Available online at www.sciencedirect.com

SCIENCE DIRECT

Virology 337 (2005) 85 – 92

VIROLOGY

www.elsevier.com/locate/yviro

Circulation of genetically distinct contemporary human coronavirus OC43 strains

Leen Vijgen, Els Keyaerts, Philippe Lemey, Elien Moës, Sandra Li, Anne-Mieke Vandamme, Marc Van Ranst*

Laboratory of Clinical and Epidemiological Virology, Department of Microbiology and Immunology, Rega Institute for Medical Research, University of Leuven, Minderbroedersstraat 10, B-3000 Leuven, Belgium

Received 8 March 2005; returned to author for revision 5 April 2005; accepted 7 April 2005

Abstract

In this study, we report the complete genome sequence of two contemporary human coronavirus OC43 (HCoV-OC43) strains detected in 2003 and 2004, respectively. Comparative genetic analyses of the circulating strains and the prototype HCoV-OC43 strain (ATCC VR759) were performed. Remarkably, a lower than expected similarity is found between the complete genomes and more in particular between the spike genes of the BE03 and BE04 strains. This finding suggests the existence of two genetically distinct HCoV-OC43 strains, circulating in Belgium in 2003 and 2004. Spike gene sequencing of three additional 2003 and two additional 2004 HCoV-OC43 strains, and subsequent phylogenetic analysis confirm this assumption. Compared to the ATCC prototype HCoV-OC43 strain, an important amino acid substitution is present in the potential cleavage site sequence of the spike protein of all contemporary strains, restoring the N-RRXRR-C motif, associated with increased spike protein cleavability in bovine coronaviruses. We here describe specific characteristics associated with circulating HCoV-OC43 strains, and we provide substantial evidence for the genetic variability of HCoV-OC43.

Keywords: Human coronavirus OC43; Contemporary strain; Genetic variability; Phylogenetic analysis

Introduction

Coronaviruses (family *Coronaviridae*, order *Nidovirales*) are large, roughly spherical particles with a linear, non-segmented, capped, and polyadenylated positive-sense single-stranded RNA genome (Cavanagh, 1997; Lai and Holmes, 2001). The virions contain four major structural proteins: the nucleocapsid (N) protein, the membrane (M) glycoprotein, the spike (S) glycoprotein and the small membrane or envelope (E) protein. A hemagglutinin-esterase (HE) glycoprotein gene is only present in group 2 coronaviruses, which include human coronavirus OC43 (HCoV-OC43), bovine coronavirus (BCoV), porcine hemagglutinating encephalomyelitis virus (PHEV), canine respiratory coronavirus (CRCoV), mouse hepatitis virus (MHV), rat sialodacryoadenitis virus (SDAV) and equine coronavirus (ECoV) (Spaan et al., 1988; Zhang et al., 1992).

Human coronaviruses (HCoV) cause respiratory infections but also gastroenteritis and neurological disorders can occur (Arbour et al., 2000; Lai and Holmes, 2001). Until now, five types of human coronaviruses have been described: HCoV-OC43, HCoV-229E, HCoV-NL63, the recently characterized HCoV-HKU1 and the causal agent of the SARS outbreak, the SARS-coronavirus (SARS-CoV). HCoV-OC43 (ICTVdb code 19.0.1.0.006) and HCoV-229E (ICTVdb code 19.0.1.0.005) were isolated in 1967 from volunteers at the Common Cold Unit in Salisbury, UK. HCoV-OC43 was initially propagated on ciliated human embryonic tracheal and nasal organ cultures (OC) (McIntosh et al., 1967). HCoV-OC43 and HCoV-229E are responsible for 10 to 30% of all common colds, and infections occur mainly during the winter and early spring

* Corresponding author. Fax: +32 16 347900.
E-mail address: marc.vanranst@uz.kuleuven.ac.be (M. Van Ranst).

0042-6822/$ - see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.virol.2005.04.010

(Larson et al., 1980). The incubation period is 2 to 4 days. During the 2002–2003 winter season, a new human coronavirus, HCoV-NL63, was isolated in The Netherlands (van der Hoek et al., 2004). The frequent detection of HCoV-NL63 in patient samples worldwide indicates that HCoV-NL63 can be considered as a new important etiologic agent in respiratory tract infections (Arden et al., 2005; Bastien et al., 2005; Moës et al., 2005). Coronaviruses infect all age groups, and reinfections are common. The infection can be subclinical and is usually mild, but there have been reports of more severe lower respiratory tract involvement in infants and elderly people (Gagneur et al., 2002; Vabret et al., 2003). Unlike HCoV-OC43 and HCoV-229E, SARS-CoV and also HCoV-NL63 appear to be associated with more severe respiratory symptoms like pneumonia and bronchiolitis. Only recently, a fifth human coronavirus type, HCoV-HKU1, has been discovered in two patients with pneumonia, and based on genomic analysis, HCoV-HKU1 was proposed to be a distant member of group 2 coronaviruses (Woo et al., 2005).

