACACTATTATGTCATG24378-24398JUSO16ACAAGTTAAATAATTAGTAACC24809-24829JUSO11AATGGATGGTTCGTGTGGGG25123-25143JUSO20TAGAAGTGAGAGGGTGTAACCC24436-24456JUSO13GTGTTGGTTAATTATGACC25488-25887JUHEO2ACATAATAAGTACCCAAAACC24031-24051JUSO17GAATGGGTGTACTCTTAGCAC26632-26652JUHEO4CATTATCATAACACACAGGC2307-23057JUSO17GAATGGTGTAACTCAAAATC26632-26652JUHEO6AGACCATAAATAACACCCCAGTG23037-23057JUSO17GAATGGTGTACTCTTAGCAC26632-26652JUHEO6AGACCATAAATAACACCCCAGTG23037-23057JUSO17GAATGGTGTACTCTTAGCAC26492-2669JUHEO6AGACCATAAATAACCCC						
JUHE01AACAATTCTTGGTTCTTCC22226-22244JUSO4ATCAAGTGACAAATCTGGTGC27384-27404JUHE03TTTAGGAGTTTTCACTTTACC22388-22608JUSO6ATATGATTACCATTACCACAG27020-27040JUHE07TCATGGAGATGCTGGTTTTAC22388-22608JUSO6ATATAGATAACGCATAAAGGGCATG26678-26698JUHE07TCATGGAGATGCTGGTTTTAC23337-23357JUSO10GCTGCCTAGACAACCTAATAC26307-26327JUSO24ATTAATAACCCTGATTTACCC23482-23012JUSO12AATGGCTCAAAATTAGTAAAC25937-25957JUSO5AATGTATAGTGAGTTCCCTGC2393-24013JUSO14TTATAAAAGTCGCATTAACC2553-25573JUSO7CTTTCACACATATTATGTCATG24752-24772JUSO16ACAAGTTAAATAATTAGTACC25188-25208JUSO11AATTGAATGGTTGTGTGAG25123-25143JUSO20TAGAAGTGAGAGGGTGTAACCC24436-24456JUSO15TAGGTAGTGGGTTACTCTGTGTGG25867-2587JUHE02ACATAATAAGTACCCAAAAC24031-24051JUSO17GAATGGTGTAACTCTAGCAC2632-26652JUHE04CATTATCATACAAACC2408-23428JUSO19TGGATGTGGCTAAGTCAAAATC26632-26652JUHE04CATTATAAAAACACCAGTG2307-23057JUS021ATTTCTGTGGTAAGTCAAAATC26632-26652JUHE06AGACCATAAATAACACCAGTG23037-23057JUS025TGGCACCAGATTTGTCAAGTCAAATC27649-27669JUNS02GAAACAACATTGGTAAGGAGGG22416-22436JUM03TAGTGCATTATTATCGGATACC27649-27669JUNS04TTCCTTATGGAAGGAGGGG22416-22436JUM04TACCACACACACAGTTGTGTATTTGTG28662-28682JUO4TACCAAAACAC						
JUHEO3TTTAGGAGTTTTCACTTTACC22588–22608JUSO6ATATGATTACCATTACCATACCAGG27020–27040JUHEO5ACCACTTTGTATTTTTAACGG22959–22979JUSO8TGAATAGCATAAAGGGCATTG26678–26698JUHEO7TCATGGAGATGCTGGTTTAC2337–2337JUSO10GCTGCCTAGACAACCTAAATAC26678–26698JUSO24ATTAATAAACCCTGATTTACC2337–2357JUSO10GCTGCCTAGACAACTAAATAC25937–25957JUSO5AATGTATAGTGAGTTCCCTGC2393–24013JUSO14TTATAAAAGTCGCATTAACC2553–25573JUSO7CTTTCACACTATTATGTCATG24378–24398JUSO16ACAAGTTAAATAATTAGTACC25188–25208JUSO9TATTCAGGCAGACTCATTTAC24752–24772JUSO18ATAGTATGGAGAGGTGTAACC24436–24456JUSO13GTGTTTGTGTTAATGTATGGAG25123–25183JUSO20TAGAAGTGAGAGGTGTAACC24436–24456JUSO15TAGGTAGTGGTTACTGTGTGG25867–25887JUHEO2ACATAATAAGTACCCAAAACC23778–23797JUSO19TGGATGTGCTAAGTCAAAATC26632–26652JUHEO4CATTATCATACCTAAAAACCC23408–24428JUSO21ATTCCTGTGGGAAATCGAAATC27016–27036JUHEO8ATGATAAGGCGTAAAATAACCCAGTG2303–23057JUSO25TGGCACCAGATTTATGGAACT27016–27036JUHEO8ATGATAAGGCGTAAAATTAAC22660–22680JUMO3TAGTGGATATAACTGGATACC27649–27669JUNSO4TTCCTAAAGACACAGTGGCGC21917–230514JUMO5ATGTGGGATTATACTGGATAGC28662–28682JUO4TACCAAACACTGCTGGAACAG29920–29940JUO8ATGTCTTTAACTGGGTAAGC29079–29100JUO6ATACCATCACAACA						
