Abstract of thesis entitled

"Molecular epidemiology of human coronavirus OC43 in Hong Kong"

Submitted by

Paul Lee

for the degree of Master of Medical Sciences

at the University of Hong Kong

in August 2007

The severe acute respiratory syndrome (SARS) epidemic in 2003 has led to a boost in interests in the study of coronaviruses in both humans and animals. Apart from SARS coronavirus (SARS-CoV), human coronavirus OC43 (HCoV-OC43), human coronavirus 229E (HCoV-229E), coronavirus HKU1 (CoV-HKU) and human coronavirus NL63 (HCoV-NL63) are associated with human respiratory tract infections. As a result of the unique mechanism of viral replication, coronaviruses have a high frequency of recombination. In a previous study on CoV-HKU1, natural recombination events were identified in the generation of a novel genotype. However, it is not known if such recombination events are also found in other human coronaviruses. Previous studies have identified the presence of two genotypes of contemporary HCoV-OC43 strains.

In this study, 23 HCoV-OC43 strains identified from nasopharyngeal aspirates (NPAs) of patients with respiratory tract infections from 2004-2006 were subject to complete RNA-dependent RNA polymerase (RdRp), spike (S) and nucleocapsid (N) gene sequencing and analysis. Phylogenetic analysis of the three genes of the 23 strains and those of strains with sequences available showed the presence of three clusters, one including ATCC and Paris isolates (clade A), and the other two corresponding to the two previously identified genotypes (clades B and C). In the S and N genes, five of the 23 strains belonged to clade B while 18 belonged to clade C. In the RdRp gene, the five strains that belonged to clade B in their S and N genes were also clustered to form clade B, while 17 strains belonged to clade C. However, the RdRp sequence of one (HK04-12 from year 2004) of the strains of clade C in its S and N gene and that of the BE04 strain were not clustered with other clade C strains, but fell into the cluster formed by the clade B strains, which is in line with results obtained from multiple alignment revealing that the two "unusual" strains displayed hig nucleotide similarity to clade B in their RdRp genes, but to clade C in their S ar N genes. The results suggested the presence of two genotypes of HCoV-OC43 Hong Kong, with at least one strain potentially belonging to an addition genotype probably arising from recombination between strains from clade B and

C, a situation similar to that reported for CoV-HKU1. Further studies on the complete genome sequences of more HCoV-OC43 strains, especially the "unusual" strains are required to localize the potential sites of recombination.

----- Total word count: 404 -----

Molecular epidemiology of human coronavirus OC43 in Hong Kong

By

Paul Lee

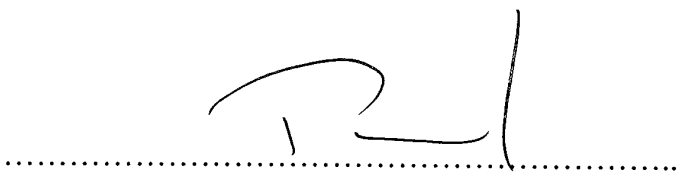
B.Sc. *UK*

This thesis is submitted to the Department of Microbiology of The University of Hong Kong in partial fulfillment of the requirements for the Degree of Master of Medical Sciences

August 2007

Declaration

I, Paul Lee, declare that this dissertation represents my own work, except where due acknowledgement is made, and that it has not been previously included in a thesis, dissertation or report submitted to this University or to any other institution for a degree, diploma or other qualifications.



Mr. Paul Lee

Acknowledgments

Many people have helped me in different ways both technically and personally, no matter what I have come through in these 2 years of part time studies. I would like to dedicate my sincere thanks to all of them in this section.

I wish to present my deepest thanks to my supervisor Dr. Susanna K. P. Lau and Professor Patrick C Y Woo. I would like to thank them for giving me a chance to have postgraduate study in this famous university in the world to learn and develop scientific skills and knowledge. Thanks for their professional guidance, patience and kind encouragement during my postgraduate studies. I would also like to thank Professor K. Y. Yuen for his valuable suggestion and discussion during my research work.

Very special thanks to Mr. H. W. Tsoi and Mr. Cyril C. Y. Yip for their technical assistance throughout my study and Dr. Herman Tse for providing clinical data.