In our previous study, we described the complete genome sequence of the HCoV-OC43 prototype strain (ATCC VR759), isolated in 1967 from an adult with common cold-like symptoms (McIntosh et al., 1967; Vijgen et al., 2005a). We demonstrated a high rate of similarity with bovine coronaviruses, and we postulated that both viruses diverged from each other (Vijgen et al., 2005a). Molecular clock analysis situated the most recent common ancestor of both viruses around 1890. The evolutionary rate of the HCoV-OC43/BCoV pair was estimated in the order of 10^{-4} nucleotide substitutions per site per year (Vijgen et al., 2005a), which is in the same range as reported for other RNA viruses (Domingo and Holland, 1988). The prototype HCoV-OC43 strain is, however, a laboratory strain, that since its isolation in 1967 was passaged seven times in human embryonic tracheal organ culture, followed by 15 passages in suckling mouse brain, and an unknown number of passages in human rectal tumor HRT-18 cells and/or Vero cells. During the passage history, it is likely that a number of mutations have accumulated. Recently, St-Jean et al. (2004) reported the complete genome sequence of a HCoV-OC43 clinical isolate, designated Paris. A high degree of genetic stability was stated for HCoV-OC43 since only 6 nucleotide variations in the whole genome could be observed in comparison to the HCoV-OC43 ATCC strain, with isolation dates 34 years apart. Based on evolutionary analyses, we demonstrated that this finding seems unlikely, and we suggested that the HCoV-OC43 Paris isolate might have been a result of cross-contamination with the ATCC HCoV-OC43 strain (Vijgen et al., 2005b). In this study, we present the complete genome sequence data of two contemporary non-cell culture adapted HCoV-OC43 strains. The HCoV-OC43 BE03 strain (isolate 87309) was detected in 2003 from a 2-year-old girl suffering from bronchiolitis, and the HCoV-OC43 BE04 strain (isolate 19572) was detected in 2004 in a 1-year old boy, who presented with rhinitis, bronchiolitis and coughing. Both patients were hospitalized at the University Hospital in Leuven, Belgium. Spike gene sequence data are also determined for three additional 2003 and two additional 2004 HCoV-OC43 strains. Phylogenetic and comparative sequence analyses of these data are performed, providing closer insights in the HCoV-OC43 genetic variability.

Results

We here present the complete nucleotide sequence of two contemporary HCoV-OC43 strains. The genome of both contemporary HCoV-OC43 strains encompasses 30723 nucleotides (excluding the 3′ poly(A) tail), being 15 nucleotides shorter than the genome of the ATCC prototype strain (GenBank accession number AY391777). Furthermore, we here report the spike gene sequences of three additional 2003 HCoV-OC43 strains (HCoV-OC43 BE03 isolates 89996, 37767 and 84020) and two additional HCoV-OC43 strains from 2004 (HCoV-OC43 BE04 34364 and 36638). These sequences were deposited in the GenBank database under accession numbers AY903454-AY903460.

Pairwise sequence alignments demonstrate an overall genome similarity of 99.0% between the HCoV-OC43 ATCC prototype strain (AY391777) and the HCoV-OC43 BE03 strain as well as between the prototype strain and the BE04 strain. Complete genome sequence comparisons to the HCoV-OC43 ATCC prototype strain sequenced by St-Jean et al. (AY585228) and the Paris isolate (AY585229) show similar results: 99.0% identity is found between the BE03 strain and the ATCC strain (AY585228) as well as between the BE03 strain and the Paris isolate. For the HCoV-OC43 BE04 strain 98.9% similarity with the complete genome sequences of both the ATCC strain (AY585228) and the Paris isolate is found. The genetic divergence between the prototype and each of the contemporary HCoV-OC43 strains roughly corresponds to an evolutionary rate of about 2.7×10^{-4} nucleotide substitutions per site per year. The 2003 and 2004 HCoV-OC43 strains show 99.4% similarity throughout their complete genomes. When comparing the genome sequence data of the HCoV-OC43 ATCC (AY391777) and contemporary strains, several nucleotide deletions and insertions are observed in the 3′ part of the genome (Fig. 1). Some of these nucleotide deletions and insertions are common for both contemporary strains in reference to the prototype strain, while others are observed in only one of the strains compared to the prototype strain. Similarity percentages between the HCoV-OC43 BE03 and BE04 strains, the ATCC prototype strain and the HCoV-OC43 Paris isolate are calculated for the major open reading frames (ORFs) and their proteins (Table 1). In the 2003 strain, the lowest percentage of identities with the other HCoV-OC43 strains is found for the envelope protein (E) gene and the corresponding amino acid