JUHEO5ACCACTTTGTATTTTTAACGG22959–22979JUS08TGAATAGCATAAAGGGCATTG26678–26698JUHEO7TCATGGAGATGCTGGTTTACC23337–23357JUS010GCTGCCTAGACAACCTAATAC2307–26327JUS024ATTAATAACCCTGATTTACCC23482–23502JUS012AATGGCTCAAAATTAGTAAAC25937–25957JUS05AATGTATÄGTGAGTTCCCTGC23993–24013JUS014TTATAATAAGTCGCATTAACC25553–25573JUS07CTTTCACACTATTATGTCATG24378–24398JUS016ACAAGTTAAATAATTAGTACC25188–25208JUS09TATTCAGGCAGACTCATTTAC24752–24772JUS018ATAGTTATGCTGGAAAAACAC24809–24829JUS011AATTGAATGGTTCGTGTGTGA25123–2518JUS020TAGAAGTGAGAGGTGTAACCC24436–24456JUS015TAGGTAGTGGTTACTGTGTGGG25867–25887JUHEO2ACATAATAAGTAACCAAAACC2307–23057JUS017GAATGGTGCTAAGTCAAAATC26632–26652JUHEO4CATTATCATACCTAAAAACACC23408–23428JUS019TGGATGTGCTAAGTCAAAATC26632–26652JUHEO6AGACCATAAATAACACCAGTG23037–23057JUS021ATTTCTGTGGGAAGCAAAATC26632–26652JUHEO6AGACCATAAATAACACCAGTG23037–23057JUS021ATTTCTGTGGGAAGCAAAATC2738–27402JUNS02GAAACAACATGGTAAAATAACACCAGTG22037–23057JUS021ATGTCATTATACTGGATAACC27649–27669JUNS04TTCCTAATGGACAGTGCTGC21917–21937JUMO3TAGTGGATTATACTGGTAAGC27649–27669JUNS04TTCCTAATGGACAGTGCTGC21917–21937JUMO3TAGTGGATTGTTATTGTG28662–28682JUO4TACCAAAACACA						
JÚSO24ATTAÁTAÁCCCTGÁTTTACCC23482–23502JÚSÓ12ÁATGGCTCÁAÁATTÁGTAAAC25937–25957JUSO5AATGTATÁGTGAGTTCCCTGC23993–24013JUSO14TTATAATAAGTCGCATTAACC25553–25573JUSO7CTTTCACACTATTATGTCATG24378–24398JUSO16ACAAGTTAAATAATTAGTAACC25188–25208JUSO9TATTCAGGCAGACTCATTTAC24752–24772JUSO18ATAGTTATGCTGGAAAAAACC24409–24829JUSO11AATGAATGGTTCGTGTGAG25123–25143JUSO20TAGAAGTGAGAGGAGGAGACTCAATTAGCC24436–24456JUSO13GTGTTTGTGTTAATTATGACC25498–25518JUSO22TATAGGAAGTGGAGAGGTGTAACCC24401–24051JUSO17GAATGGTGTTACTGTGTGG25867–25887JUHEO2ACATAATAAGTACCCAAACC23408–23428JUSO19TGGATGTGCTAAGTCAAAATC26032–26652JUHEO4CATTATCATAACTAACACAGGC23408–23428JUSO20AATGATAAGGCGTAAATCACTAGGTAAATC26032–26652JUHEO6AGACCATAAATAACACCAGTG2307–23057JUSO21ATTTCTGTGGGTAATCCACTTG27382–27402JUNSO2GAAACAACATTGGTAAGGCGTAAAATTAAC22660–22680JUNO1GTGATGATTATACTGGATACC27649–27669JUNSO4TTCCTTAATGGACAGAGCGCGCC21917–21937JUMO3TAGTGCCATTTGTTATTGG28662–28682JUO4TACCAAAACACTGGCGAGAGG29920–29940JU08ATGTCTTTACTGGGAAGC29010–29100JUO6ATACCATCGTGGCAGCAGTG29940–29450JUO3TCTACTGGGTGCCCATTCTG29999–30019JUMO4TACACAATCCACATAATAATAG28655–28675	JUHEO5	ACCACTTTGTATTTTTAACGG	22959-22979	JUSO8	TGAATAGCATAAAGGGCATTG	26678-26698