Also, I would like to thank all of my colleagues from the Department of Microbiolog for the encouragements and supports, especially Mr. Kenneth S. M. Li, Miss Beatrice

H. L. Wong, Miss Rosana W. S. Poon, Miss Candy C. Y. Lau, Dr. Ken T. K. Chong, Dr. Jade L. L. Teng, Miss Carol S. F. Lam, Miss Shirly O. T. Curreem, Mr. Kenneth K. Y. Lai, Mr. Alan K. L. Tsang, Miss Clara T. Y. Fu and Miss Joanna S. Y. Chan.

List of tables and figure

List of Abbreviations

ATCC	American type culture collection
Bat-CoV	Bat coronavirus
BCoV	Bovine coronavirus
bp	Base pair(s)
CoV-HKU1	Coronavirus HKU1
CRCoV	Canine respiratory coronavirus
DNA	Deoxyribonucleic acid
E	Envelope
ER	Endoplasmic reticulum
FIPV	Feline infectious peritonitis
HCoV	Human coronavirus
HCoV-NL63	Human coronavirus NL63
HCoV-OC43	Human coronavirus OC43
HCoV-229E	Human coronavirus 229E
HIV	Human immunodeficiency virus
HE	Hemagglutinin-esterase
IBV	Avian infectious bronchitis virus
M	Membrane
MHV	Murine hepatitis virus
N	Nucleocapsid
NPAs	Nasopharyngeal aspirates
nsps	Non-structural proteins
no.	Number

ORF	Open reading frames
PCR	Polymerase chain reaction
RdRp	RNA-dependent RNA polymerase
PEDV	Porcine epidemic diarrhea virus
PHEV	Porcine hemagglutinating encephalomyelitis virus
RNA	Ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
S	Spike
SARS	Severe acute respiratory syndrome
SARS-CoV	Severe acute respiratory syndrome coronavirus
SDVA	Rat sialodacryoadenitis virus
TCoV	Turkey coronavirus
TGEV	Porcine transmissible gastroenteritis virus

List of Symbols

^{o}C	Degree Celsius
µg	Microgram
µl	Microliter
ml	Milliliter
min	Minutes
%	Percent
s	Second
U	Units

Contents

Chapter 1: Introduction

1.1 Features of coronavirus

1.1.1 Classification

Coronavirus belongs to the family Coronaviridae, under the order Nidovirales. Before the SARS epidemics in 2002-2003, coronavirus was somehow neglected in human medicine. Coronavirus infect a variety of hosts, including mouse, feline, human, canine, bird porcine and bat. It causes mild to serious and sometimes fatal respiratory, enteric, cardiovascular and neurologic diseases. Based on the genetic and serological characterization, coronavirus can be classified into three groups. Group 1 contains human coronavirus 229E (HCoV-229E), human coronavirus NL63 (HCoV-NL63), porcine epidemic diarrhea virus (PEDV), porcine transmissible gastroenteritis virus (TGEV), feline infectious peritonitis (FIPV) and bat coronavirus HKU2 (Bat-CoV HKU2). Group 2 contains human coronavirus OC43 (HCoV-OC43), bovine coronavirus (BCoV), coronavirus HKU1 (CoV-HKU1), murine hepatitis virus (MHV), porcine hemagglutinating encephalomyelitis virus (PHEV), rat

sialodacryoadenitis virus (SDAV), canine respiratory coronavirus (CRCoV), bat coronavirus: HKU4, HKU5 and HKU9 (Bat-CoV HKU4, 5, and 9). Group 3 contains turkey coronavirus (TCoV) and avian infectious bronchitis virus (IBV) (1, 2, 3, 4). For the SARS coronavirus, it does not closely resemble to any of these 3 groups. Previous studies suggested that, SARS-CoV represent an early split-off from group 2 lineage (5).

1.1.2 Genome organization

Coronavirus is an enveloped virus with positive-sense single-stranded RNA genome. The genome size of coronavirus can be up to 31kb which is the largest known among all RNA viruses (6). The order of the gene encoding the viral RNA-dependent RNA-polymerase (RdRp) and the four common structure proteins: spike (S), envelope (E), membrane (M) and nucleocapsid (N) is arranged as RdRp-S-E-M-N. The genome of coronavirus also included few additional open reading frames (ORFs) that encode non-structural protein that with unknown function.

1.1.3 Morphology and structure

Coronavirus was first recognized as a distinct virus group in negative stained preparation. Its morphology is like a crown under electron microscope. The virion of coronavirus is spherical enveloped particles with 100-200nm in diameter. Inside the envelope, there is a helical nucleocapsid core structure including the viral RNA and N phosphoprotein. Two type of spikes (S and HE) line outside the virion. S protein is present in all coronaviruses while the HE protein is present in group 2 coronavirus only. M protein is a component in both internal core structure and the envelope. Small amount of E proteins are present in the envelope.