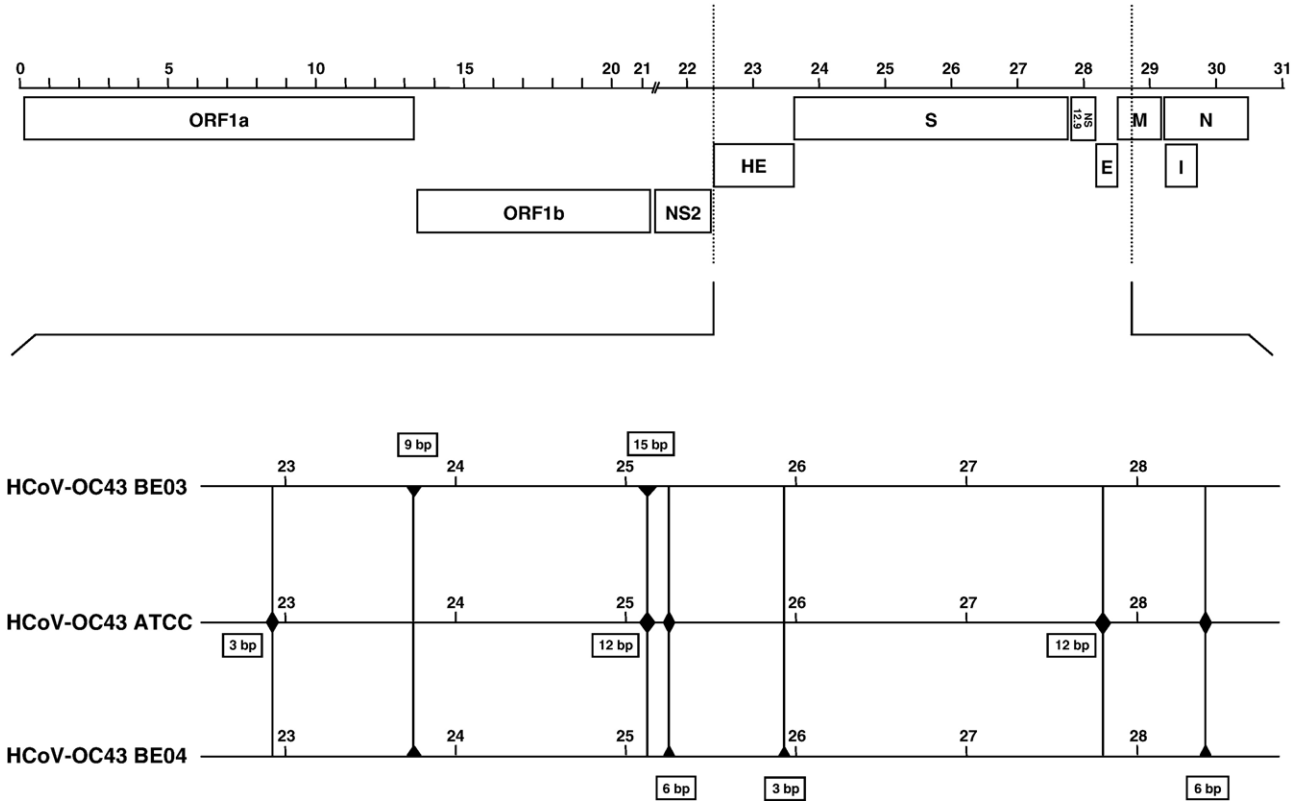

Fig. 1. Overview of the genome organisation of the ATCC prototype (GenBank accession number AY391777) and contemporary HCoV-OC43 strains. A stretch of nucleotides, present in the genome of one strain but not in that of another, is indicated by a triangle. The boxed regions indicate the size of the nucleotide stretch.

sequence. However, the low rate of similarity is not due to a high number of mutations but to a 6 nucleotide deletion at the 3′ end of the HCoV-OC43 BE03 E gene leading to the deletion of a stop codon. The ORF is thereby elongated until the next stop codon, which is present 14 base pairs (bp) downstream. The HCoV-OC43 BE03 E protein is therefore 4 amino acids longer than its HCoV-OC43 ATCC, BE04 and Paris counterparts. Sequencing of the E gene of the other HCoV-OC43 2003 strains demonstrated the same elongation of the E ORF.

Table 1
Nucleotide (and amino acid) similarity percentages of the major ORFs of the HCoV-OC43 ATCC prototype strain (GenBank acc. nr. AY391777), the Paris isolate, and the BE03 and BE04 strains

	ATCC/BE03	ATCC/BE04	Paris/BE03	Paris/BE04	BE03/BE04
ORF1a	99.4 (99.4)	99.4 (99.3)	99.4 (99.4)	99.4 (99.3)	99.8 (99.8)
ORF1b	99.5 (99.8)	99.5 (99.7)	99.5 (99.8)	99.5 (99.7)	99.8 (99.8)
ns2	99.6 (100.0)	99.8 (100.0)	99.5 (100.0)	99.6 (100.0)	99.6 (100.0)
HE	98.0 (97.2)	97.7 (96.5)	98.4 (97.2)	98.1 (96.5)	99.2 (98.8)
S	97.3 (96.2)	97.0 (95.5)	97.1 (96.0)	96.5 (95.2)	97.2 (96.9)
ns12.9	99.4 (98.2)	99.7 (99.1)	99.4 (98.2)	99.7 (99.1)	99.7 (99.1)
E[a]	93.3 (94.3)	99.6 (98.8)	93.3 (94.3)	99.6 (98.8)	93.6 (95.5)
M	98.8 (98.3)	98.7 (98.3)	99.0 (98.3)	98.8 (98.3)	99.6 (100.0)
N	99.2 (98.7)	98.8 (98.2)	99.3 (98.9)	98.9 (98.4)	99.2 (98.2)
Ia/Ib[b]	99.2 (ND[c])	99.0 (ND[c])	99.4 (99.0)	99.2 (98.6)	98.9 (98.6)

[a] A deletion of a stop codon in the HCoV-OC43 BE03 E gene leads to an elongation of the ORF with 12 base pairs, and thus 4 amino acids.
[b] The HCoV-OC43 ATCC I coding region contains a stop codon at position 29345, resulting in two potential coding regions of 60 amino acids (Ia) and 115 amino acids (Ib). The percentage nucleotide similarity is calculated for the continuous Ia/Ib region.
[c] Not done.

The S gene is the most polymorphic ORF, with nucleotide (and amino acid) identity percentages of 97.3% (96.2%) between the ATCC prototype strain (AY391777) and the BE03 strain, 97.1% (96.0%) between the Paris isolate and the BE03 strain, 97.0% (95.5%) for the prototype strain/BE04 pair, and 96.5% (95.2%) for the Paris/BE04 pair. SimPlot analysis of the complete genomes of the HCoV-OC43 BE03 and BE04 strains, the Paris isolate and the HCoV-OC43 ATCC strain (AY585228) in reference to the ATCC prototype strain (AY391777), also demonstrates for the BE03 and BE04 strains that the highest variability is found in the genome region containing the spike gene (Fig. 2). Most remarkably, the spike genes of the BE03 and BE04 strains, with detection dates only one year apart, show only 97.2% nucleotide and 96.9% amino acid similarity which corresponds to a total of 93 nucleotide variations leading to 34 amino acid changes, 15 nucleotide (or 5 amino acid) insertions in the BE03 strain spike gene and 9 nucleotide (or 3 amino acid) insertions in the BE04 strain spike gene. Out of 93 polymorphic nucleotides, 60 are located in the S1 subunit and these include 29 synonymous substitutions and 31 nonsynonymous substitutions leading to 23 amino acid changes. Nucleotide insertions and