JUSO5AATGTATAGTGAGTTCCCTGC23993-24013JUSO14TTATAATAAGTCGCATTAACC25553-25573JUSO7CTTTCACACTATTATGTCATG24378-24398JUSO16ACAAGTTAAATAATTAGTACC25188-25208JUSO9TATTCAGGCAGACTCATTTAC24752-24772JUSO18ATAGTTATGCTGGGAAAAACAC24409-24829JUSO11AATTGAATGGTTCGTGTGAG25123-25143JUSO20TAGAAGTGAGAGGGGTAAACCC24436-24456JUSO15TAGGTAGTGGTTACTGTGTGG25867-25887JUBO22TATAGGATGTATTTAACAAAAC24031-24051JUSO17GAATGGTGTTACTCTTAGCAC26234-26254JUHEO2ACATAATAAGTACCCAAACC23408-23428JUSO19TGGATGTGCTAAGTCAAAATC26632-26652JUHEO4CATTATCATACCTAAAAACACC23408-23428JUSO21ATTTCTGTGGGTAATGGTAATC27036-27036JUHEO4CATTATCATACTAGCTAAAAACACC2260-22680JUSO25TGGCACCAGATTTGTCACTTG27382-27402JUNSO2GAAACAACATTGGTAGCGGC22416-22436JUMO1GTGATGGCATTGGTATCC27649-27669JUNSO4TTCCTTAATGGAAGGGGCC21917-21937JUMO3TAGTGGCATTGTGTAATTTGTG28662-28682JUO4TACCAAAACACTGCTGC21917-21937JUMO5ATGTGGATTGTGTAAACC29512-29100JUO6ATACCATCGTGGCAGCAGTGC29920-29440JUO3TCTACTGGGGTGCCCATTTGTCACCC29512-29532JUMO4TACCAAAACACTGCGGCAGCAGTG29920-29450JUO5CCCACAGTTCCCCATTCTTGC29999-30019JUMO4TACACAATCCACATAATAAC28655-28675						
JUSO9TATTCAGGCAGACTCATTTAC24752-24772JUSO18ATAGTTATGCTGGAAAAACAC24809-24829JUS011AATTGAATGGTTCGTGTGTAG25123-25143JUSO20TAGAAGTGAGAGGGGTGTAACCC24436-24456JUS013GTGTTGTGTTAATTATGACC2548-25518JUSO22TATAGGATGTATTTACAAAAG24031-24051JUS015TAGGTAGTGTTACTGTGGGC25867-25887JUHEO2ACATAATAAGTACCCAAACC2378-23797JUS017GAATGGTGTTACTCTTAGCAC26234-26254JUHEO4CATTATCATACCTAAAAACG2408-23428JUS019TGGATGTGCTAAGGTAATC26632-26652JUHEO4CATTATAGGGTAATACCCAAGTG23037-23057JUS021ATTTCTGTGGGTAATGGTAATC27016-27036JUHEO8ATGATAAGGCGTAAAAATAAC2660-22680JUS025TGGCACCAGATTTGTCACTTG27382-27402JUNS02GAAACAACATTGGTAAGGAGGG22416-22436JUM03TAGTGGCATTATATCTGGATACC27649-27669JUNS04TTCCTTAATGGAAGGGTGACC21917-21937JUM05ATGTGGATTGTGTAATTGTGTAAGC28662-28682JUO4TACCAAAACACTGCTGAACAG29920-29940JU08ATGTCTTTTACTCCGGGTAAGC29919-29100JUO6ATACCATCATGCGGAGCAGTTG29430-29450JU03TCTACTGGGTCGCTAGTAACC29512-29532JUMO4TACCACAAACACTCCATTACC29136-29156JUO5CCCACAGTTCCCCATTTCTTG2999-30019JUMO4TACACAATCCACAATAATAAC24655-28675						
JUS011AATTGAATGGTTCGTGTGTAG25123–25143JUS020TAGAAGTGAGAGGGTGTAACCC24436–24456JUS013GTGTTTGTGTTAATTATGACC25498–25518JUS022TATAGGATGTATTTACAAAAG24031–24051JUS015TAGGTAGTGGTTACTCTTAGCAC26234–26254JUHEO2ACATAATAAGTACCCAAAACC23778–23797JUS017GAATGGTGTTACTCTTAGCAC26234–26254JUHEO4CATTATCATACCTAAAAACC23408–23428JUS019TGGATGTGCTAAGGTAATC26632–26652JUHEO4AGACCATAAATAACACCCAGTG23037–23057JUS021ATTTCTGTGGGTAATGGTAATC27016–27036JUHEO8ATGATAAGGCGTAAAAATCACC26GG22416–22436JUS025TGGCACCAGATTTGTCACTTG27322–27402JUNS02GAAACAACATTGGTAGGTAGGG22416–22436JUM01GTGATGATATAACTGGATACC27649–27669JUNS04TTCCTTAATGGACAGTGGTGGC21917–21937JUM03TAGTGGATTGTGTATTTGTG28662–28682JUO4TACCAAAAACACTGCTGAACAG29920–29940JU08ATGTCCTTTAATCGGTAACC29512–29532JUMO2CCACTTGGGAACAGC29920–29400JU03TCTACTGGGTCGCTAGTAACC29512–29532JUMO2CCACTTGAGGATGCCATTACC29136–29156JUO5CCCACAGTTCCCCATTCTTG2999–30019JUM04TACCACAATCCACATAAATAAC2403–29450						