1.1.4 Epidemiology and clinical features

Coronavirus shows marked species specificity and strong tissue tropism. They usually infect via respiratory or gut tract. Even the same coronavirus species, different isolates show different tropisms for respiratory and gut tract. The tropism is determined by the S protein and the type of receptors. The prevalence of coronavirus may have been underestimated because of lacking

sensitivity diagnostic methods and the fastidious nature of viral growth. Epidemiological and volunteer inoculation studies found that coronavirus was associated with variety of respiratory illnesses but only with low pathogenicity except SARS-CoV. They cause upper respiratory infection with infrequent cases of pneumonia asthma exacerbations in children as well as chronic bronchitis in adults and elderly (7, 8).

1.2 Coronavirus HKU1 (CoV-HKU1)

1.2.1 Discovery of CoV-HKU1

Coronavirus HKU1 was first discovered from a 71-year-old man with pneumonia in Hong Kong who had just returned from Shenzhen, China in January 2005. This virus was also detected in United States, Australia and France shortly afterwards (9, 10, 11). In order to examine the epidemiology and clinical spectrum of illness of CoV-HKU1, prospective studies have been carried out in Hong Kong since the SARS epidemic in 2003.

The genome of CoV-HKU1 found in the index patient is a 29,926-nucleotide, polyadenylated RNA. The genome organization is same as other coronaviruses. Sequence analysis showed that CoV-HKU1 has G+C content of 32%, which is the lowest among all known coronaviruses. The replicase 1a and 1b ORF occupy 21.5kb of the genome. The *pol* gene encoded protein with 928 amino acid has about 90% identity compare with other group 2 coronaviruses. *In silico* analysis of ORF1ab revealed a unique putative cleavage site of coronavirus HKU1 3C-Like protease (12).

1.2.2 Clinical features and pathogenesis

Previous study by Lau et al revealed the prevalence of CoV-HKU1 in acute respiratory tract infection is 0.3% in 2005 (1). Upper respiratory infection was the most common case of CoV-HKU1 infection, although pneumonia, acute bronchiolitis and asthmatic exacerbation also occurred. CoV-HKU1 accounts for 2.4% of community-acquired pneumonia with 2 genotypes in another study (13). Similar to other human coronavirus infections, cases of CoV-HKU1 associated pneumonia occurred during winter and spring. In contrast to HCoV-NL63, most patients with CoV-HKU1-associated pneumonia were old (80% were > 65 years old) and had major underlying diseases, especially those of the respiratory and cardiovascular system. Compared with SARS-CoV pneumonia, CoV-HKU1 associated pneumonia had relatively mild symptoms that were localized to the respiratory tract and were hospitalized only briefly. Lau's study also showed a high incidence of febrile seizures (50%) is associated with CoV-HKU1 infection, which was significantly higher than other viral respiratory infections (14).

1.2.3 Natural recombination leading to the generation of a different genotype of CoV-HKU1

In previous study on phylogeny of the RNA-dependent RNA polymerase (RdRp), spike (S) and nucleocapsid (N) genes of nine isolates of CoV-HKU1 recovered from patients with pneumonia, it was discovered that the sequences of the S and N genes fell into two distinct genotypes, with seven strains belonging to genotype A and two belonging to genotype B. On the other hand, for the RdRp gene, one of the two "genotype B" strains by the S and N sequences (from patient 8) was clustered with the other seven "genotype A" strains (13). Furthermore, the same phenomenon was also observed in their subsequent prospective study on CoV-HKU1 associated respiratory tract infections (14). Based on these observations, it was suspected that there is an additional CoV-HKU1 genotype which has arisen from recombination between genotypes A and B of CoV-HKU1. Subsequently, complete genome sequencing of 22 strains of CoV-HKU1 obtained from nasopharyngeal aspirates of patients wit respiratory tract infections showed that the 22 CoV-HKU1 strains fell into thre clusters (genotype A, 13 strains; genotype B, three strains and genotype C, si strains) (2). However, different phylogenetic relationships among the thre