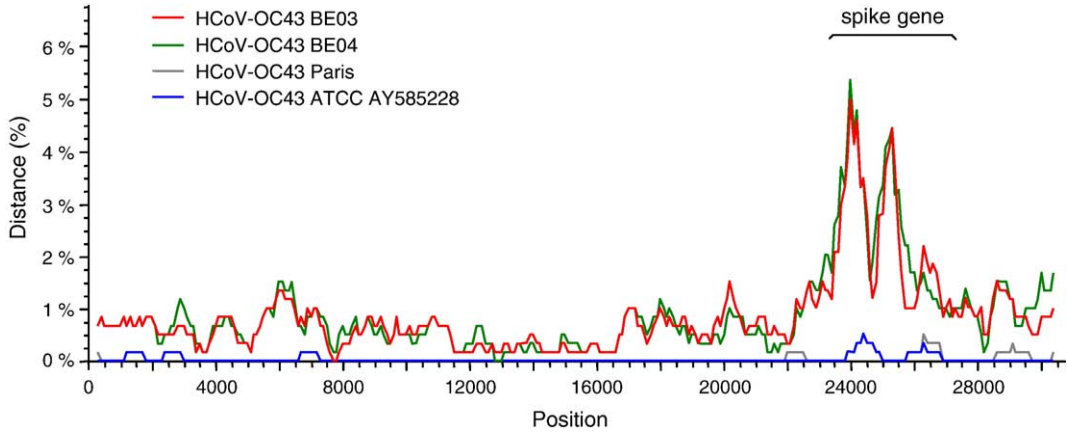

Fig. 2. SimPlot analysis of complete genome sequence data of the HCoV-OC43 BE03, BE04, Paris, and ATCC (GenBank accession number AY585228) strains in reference to HCoV-OC43 ATCC strain GenBank accession number AY391777. Each point plotted is the percentage genetic distance within a sliding window of 600 bp wide, centered on the position plotted, with a step size of 100 bp. Each curve is a comparison between the genome of one of the strains and the reference genome of the ATCC HCoV-OC43 strain (AY391777).

deletions are also exclusively found in the S1 subunit, which codes for the fragment of the spike glycoprotein that binds to its cellular receptor. The putative N-glycosylation pattern of the spike proteins of the contemporary strains also differs slightly between both strains and in comparison to that of the HCoV-OC43 prototype spike protein. There are 15

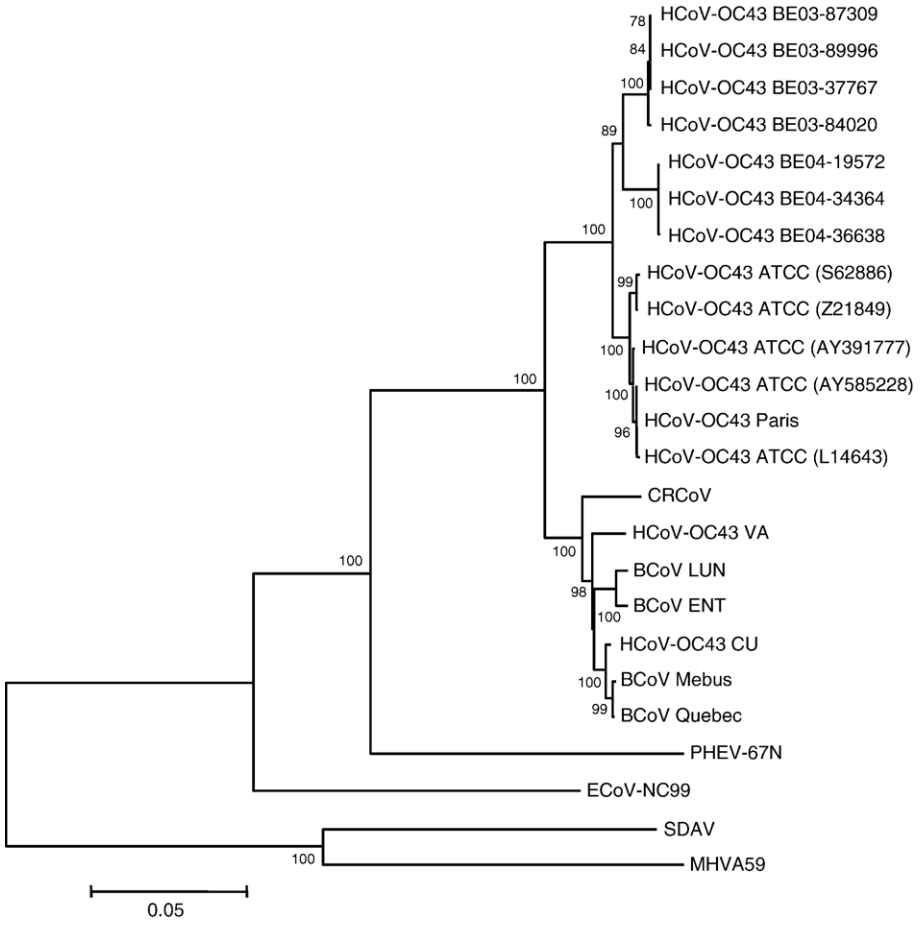

Fig. 3. Neighbor joining phylogenetic tree of the complete spike gene nucleotide sequence data of group 2 coronaviruses: HCoV-OC43 BE03 strain (GenBank accession numbers AY903454, AY903456, AY903457, AY903459), HCoV-OC43 BE04 strain (AY903455, AY903458, AY903460), HCoV-OC43 ATCC VR759 strain (AY391777, AY585228, L14643, Z21849, S62886, Z32768, Z32769), HCoV-OC43 serotype OC43 Paris (AY585229), BCoV LUN (AF391542), BCoV ENT (AF391541), BCoV Mebus (U00735), BCoV Quebec (AF220295), CRCoV (AY150272), PHEV strain 67N (AY078417), ECoV strain NC99 (AY316300), SDAV (AF207551) and MHV strain A59 (AY700211). Bootstrap values over 75% are shown.

potential N-glycosylation sites (Asn-Xxx-Ser/Thr) in the spike protein of the 2003 strain, and 17 are predicted to be present in the spike protein of the 2004 strain, while the prototype HCoV-OC43 spike protein contains 14 potential N-glycosylation sites (NetNGlyc 1.0 Server).