JUS013GTGTTTGTGTTAATTATGACC25498–25518JUS022TATAGGATGTATTTACAAAAG24031–24051JUS015TAGGTAGTGGTTACTGTGTGG25867–25887JUHEO2ACATAATAAGTACCCAAACC23778–23797JUS017GAATGGTGTTACTCTTAGCAC26234–26254JUHEO2ACATAATAAGTACCCAAACC23408–23428JUS019TGGATGTGCTAAGTCAAAATC26632–26652JUHEO6AGACCATAAATAACACCAGTG23037–23057JUS021ATTTCTGTGGTAATGGTAATC27016–27036JUHEO8ATGATAAGGCACAAAATAAAAATC22660–22680JUS025TGGCACCAGATTTGTCACTTG27382–27402JUNS02GAAACAACATTGGTAAGGAGGG22416–22436JUMO1GTGATGATTATACTGGATACC27649–27669JUNS04TTCCTTAATGGACAGTGCTGC21917–21937JUMO3TAGTTGCCATTTGTTATTGG28169–28189JUO2GCAGCAAGACATCCATTCG30495–30514JUM05ATGTGGATTGTGTAATC2960-29100JUO6ATACCATCGTGGCAGCAGTG29920–29940JU08ATGTCTTTACTCCTGGTAAGC29912–29532JUMO2CCACTTGAGGATGCCATTACC29136–29156JUO5CCCACAGTTCCCCATTCTTGC29999–30019JUMO4TACACAATCCACATAAAATAACG28655–28675						
JUS017 GAATGGTGTTACTCTTAGCAC 26234–26254 JUHEO4 CATTATCATACCTAAAAACGC 23408–23428 JUS019 TGGATGTGCTAAGTCAAAATC 26632–26652 JUHEO6 AGACCATAAATAACACCAGTG 23037–23057 JUS021 ATTTCTGTGGTAATGGTAATC 27016–27036 JUHEO8 ATGATAAGGCGTAAAAATAAC 22660–22680 JUS025 TGGCACCAGATTTGTCACTTG 27382–27402 JUNS02 GAAACAACATTGGTAGGGGGC 22416–22436 JUMO1 GTGATGATTATACTGGATACC 27649–27669 JUNS04 TTCCTTAATGGACAGTGCTGC 21917–21937 JUMO3 TAGTTGCATTTGTGTATTTGG 28662–28682 JUO2 GCAGCAAGACATCCATTCG 30495–30514 JUM05 ATGTCGTTTACTCGGTAAGC 29912–29100 JUO6 ATACCATCGTGGCAGCAGTG 29920–29940 JU03 TCTACTGGGTCGCTAGTAACC 29512–29532 JUMO2 CCACTTGAGGATGCCATTACC 29136–29156 JUO5 CCCACAGTTCCCCATTCTTG 29999–30019 JUMO4 TACCACAATCCACATAAATAAC 29136–292156	JUSO13	GTGTTTGTGTTAATTATGACC	25498-25518	JUSO22	TATAGGATGTATTTACAAAAG	24031-24051
JUSO19 TGGATGTGCTAAGTCAAAATC 26632–26652 JUHEO6 AGACCATAAATAACACCAGTG 23037–23057 JUSO21 ATTTCTGTGGGTAATGGTAATC 27016–27036 JUHEO8 ATGATAAGGCGTAAAAATAAC 22660–22680 JUSO25 TGGCACCAGATTTGTCACTTG 27382–27402 JUNSO2 GAAACAACATTGGTAAGGAGGG 22416–22436 JUMO1 GTGATGATATACTGGATACC 27649–27669 JUNSO2 GAACAACATTGGGACGGTGGCC 21917–21937 JUMO3 TAGTTGCCATTTGTTTATTGG 28662–28682 JUO4 TACCAAAACACTGCTGAACAG 29920–29940 JU08 ATGTCTTTTACTCCTGGTAAGC 29019–29100 JUO6 ATACCAACACTGCGGCAGCAGTTG 29430–29450 JUO3 TCTACTGGGTCGCTAGTAACC 29512–29532 JUMO2 CCACTTGAGATGCCATTACC 29136–29156 JUO5 CCCACAGTTCCCCATTCTTG 29999–30019 JUMO4 TACCACAATCCACATAAATAACG 29136–29156						