clusters were observed in different regions of their genomes. Bootscan analysis showed possible recombination between genotypes B and C from nucleotide positions 11500 to 13000, corresponding to the nsp6/nsp7 junction, giving rise to genotype A; and between genotypes A and B from nucleotide positions 21500 to 22500, corresponding to the nsp16/HE junction, giving rise to genotype C. Multiple alignments further narrowed the sites of cross-over to a 143-bp region between nucleotide positions 11750 and 11892, and a 29-bp region between nucleotide positions 21502 and 21530. This was the first evidence for natural recombination in coronavirus associated with human infection. The results also suggested that analysis of a single gene is not sufficient for genotyping of CoV-HKU1, but would require amplification and sequencing of at least two gene loci, one from nsp10 to nsp16 (e.g RdRp or helicase) and another from HE to N (e.g spike or N)

1.3 Human coronavirus OC43 (HCoV-OC43)

1.3.1 Discovery of HCoV-OC43

Human coronavirus OC43 belongs to group 2 coronaviruses. It was first isolated in 1967 from volunteers at the Common Cold Unit in United Kingdom, it was initially propagated on ciliated human embryonic tracheal and nasal organ cultures (OC). HCoV-OC43 is responsible for around 10% - 30% of all common colds with the infections occurring mainly in winter (1, 14). The HCoV-OC43 virion is composed of 4 major structural proteins: spike (S), membrane (M), haemagglutinin-esterase (HE) and Nucleocapsid (N) protein. The N protein binds to the virion RNA to form the nucleocapsid of virion, S and M proteins are synthesized in the endoplasmic reticulum (ER). The integral M protein interacts with viral nucleocapsid and believed for the determination the site of virus budding. The S protein mediated binding of virons to host cell receptor (15, 16). During virion maturation and intracellular transport, S protein will be cleaved by host cell proteases to yield S1 (N-terminal half) and S2 (C-terminal half) subunits (17).

1.3.2 Clinical features and pathogenesis

HCoV-OC43 is mainly associated with upper respiratory tract disease, and causes up to 30% of common cold. Infection is often mild or sometimes sub-clinical. The major pathogenesis of HCoV is direct cytolysis of infected cells as a result of viral replication. Since HCoV can cause disease in re-infection soon after primary infection, it cannot be excluded that humoral antibody or some other component of the immune response has some role in causing or aggravating acute disease. They have the potential to cause disease by immunopathological mechanisms which already given of CNS disease in mice, rats and cats (18).

1.3.3 Complete genome sequence of HCoV-OC43

The first HCoV-OC43 complete genome sequence was published in 2004. The whole genome consist of 30,738 nucleotides with 36.8% G+C content. Interestingly, it shows more than 93% identity on nucleotide and more than 91% identity on protein to BCoV in all ORFs and was suggested that BCoV might have jumped the species barrier and became able to infect humans in around

1890, as a result emergence of new type of human coronavirus (HCoV-OC43) (1).

1.3.4 Genotype of HCoV-OC43

It is well known that all RNA virus have very high mutation rates. The average nucleotide base miss-incorporation rates are of the order of 10^{-4} to 10^{-5} (19). However, in 2004, a research team in France reported their clinical isolate of OC43 in 2001 designated Paris (AY585229) to show only 6 nucleotide differences leading to 1 amino acid change in S gene and 1 in N gene when compared with laboratory strain from ATCC (NC005147) which was isolated 34 years ago. Their results suggested that the genome of HCoV-OC43 is very stable which was unusual for an RNA virus (20). In 2005, a research group from Belgium reported very different findings. They determined the complete genome sequence of two contemporary human coronavirus OC43 strains which were detected in 2003 (BE03) and 2004 (BE04) respectively. The two genomes wer found to possess 193 (0.63%) nucleotide differences. They have also determine the complete S gene sequences on other three BE03 strains from year 2003 an two BE04 strains from year 2004. According to phylogenetic analysis, thre

distinct clusters were identified: (I) a cluster containing all ATCC strains and the Paris isolate, (II) a cluster containing all HCoV-OC43 BE03 strains from year 2003 and (III) a cluster containing all HCoV-OC43 BE04 strains from year 2004. Based on their findings, they suggested that two genetically distinct human coronavirus strains have circulated in Belgium in 2003 and 2004 respectively (21).