The observed variability in the spike gene of the 2003 and 2004 HCoV-OC43 strains is further investigated by sequencing this gene in three additional 2003 and two additional 2004 HCoV-OC43 positive samples. Among the strains sampled in the same year, a high degree of conservation is found in their spike genes: no more than 0.15% difference between the 2003 strains, and no more than 0.05% between the 2004 strains. A multiple sequence alignment of these sequence data and spike gene sequence data of HCoV-OC43 and of other group 2 coronaviruses, available in Genbank, is performed. Based on this alignment, a neighbor joining phylogenetic tree is constructed and evaluated by 1000 bootstrap pseudoreplicates (Fig. 3). In the HCoV-OC43 branch, three phylogenetic clusters can clearly be demonstrated: the ATCC cluster, containing all ATCC strains as well as the Paris isolate, a second cluster containing all HCoV-OC43 BE03 strains, and a third cluster, in which all HCoV-OC43 BE04 strains are found. Molecular clock analysis is performed using the spike gene sequence data of HCoV-OC43 and BCoV strains for which the sampling date is known (Table 2). A likelihood ratio test indicates that the molecular clock hypothesis, which assumes that the evolutionary rate is roughly constant among lineages, cannot be rejected for these serially sampled data ($P = 0.62$). Using a Bayesian coalescent approach, the time to the most recent common ancestor (TMRCA) of the HCoV-OC43 2003 and 2004 strains is estimated in 1971 with a 95% highest posterior density interval of 1962–1979. The evolutionary rate is estimated to be 3.5×10^{-4} substitutions per site per year (95% highest posterior density interval of 2.6×10^{-4}–4.5×10^{-4}).

Table 2
Sampling data of bovine and human coronaviruses to calculate the TMRCA

Strain	GenBank accession nr.	Sampling date
HCoV-OC43 ATCC	AY391777	1967
HCoV-OC43 ATCC	AY585228	1967
HCoV-OC43 ATCC	Z21849	1967
HCoV-OC43 BE03 isolate 37767	AY903457	2003
HCoV-OC43 BE03 isolate 84020	AY903456	2003
HCoV-OC43 BE03 isolate 87309	AY903459	2003
HCoV-OC43 BE03 isolate 89996	AY903454	2003
HCoV-OC43 BE04 isolate 19572	AY903460	2004
HCoV-OC43 BE04 isolate 34364	AY903455	2004
HCoV-OC43 BE04 isolate 36638	AY903458	2004
BCoV Mebus	U00735	1972
BCoV Quebec	AF220295	1972
BCoV LUN	AF391542	1998
BCoV ENT	AF391541	1998
BCoV-LSU94	AF058943	1994
BCoV M80844	M80844	1989
BCoV-LY138	AF058942	1965
BCoV BECS	D00731	1979

In all 2003 and 2004 contemporary HCoV-OC43 strains, an important amino acid change, in reference to the ATCC strains, is present in the proteolytic cleavage site of the HCoV-OC43 S protein. Unlike the HCoV-OC43 ATCC prototype strain, in which this sequence is RRSRG, the contemporary strains have a G to R amino acid change in the last position, leading to an RRSRR motif.

In the HCoV-OC43 ATCC prototype strain, the internal ORF (I) coding region contains a stop codon at position 29345, resulting in two potential coding regions of 60 amino acids and 115 amino acids. This stop codon is not present in the BE03 and BE04 strains, which have the capacity to code for a 207-amino acid protein.

Discussion

Human coronaviruses are an important cause of respiratory tract infections. Before the SARS outbreak coronaviruses have been somewhat neglected in human medicine, although they have always played a significant role in animal virology. Only recently, we reported the complete genome sequence of the prototype HCoV-OC43 strain (ATCC VR759; GenBank accession number AY391777) (Vijgen et al., 2005a). This strain, however, has been passaged several times in organ culture, suckling mouse brain and cell culture, and the presence of culture-related polymorphisms is likely. In this study, we screened a collection of RSV-negative bronchiolitis samples in order to identify contemporary HCoV-OC43 strains and we determined the full-length genome sequence of two circulating strains (HCoV-OC43 BE03 and BE04) detected in 2003 and 2004, respectively. When comparing the complete genome sequence of the contemporary strains to that of the ATCC prototype strain and to the HCoV-OC43 Paris isolate, we found similarity percentages of 98.9% to 99.0%. The genetic divergence between each of the circulating strains and the ATCC strain corresponds roughly to an evolutionary rate of about 2.7×10^{-4} nucleotide substitutions per site per year. This value is consistent with the previously estimated evolutionary rate of the BCoV/HCoV-OC43 pair, and that of other coronaviruses (TGEV and SARS-CoV) (Salemi et al., 2004; Sanchez et al., 1992; Vijgen et al., 2005a).