JUSO21 ATTTCTGTGGTAATGGTAATC 27016–27036 JUHEO8 ATGATAAGGCGTAAAGGCGTAAAATTAAC 22660–22680 JUSO25 TGGCACCAGATTTGTCACTTG 27382–27402 JUNSO2 GAAACAACATTGGTAGGAGGG 22416–22436 JUMO1 GTGATGATTATACTGGATACC 27649–27669 JUNSO4 TTCCTTAATGGACAGTGCTGC 21917–21937 JUMO3 TAGTTGCCATTTGTTATTGG 28169–28189 JUO2 GCAGCAAGACATCCATTCG 30495–30514 JUMO5 ATGTGGATTGTGTATTTGTG 28662–28682 JUO4 TACCAAAACACTGCTGAACAG 29920–29940 JUO8 ATGTCTTTTACTGGTAAGC 29079–29100 JUO6 ATACCATCGTGGCAGCAGTTG 29430–29450 JUO3 TCTACTGGGTCGCTAGTAACC 29512–29532 JUMO2 CCACTTGAGGATGCCATTACC 29136–29156 JUO5 CCCACAGTTCCCCATTCTTGC 29999–30019 JUMO4 TACACAATCCACAATAATAATG 28655–28675						
JUMO1 GTGATGATTATACTGGATACC 27649–27669 JUNSO4 TTCCTTAATGGACAGTGCTGC 21917–21937 JUMO3 TAGTTGCCATTTGTTTATTGG 28169–28189 JUO2 GCAGCAAGACATCCATTCTG 30495–30514 JUMO5 ATGTGGATTGTGTATTTGTG 28662–28682 JUO2 GCAGCAAGACATCCATTCTG 30495–30514 JU08 ATGTCTTTACTCCTGGTAAGC 2907–29100 JUO6 ATACCATCGTGGCAGCAGTG 29920–29940 JUO3 TCTACTGGGTCGCTAGTAACC 29512–29532 JUMO2 CCACATGGGATGCCATTACC 29136–29156 JUO5 CCCACAGTTCCCCATTCTTGC 29999–30019 JUMO4 TACACAATCCACAATAATAATG 28655–28675	JUSO21	ATTTCTGTGGTAATGGTAATC	27016-27036	JUHEO8	ATGATAAGGCGTAAAATTAAC	22660-22680
JUMO3 TAGTTGCCATTTGTTTATTGG 28169–28189 JUO2 GCAGCAAGACATCCATTCTG 30495–30514 JUMO5 ATGTGGATTGTGTATTTTGTG 28662–28682 JUO4 TACCAAAACACTGCTGAACAG 29920–29940 JU08 ATGTCTTTACTCCTGGTAAGC 2907–29100 JUO6 ATACCATCGTGGCAGCAGTGC 29430–29450 JUO3 TCTACTGGGTCGCTAGTAACC 29512–29532 JUMO2 CCACATGGGGCAGCAGTACC 29136–29156 JUO5 CCCACAGTTCCCCATTCTTGC 29999–30019 JUMO4 TACACAATCCACATAATAATG 28655–28675		TGGCACCAGATTTGTCACTTG GTGATGATTATACTGGATACC				
JUMO5 ATGTGGATTGTGTATTTTGTG 28662–28682 JUO4 TACCAAAACACTGCTGAACAG 29920–29940 JUO8 ATGTCTTTTACTCCTGGTAAGC 29079–29100 JUO6 ATACCATCGTGGCAGCAGCTG 29430–29450 JUO3 TCTACTGGGTCGCTAGTAACC 29512–29532 JUMO2 CCACTTGAGGATGCCATTACC 29136–29156 JUO5 CCCACAGTTCCCCATTCTTGC 29999–30019 JUMO4 TACACAATCCACATAATAATG 28655–28675						
JUO3 TCTACTGGGTCGCTAGTAACC 29512–29532 JUMO2 CCACTTGAGGATGCCATTACC 29136–29156 JUO5 CCCACAGTTCCCCATTCTTGC 29999–30019 JUMO4 TACACAATCCACATAATAATG 28655–28675	JUMO5	ATGTGGATTGTGTATTTTGTG	28662-28682		TACCAAAACACTGCTGAACAG	
JUO5 CCCACAGTTCCCCATTCTTGC 29999–30019 JUMO4 TACACAATCCACATAATAATG 28655–28675						
JUO7 CTCTCTATCAGAATGGATGTC 30487–30507 JUMO6 TATAAAAATTATTTGCCCCAC 28150–28170	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.