1.3 Scope of study

The present study aims to investigate for the presence of different genotypes among HCoV-OC43 strains in Hong Kong and to identify potential recombination events that could lead to the generation of novel genotypes, a situation analogous to that observed for CoV-HKU1. Twenty-three HCoV-OC43 strains identified from nasopharyngeal aspirates (NPAs) of patients with respiratory tract infections from 2004-2006 were subject to RT-PCR and sequencing of their complete RNA-dependent RNA polymerase (RdRp), spike (S) and nucleocapsid (N) genes. The RdRp gene is the gene encoding for the enzyme polymerase essential for replication. It is the most conserved region in the entire coronavirus genome. This gene is mainly used for primer design for the detection of virus. S gene is the largest structural gene which codes for the spike protein for the attachment of virus to host cell surface receptor. This gene is the most variable region and is commonly used for serological study. N gene is the second largest structural gene encoding for the nucleocapsid protein which is a component of the helical nucleocapsid and is thought to bind to the viral genomic RNA. It is relatively conserved among the four major structural proteins. In this study, all the three genes were used based on their function and

genetic structure. Multiple alignments and phylogenetic analysis were also performed.

Chapter 2: Materials and methods

2.1 Human coronavirus OC43 strains

All 23 HCoV-OC43 strains were detected from the nasopharyngeal aspirates (NPAs) of patients with respiratory tract infections in Hong Kong in 15-month (November 2004 - January 2006) (Table 1) (12, 13, 14).

2.2 RNA extraction

Viral RNA was extracted from NPAs using QIAamp Viral RNA Mini Kit (QIAgen, Hilden, Germany) within 10 hours upon receipt of specimens. 560 µl of prepared buffer AVL containing carrier RNA was pipetted into a 1.5 ml microcentrifuge tube. 140 µl of sample NPA was added and mixed by pluse vortexing for 15s. The mixture was incubated at room temperature for 10mins and mixing during incubation period. 560 µl of absolute ethanol was then added to the sample and mixed by pulse vortexing for 15s. 630 µl of the reaction mixture was added to the QIAamp spin column and centrifuge at 6000 x g for 1min. the filtrate was discarded and this step was repeated until all the samples

had applied to the column. After that, 500 µl of AW1 wash buffer was added and centrifuge at 6000 x g for 1 min. the filtrate was discarded and 500 µl of AW2 wash buffer was added and centrifuge at 16100 ×g for 3 mins. Discard the filtrate and 60 µl of elution buffer AVE was added. The sample was incubated at room temperature for 1 min. and centrifuged at 6000 ×g for 1 min. the eluted RNA (template for RT-PCR) was collected and stored immediately at -70°C until use.

2.3 Reverse transcription

Reverse transcription was performed using random hexamers and the SuperScript III kit (Invitrogen, San Diego, CA, USA) each reaction mixture contained 2 µl 5× FS buffer, 0.5 µl DTT, 0.025 µl random hexamers (50 ng/µl), 0.5 µl dNTP, 0.975 µl DEPC treated water, 0.5 µl SuperScript III and 5.5 µl RNA. The reaction mixture was incubated at 20°C for 10 mins and then 50°C for 50mins. After that, the reaction was inactivated by heating at 94°C for 3 mins. The cDNA (template for PCR) was stored immediately at -20°C until use.

2.4 PCR for the complete RdRp, S and N genes of HCoV-OC43 strains

Polymerase chain reaction was performed using iProof High Fidelity DNA polymerase (Bio-Rad, CA, USA) each reaction mixture contained 5 μl Buffer, 0.5 μl dNTP, 1.25 μl forward primer (Table 2), 1.25 μl reverse primer (Table 1), 15.75 μl DEPC treated water, 0.25 μl iProof DNA polymerase and 1 μl template. The mixtures were amplified in 40 cycles of 98°C for 30s, 98°C for 10s min and 55°C for 30s, and a final extension at 72°C for 10 min. The amplified products were detected by agarose gel electrophoresis.

2.5 DNA sequencing

The PCR products were gel purified using the QIAquick gel extraction kit (QIAgen, Hilden, Germany). Both strands of all PCR products were sequenced twice with an ABI Prism 3700 DNA Analyzer (Applied Biosystems, Foster City, CA, USA), using the PCR primers.

2.6 Sequence analysis

The nucleotide and the deduced amino acid sequences of the RdRp, S and N genes were manually edited to produce final sequence and compared to those of ATCC, Paris, BE03 and BE04 strains. Phylogenetic tree construction was performed using neighbor-joining method with ClusterX v.1.81. The corresponding nucleotide sequences of CoV-HKU1 (GenBank accession no. AY597011) were used as outgroups.