In reference to the prototype strain, in frame nucleotide insertions and deletions are present in the HE, S and E genes of the contemporary strains, leading to variability in their predicted protein lengths. HE, S and E proteins are expressed at the surface of the virus particles and host immune pressures can attribute to the occurrence of variations, insertions and deletions in their genes. Deletion of a stop codon at the 3′ end of the HCoV-OC43 E gene of the BE03 strain elongates this ORF with 12 bp or 4 amino acids. The same elongation of the E ORF by deletion of a stop codon is also present in the other HCoV-OC43 strains

from 2003, but not in the 2004 samples, indicating that this phenomenon might be characteristic for the HCoV-OC43 strain circulating in Belgium in 2003. These four additional amino acids (IQTL) are present at the carboxyterminus, the part of the E protein that is located inside the virion. Whether this E protein elongation has a functional significance remains to be elucidated. The spike gene and protein are the most polymorphic when comparing the major ORFs of the prototype and contemporary HCoV-OC43 strains. The spike glycoprotein is the major coronavirus antigen and plays an important role in attachment of the virus to cell surface receptors and induces the fusion of viral and cellular membranes (Spaan et al., 1988). Host immunity escape mechanisms, variation in host range and tissue tropism of coronaviruses are largely attributed to variations in the spike glycoprotein (Gallagher and Buchmeier, 2001). Interestingly, the similarity between the spike genes and proteins of the 2003 and 2004 circulating HCoV-OC43 strains is lower (97.2% and 96.9%, respectively) than would be expected based on the HCoV-OC43 evolutionary rate. The observed spike gene variability provides evidence for the existence of several genetically distinct HCoV-OC43 strains with different temporal and possible geographical circulation patterns. Spike gene sequencing of the additional 2003 and 2004 HCoV-OC43 strains reveals more characteristics of the HCoV-OC43 strains circulating in 2003 and 2004. All BE03 strains have a 4092-nt spike gene coding for a 1363-amino acid protein, while the spike gene of all BE04 strains is six nucleotides shorter encoding a 1361-amino acid protein. The HCoV-OC43 BE04 spike gene is of the same length as the HCoV-OC43 ATCC prototype strain spike gene, but shows a different nucleotide insertion and deletion pattern. Strain-specific nucleotide variations, insertions and deletions are present among all samples of the same year of detection. Phylogenetic analysis of spike gene sequences of these four Belgian BE03 and three Belgian BE04 HCoV-OC43 strains, of the HCoV-OC43 Paris isolate spike gene, and of all ATCC strain spike sequence data, confirms the existence of genetically different HCoV-OC43 strains. Three distinct phylogenetic clusters can be demonstrated: the ATCC cluster, containing all ATCC laboratory strains as well as the Paris isolate, a second cluster, containing all HCoV-OC43 BE03 strains, and a third cluster, in which all HCoV-OC43 BE04 strains are found (Fig. 3). The HCoV-OC43 Paris isolate clusters with all ATCC strains and not with the contemporary strains, implicating that this isolate might not be a circulating strain, but rather a result of cross-contamination with the ATCC HCoV-OC43 strain (Vijgen et al., 2005b).

At a certain time in history the HCoV-OC43 2003 and 2004 circulating strains diverged from each other, and based on molecular clock analysis of spike gene sequence data, their most recent common ancestor can be dated back to 1971 (95% highest posterior density interval: 1962–1979). According to this molecular clock model, an evolutionary rate of 3.5×10^{-4} nucleotide substitutions per site per year is estimated (95% highest posterior density interval: 2.6×10^{-4}–4.5×10^{-4}), a value consistent with our previous findings (Vijgen et al., 2005a). The existence of two different HCoV-OC43 strains circulating in Belgium in 2003 and in 2004 is also supported by the difference in length of the E protein, which is four amino acids longer in the BE03 samples, while this elongation of the E ORF is not present in the BE04 samples.

Cleavage of the coronavirus spike protein into the subunits S1 and S2 is mediated by cellular trypsin-like proteases acting at the C-terminus of the sequence N-Arg-Arg-Xxx-Arg-Arg-C (Abraham et al., 1990). This cleavage process is believed to play an important, although not obligatory role in the fusion activity and viral infectivity of BCoV and MHV (Stauber et al., 1993; Storz et al., 1981). In all BE03 and BE04 HCoV-OC43 strains, a glycine to arginine variation compared to all spike protein sequences of the ATCC HCoV-OC43 strain and the Paris isolate is present in this proteolytic cleavage site. This observation might have possible important functional consequences, as the cleavage site sequence is RRSRG in the ATCC strains and these have been reported to have an uncleaved spike protein (Hogue and Brian, 1986; Künkel and Herrler, 1993). All 2003 and 2004 contemporary strains, however, have a G to R amino acid change in the last position, leading to an RRSRR motif, identical to that of BCoV-Mebus, which therefore might lead to an increased cleavability compared to the ATCC prototype HCoV-OC43 spike protein. This amino acid change is also present in two HCoV-OC43 strains described by Künkel and Herrler (1996) (OC43-CU and OC43-VA), which have been shown to be cleaved to an extent of nearly 40% in infected cells. Analysis of the spike gene of these two strains in comparison to the ATCC prototype HCoV-OC43 strains and to BCoV-Mebus reveals a closer relationship to BCoV than to the ATCC HCoV-OC43 strains.

Another observation when comparing the HCoV-OC43 contemporary and prototype strains is present in the internal (I) ORF, localized within the nucleocapsid gene. In both contemporary HCoV-OC43 strains, the I gene encodes a 207-amino acid protein, as is also observed in BCoV, PHEV and MHV (Fischer et al., 1997; Lapps et al., 1987). In the HCoV-OC43 prototype strain, however, an early stop codon is present in the I coding region leading to two potential coding regions of 60 amino acids and 115 amino acids. The occurrence of this early stop codon is due to a single nucleotide variation, which probably occurred during cell culture passaging.

In this study, we performed a comparative genomic analysis of two circulating non-cell culture adapted HCoV-OC43 strains and the cell cultured prototype strain. Furthermore, based on spike gene sequence data of four 2003 and three 2004 strains, we demonstrate the circulation of two genetically different HCoV-OC43 strains in Belgium in 2003 and 2004, respectively. We provide substantial evidence for the genetic variability of HCoV-OC43,

supporting our previous conclusions based on evolutionary analyses (Vijgen et al., 2005b).

Materials and methods

Screening for contemporary coronavirus strains

A collection of respiratory tract specimens from neonates and infants hospitalized with respiratory syncytial virus (RSV)-negative respiratory tract infections was screened using coronavirus consensus primers amplifying a 251-bp amplicon in a RT-PCR reaction (Moës et al., 2005). No prior amplification by cell culture was performed. Samples found positive were sequenced in both directions using the RT-PCR primer set.