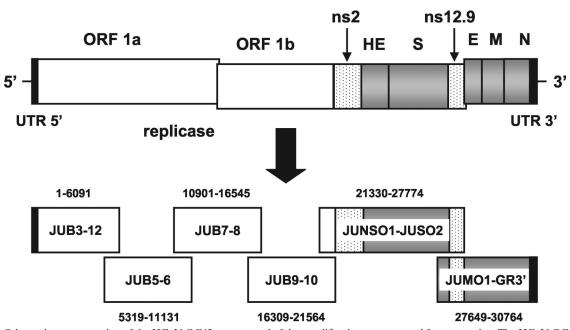


FIG. 1. Schematic representation of the HCoV-OC43 genome and of the amplification strategy used for sequencing. The HCoV-OC43 genome is 30,713 nt long and comprises nine main ORFs: ORF1a, ORF1b, ns2 (the gene encoding the nonstructural protein 2), HE (hemagglutininesterase gene), S (spike gene), ns12.9 (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

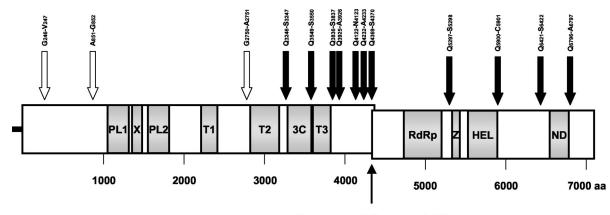
Genome region	Location (nt)	TRS location (nt)	TRS sequence ^a
Leader and 5'UTR	1-209	63–69	UCUAAAC139 ntAUG
ORF1a	210-13361		
$ORF1b^b$	13361-21496		
Intergenic region	21497-21505		
ns2 gene	21506-22342	21492-21498	UCUAAACUUUAAAAAUG
Intergenic region	22343-22353		
HE gene	22354-23628	22339-22344	UUAAAC UCAGUGAAAAUG
Intergenic region	23629-23642		
S gene	23643-27704	23636-23642	UCUAAACAUG
Intergenic region	27705-27791		
ns12.9 gene	27792-28121	27771-27777	UCUUAAGGCCACGCCCUAUUAAUG
$E gene^{\tilde{c}}$	28108-28362		
Intergenic region	28363-28376		
M gene	28377-29069	28367-28373	UCCAAACAUUAUG
Intergenic region	29070-29078		
N gene	29079-30425	29065-29070	UCUAAAUUUUUAAGGAUG
3'ŬTR	30426-30713		
Poly(A) tail of 28 nt	30714-30741		

TABLE 3. Organization of the HCoV-OC43 genome

^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.



ribosomal frameshift

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 \rightarrow 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 $10^{1.5}$ TCID₅₀/g due to a lung extract toxicity on HRT-18 cells) and only when brain titers reached at least 10⁴ TCID₅₀/g (data not shown). HCoV-OC43 ATCC infectious virus could be detected in mouse CNS, as early as 2

 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 (Åla81)
30632	3'ŬTR	C→A

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.