Sequencing of the contemporary HCoV-OC43 genomes

To determine the genomic sequence of the contemporary HCoV-OC43 strains, a set of overlapping RT-PCR products with an average size of 1.5 kb encompassing the entire genome were generated. For both RT-PCR and sequencing, oligonucleotide primers were designed in regions that were conserved between the BCoV and MHV genomes. The forward PCR primer in the 5′-terminal sequence (OC43F1: 5′-GATTGTGAGCGATTTGC-3′) was based on the HCoV-OC43 5′UTR partial sequence (Wu, H.Y., Guy, J.S., Yoo, D., Vlasak, R. and Brian, D.A., University of Tennessee, Knoxville, TN, unpublished; GenBank accession number AF523847). To generate RT-PCR products containing the exact 3′-terminal sequence, we used oligonucleotide OC43R74 (5′-TTTTTTTTTTGTGATTCTTCCA-3′) based on the conserved 3′-end sequence of all known group-2 coronaviruses. Using 150 sequencing primers, sequencing in both directions was performed on an ABI Prism 3100 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA) using the BigDye terminator v3.1 cycle sequencing kit. Chromatogram sequencing files were inspected with Chromas 2.2 (Technelysium, Helensvale, Australia), and contigs were prepared using SeqMan II (DNASTAR, Madison, WI).

DNA sequence submission

The nucleotide sequence data reported in this paper were deposited in GenBank using the National Center for Biotechnology Information (NCBI, Bethesda, MD) BankIt v3.0 and Sequin v5.35 submission tools under accession numbers AY903454–AY903460.

DNA and protein sequence analysis

DNA and protein similarity searches were performed using the NCBI BLAST (Basic local alignment search tool) server on GenBank DNA database release 118.0 (Altschul et al., 1990). Pairwise nucleotide and protein sequence alignments were performed using FASTA algorithms in the ALIGN program on the GENESTREAM network server (http://www2.igh.cnrs.fr) at the Institut de Génétique Humaine in Montpellier, France (Pearson et al., 1997). Multiple sequence alignments were prepared using CLUSTALX version 1.82 (Thompson et al., 1997), and manually edited in GENEDOC (Nicholas et al., 1997). The SimPlot program version 3.2 was used to plot the genetic distance between two HCoV-OC43 strains versus nucleotide positions (Lole et al., 1999). Phylogenetic analyses were conducted using MEGA version 2.1 (Kumar et al., 2001). Potential N-glycosylation sites in the HCoV-OC43 spike proteins were predicted using the CBS NetNGlyc 1.0 Server (http://www.cbs.dtu.dk/services/NetNGlyc/).

Timing the most recent common ancestor

The relationship between sampling date and genetic divergence was investigated using a linear regression, based on a maximum likelihood tree, as implemented in the Path-O-Gen software, kindly provided by Andrew Rambaut (University of Oxford, UK). The molecular clock hypothesis was tested using a likelihood ratio test that evaluates the relative goodness-of-fit of a model assuming a molecular clock for serially-sampled data compared to a model that does not assume rate constancy (Rambaut, 2000). Evolutionary rates and divergence times were estimated using Bayesian inference in BEAST v1.03 (Drummond, A., Rambaut, A. BEAST v1.0, available from http://evolve.zoo.ox.ac.uk/beast/, Drummond et al., 2002). Markov chain Monte Carlo (MCMC) inferences were made under a constant population size demographic function. A chain was run for 10×10^6 generations and sampled every 1000th generation after burn-in (10%).

Acknowledgments

We would like to thank all colleagues of the laboratory of Clinical Epidemiological Virology, Department of Microbiology Immunology, Rega Institute for Medical Research, University of Leuven, Belgium, for helpful comments and discussion. This work was supported by a fellowship of the Flemish Fonds voor Wetenschappelijk Onderzoek (FWO) to Leen Vijgen, and by FWO-grant G.0288.01.

References

Abraham, S., Kienzle, T.E., Lapps, W., Brian, D.A., 1990. Deduced sequence of the bovine coronavirus spike protein and identification of the internal proteolytic cleavage site. Virology 176, 296–301.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.

Arbour, N., Day, R., Newcombe, J., Talbot, P.J., 2000. Neuroinvasion by human respiratory coronaviruses. J. Virol. 74, 8913–8921.

Arden, K.E., Nissen, M.D., Sloots, T.P., Mackay, I.M., 2005. New human coronavirus, HCoV-NL63, associated with severe lower respiratory tract disease in Australia. J. Med. Virol. 75, 455–462.

Bastien, N., Anderson, K., Hart, L., Van Caeseele, P., Brandt, K., Milley, D., Hatchette, T., Weiss, E.C., Li, Y., 2005. Human coronavirus NL63 infection in Canada. J. Infect. Dis. 191, 503–506.

Cavanagh, D., 1997. Nidovirales: a new order comprising *Coronaviridae* and *Arteriviridae*. Arch. Virol. 142, 629–633.

Domingo, E., Holland, J.J., 1988. High error rates, population equilibrium and evolution of RNA replication systems. In: Domingo, E., Holland, J.J., Ahlquist, P. (Eds.), RNA Genetics, vol. 3. CRC Press, Boca Raton, pp. 3–36.

Drummond, A.J., Nicholls, G.K., Rodrigo, A.G., Solomon, W., 2002. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. Genetics 161, 1307–1320.

Fischer, F., Peng, D., Hingley, S.T., Weiss, S.R., Masters, P.S., 1997. The internal open reading frame within the nucleocapsid gene of mouse hepatitis virus encodes a structural protein that is not essential for viral replication. J. Virol. 71, 996–1003.