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-

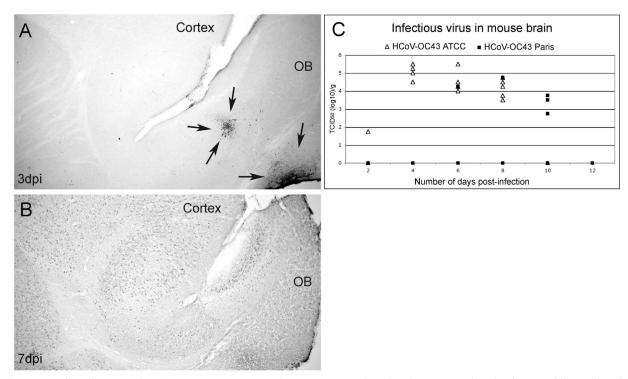


FIG. 3. Neuroinvasive properties of HCoV-OC43 ATCC and HCoV-OC43 Paris variant in BALB/c mice after intranasal inoculation. (A) At 3 dpi, cells positive for viral antigens (arrows) were first observed in the olfactory bulb (OB). No infected cells could be detected in the cortex or other brain structures, 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), $\times 32$. (C) Quantification of infectious virus in the brain of each mouse at different times postinfection. Virus titers are presented as logarithmic value of TCID₅₀ per gram of tissue (the limit of detection was 10^{0.5} TCID₅₀/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 at 6 dpi. Infectious particles were detected in some of the brains up to 10 dpi. 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-

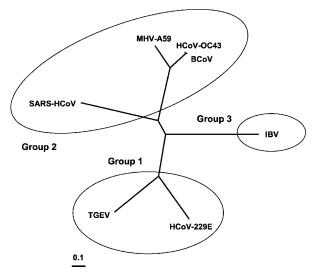


FIG. 4. Phylogenic unrooted tree regrouping seven coronavirus complete genomes from the three genetic groups. Circles regroup members of each three genetic groups. The 0.1 sliding bar represents the genetic distance between the species (i.e., nucleotide substitution units per studied site). Strains: 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).

OC43 BCoV MHV-A59 SARS Tor2 229E IBV	SGIVKMVNPT SGIVKMVSPT SGFRKMAFPS AGLRKMAQPS	SKVEPCIVSV SKVEPCIVSV GKVEGCMVQV GFVEKCVVRV	TYGNMTLNGL TYGNMTLNGL TCGTTTLNGL CYGNTVLNGL	WLDDKVYCPR WLDDKVYCPR WLDDTVYCPR WLGDIVYCPR	HVICSASDMT HVICSSADMT HVICTAEDML HVIAS-NTTS	NPDYTNLLCR DPDYPNLLCR NPNYEDLLIR AIDYDHEYSI	VTSSDFTVLF VTSSDFTVLF VTSSDFCVMS KSNHSFLVQA MRLHNFSIIS ANNHEFEVTT	DR-LSLTVMS GR-MSLTVMS GN-VQLRVIG GT-AFLGVVG
OC43 BCoV MHV-A59 SARS Tor2 229E IBV	YQMQGCMLVL YQMQGCQLVL HSMQNCLLRL ATMHGVTLKI	TVTLQNSRTP TVTLQNPNTP KVDTSNPKTP KVSQTNMHTP	KYTFGVVKPG KYSFGVVKPG KYKFVRIQPG RHSFRTLKSG	ETFTVLAAYN ETFTVLAAYN QTFSVLACYN EGFNILACYD	GKPQGAFHVT GRPQGAFHVT GSPSGVYQCA GCAQGVFGVN	MRSSYTIKGS LRSSHTIKGS MRPNHTIKGS MRTNWTIRGS	FLCGSCGSVG FLCGSCGSVG FLCGSCGSVG FLNGSCGSVG FINGACGSPG FLAGACGSVG	YVLMGDCVKF YVLTGDSVRF FNIDYDCVSF YNLKNGEVEF
OC43 BCoV MHV-A59 SARS Tor2 229E IBV	VYMHQLELST VYMHQLELST CYMHHMELPT VYMHQIELGS	GCHTGTDFNG GCHTGTDFSG GVHAGTDLEG GSHVGSSFDG	DFYGPYKDAQ NFYGPYRDAQ KFYGPFVDRQ VMYGGFEDQP	VVQLPVQDYI VVQLPVQDYT TAQAAGTDTT NLQVESANQM	QSVNFVAWLY QTVNVVAWLY ITLNVLAWLY LTVNVVAFLY	AAILNNCN AAIFNRCN AAVINGDR AAILNGCT	WFVQSD WFVQSD WFVQSD WFLNRF WWLKGE FSLPKWLEST	KCSVEDFNVW SCSLEEFNVW TTTLNDFNLV KLFVEHYNEW
OC43 BCoV MHV-A59 SARS Tor2 229E IBV	ALSNGFSQVK AMTNGFSSIK AMKYNYEPLT AQANGFTAMN	SDLV IDAL ADLV LDAL QDHVDILGPL GEDA FSIL	ASMTGVSLET ASMTGVTVEQ SAQTGIAVLD AAKTGVCVER	LLAAIKRLK- VLAAIKRLH- MCAALKELLQ LLHAIQVLN-	NGFQGRQIMG SGFQGKQILG NGMNGRTILG NGFGGKQILG	SCSFEDELTP SCVLEDELTP STILEDEFTP YSSLNDEFSI	SDVYQQLAGV FDVVRQCSGV	KLQ KLQ TFQ NLQ

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, His⁴¹ and Cys¹⁴⁵, 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 His⁴¹ and a Cys¹⁴⁵ (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, aerosolaware, 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

- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Arbour, N., R. Day, J. Newcombe, and P. J. Talbot. 2000. Neuroinvasion by human respiratory coronaviruses. J. Virol. 74:8913–8921.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- Gorbalenya, A. E. 2001. Big nidovirus genome. When count and order of domains matter. Adv. Exp. Med. Biol. 494:1–17.
- Hegyi, A., and J. Ziebuhr. 2002. Conservation of substrate specificities among coronavirus main proteases. J. Gen. Virol. 83:595–599.
- 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.
- Hung, M., C. S. Gibbs, and M. Tsiang. 2002. Biochemical characterization of rhinovirus RNA-dependent RNA polymerase. Antivir. Res. 56:99–114.
- 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.

- 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.
- Kündel, F., and G. Herrler. 1993. Structural and functional analysis of the surface protein of human coronavirus OC43. Virology 195:195–202.
- Kündel, F., and G. Herrler. 1996. Structural and functional analysis of the S
 proteins of two human coronavirus OC43 strains adapted to growth in
 different cells. Arch. Virol. 141:1123–1131.
- 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.
- 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.
- Lai, M. M. C., and D. Cavanagh. 1997. The molecular biology of coronaviruses. Adv. Virus Res. 48:1–97.
- 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.
- 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.
- 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.
- Masters, P. S. 1999. Reverse genetics of the largest RNA viruses. Adv. Virus Res. 53:245–264.
- 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.
- 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.
- Mounir, S., and P. J. Talbot. 1993. Molecular characterization of the S protein gene of human coronavirus OC43. J. Gen. Virol. 74:1981–1987.
- 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.
- 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.
- Myint, S. H. 1994. Human coronaviruses: a brief review. Rev. Med. Virol. 4:35–46.
- Novella, I., S. 2003. Contributions of vesicular stomatitis virus to the understanding of RNA virus evolution. Curr. Opin. Microbiol. 6:399–405.
- 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.
- 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.

- Perlman, S., G. Evans, and A. Affi. 1990. Effect of olfactory bulb ablation on spread of a neurotropic coronavirus into the mouse brain. J. Exp. Med. 172:1127–1132.
- 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.
- Sawicki, S. G., and D. L. Sawicki. 1998. A new model for coronavirus transcription. Adv. Exp. Med. Biol. 440:215–219.
- 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.
- 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.
- 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.
- Taguchi, F. 1993. Fusion formation by the uncleaved spike protein of murine coronavirus JHMV variant cl-2. J. Virol. 67:1195–1202.
- Taguchi, F. 1995. The S2 subunit of the murine coronavirus spike protein is not involved in receptor binding. J. Virol. 69:7260–7263.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.
- 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.