Gagneur, A., Sizun, J., Vallet, S., Legr, M.C., Picard, B., Talbot, P.J., 2002. Coronavirus-related nosocomial viral respiratory infections in a neonatal and paediatric intensive care unit: a prospective study. J. Hosp. Infect. 51, 59–64.

Gallagher, T.M., Buchmeier, M.J., 2001. Coronavirus spike proteins in viral entry and pathogenesis. Virology 279, 371–374.

Hogue, B.G., Brian, D.A., 1986. Structural proteins of human respiratory coronavirus OC43. Virus Res. 5, 131–144.

Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17, 1244–1245.

Künkel, F., Herrler, G., 1993. Structural and functional analysis of the surface protein of human coronavirus OC43. Virology 195, 195–202.

Künkel, F., Herrler, G., 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.

Lai, M.M., Holmes, K.V., 2001. Coronaviridae: the viruses and their replication. In: Fields, B.N., Knipe, D.M., Howley, P.M. (Eds.), Fields Virology, 4th ed. Lippincott Williams and Wilkins, Philadelphia, pp. 1163–1185.

Lapps, W., Hogue, B.G., Brian, D.A., 1987. Sequence analysis of the bovine coronavirus nucleocapsid and matrix protein genes. Virology 157, 47–57.

Larson, H.E., Reed, S.E., Tyrell, D.A.J., 1980. Isolation of rhinoviruses and coronaviruses from 38 colds in adults. J. Med. Virol. 5, 221–229.

Lole, K.S., Bollinger, R.C., Paranjape, R.S., Gadkari, D., Kulkarni, S.S., Novak, N.G., Ingersoll, R., Sheppard, H.W., Ray, S.C., 1999. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. J. Virol. 73, 152–160.

McIntosh, K., Becker, W.B., Chanock, R.M., 1967. Growth in suckling mouse brain of "IBV-like" viruses from patients with upper respiratory tract disease. Proc. Natl. Acad. Sci. U.S.A. 58, 2268–2273.

Moës, E., Vijgen, L., Keyaerts, E., Zlateva, K., Maes, P., Pyrc, K., Berkhout, B., van der Hoek, L., Van Ranst, M., 2005. A novel pancoronavirus RT-PCR assay: frequent detection of human coronavirus NL63 in children hospitalized with respiratory tract infections in Belgium. BMC Infect. Dis. 5, 6.

Nicholas, K.B., Nicholas, H.B., Deerfield, D.W., 1997. GeneDoc: analysis and visualization of genetic variation. Embnet News 4, 14.

Pearson, W.R., Wood, T., Zhang, Z., Miller, W., 1997. Comparison of DNA sequences with protein sequences. Genomics 46, 24–36.

Rambaut, A., 2000. Estimating the rate of molecular evolution: incorporating non-contemporaneous sequences into maximum likelihood phylogenies. Bioinformatics 16, 395–399.

Salemi, M., Fitch, W.M., Ciccozzi, M., Ruiz-Alvarez, M.J., Rezza, G., Lewis, M.J., 2004. Severe acute respiratory syndrome coronavirus sequence characteristics and evolutionary rate estimate from maximum likelihood analysis. J. Virol. 78, 1602–1603.

Sanchez, C.M., Gebauer, F., Sune, C., Mendez, A., Dopazo, J., Enjuanes, L., 1992. Genetic evolution and tropism of transmissible gastroenteritis coronaviruses. Virology 190, 92–105.

Spaan, W., Cavanagh, D., Horzinek, M.C., 1988. Coronaviruses: structure and genome expression. J. Gen. Virol. 69, 2939–2952.

Stauber, R., Pfleiderera, M., Siddell, S., 1993. Proteolytic cleavage of the murine coronavirus surface glycoprotein is not required for fusion activity. J. Gen. Virol. 74, 183–191.

St-Jean, J.R., Jacomy, H., Desforges, M., Vabret, A., Freymuth, F., Talbot, P.J., 2004. Human respiratory coronavirus OC43: genetic stability and neuroinvasion. J. Virol. 78, 8824–8834.

Storz, J., Rott, R., Kaluza, G., 1981. Enhancement of plaque formation and cell fusion of an enteropathogenic coronavirus by trypsin treatment. Infect. Immun. 31, 1214–1222.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4878.

Vabret, A., Mourez, T., Gouarin, S., Petitjean, J., Freymuth, F., 2003. An outbreak of coronavirus OC43 respiratory infection in Normandy, France. Clin. Infect. Dis. 36, 985–989.

van der Hoek, L., Pyrc, K., Jebbink, M.F., Vermeulen-Oost, W., Berkhout, R.J., Wolthers, K.C., Wertheim-van Dillen, P.M., Kaandorp, J., Spaargaren, J., Berkhout, B., 2004. Identification of a new human coronavirus. Nat. Med. 10, 368–373.

Vijgen, L., Keyaerts, E., Moës, E., Thoelen, I., Wollants, E., Lemey, P., Van Damme, A.M., Van Ranst, M., 2005a. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. J. Virol. 79, 1595–1604.

Vijgen, L., Lemey, P., Keyaerts, E., Van Ranst, M., 2005b. Genetic variability of human respiratory coronavirus OC43. J. Virol. 79, 3223–3224 (Authors' reply 3224–3225).

Woo, P.C., Lau, S.K., Chu, C.M., Chan, K.H., Tsoi, H.W., Huang, Y., Wong, B.H., Poon, R.W., Cai, J.J., Luk, W.K., Poon, L.L., Wong, S.S., Guan, Y., Peiris, J.S., Yuen, K.Y., 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J. Virol. 79, 884–895.

Zhang, X.M., Kousoulas, K.G., Storz, J., 1992. The hemagglutinin/esterase gene of human coronavirus strain OC43: phylogenetic relationships to bovine and murine coronaviruses and influenza C virus. Virology 186, 318–323.