Chapter 3: Results

3.1 Sequencing of the complete RdRp, S and N genes

The complete RdRp, S and N genes of HCoV-OC43 from the 23 NPAs were amplified and sequenced. The sequences were compared to those of HCoV-OC43 strain with sequences available in GenBank. The nucleotide substitutions among the different strains in RdRp, S and N genes are shown in Fig. 1, 2 and 3 respectively. Phylogenetic trees are shown in Fig. 4.

3.2 Phylogenetic analysis of S genes

Analysis of the S genes showed that five of the 23 sequences possessed high nucleotide identities (99.1-99.5%) to BE03 strains while the other 18 sequences possessed high nucleotide identities (99.8-99.9%) to BE04 strains. Phylogenetic analysis showed the existence of three clusters of sequences, all with high bootstrap values of 1000. One cluster (clade A) was formed by the two ATCC strains and the Paris strain. Five (HK04-14, HK04-20, HK04-21, HK04-23 and HK04-32 from year 2004) of the 23 sequences formed a cluster (clade B) with

BE03 strains; while the other 18 sequences (from years 2004, 2005 and 2006) formed another cluster with BE04 strains (clade C). Multiple alignments revealed that there were 239 nucleotide substitutions among the S genes of the different HCoV-OC43 strains.

3.3 Phylogenetic analysis of N genes

Phylogenetic analysis of the N genes of the 23 strains showed similar results to that of the S genes. The N sequences of the five strains from year 2004 that belonged to clade B in their S genes possessed 99.3-99.8% nucleotide identities to that of the BE03 strain, forming a cluster (clade B) upon phylogenetic analysis. The other 18 sequences possessed 99.6-99.8% nucleotide identities to that of strain BE04, forming another cluster (clade C), while the ATCC and Paris strains belonged to a separate cluster (clade A). Multiple alignments revealed that there were 34 nucleotide substitutions among the N genes of the different HCoV-OC43 strains.

3.4 Phylogenetic analysis of RdRp genes

For the RdRp gene, the 23 sequences possessed 99.6% to 99.9% nucleotide identity to those of strains BE04 and BE03 from Belgium. Phylogenetic analysis showed the existence of three clusters of sequences, all with high bootstrap values of >800. One cluster was formed by two ATCC strains and the Paris strain. The five strains that belonged to clade B in their S and N genes were also clustered with the BE03 strain. The nucleotide sequences of 17 of the 18 strains that belonged to clade C in their S and N genes were also clustered together. However, the sequence of one (HK04-12 from year 2004) of the strains of clade C in its S and N gene and that of the BE04 strain were not clustered with other clade C strains, but fell into the cluster formed by the clade B strains. Multiple alignments revealed that there were 34 nucleotide substitutions among the RdRp genes of the different HCoV-OC43 strains.

3.5 Comparison of "unusual" strains

Comparison of the sequences of the HK04-12 strain and BE04 strain w those of other strains in the three clades by multiple alignment revealed that t

two "unusual" strains displayed high nucleotide similarity to clade B in their RdRp genes, but to clade C in their S and N genes, which is in line with results obtained by phylogenetic analysis. For RdRp genes, the HK04-12 strain and BE04 strain possessed identical nucleotide sequences to those of clade B strains, but 13 and 3 nucleotide substitutions when compared to those of clade C strains respectively. For S genes, the HK04-12 strain and BE04 strain possessed 2 and 0 nucleotide differences to those of clade C strains, but 92 and 99 nucleotide substitutions when compared to those of clade B strains respectively. For N, the HK04-12 strain and BE04 strain possessed identical nucleotide sequences to those of clade C strains, but 5 and 8 nucleotide substitutions when compared to those of clade B strains respectively.

Chapter 4: Discussion

Coronaviruses are unique in having high frequency of homologous RNA recombination, as a result of random template switching during RNA replication, thought to be mediated by a "copy-choice" mechanism (22, 23, 24, 25, 26, 27). In feline coronavirus (FCoV), it has been documented that FCoV type II strains originated from a double recombination between FCoV type I and canine coronavirus, and the site of recombination has been pinpointed to a region of about 50 nucleotides in the M gene by multiple alignment (28). As for recombination between different strains of MHV, in vitro studies have shown variations in both sites and rates of recombination, with the S gene having a frequency three fold that of the polymerase gene (26, 29).

A previous study on nine strains of CoV-HKU1 from patients with pneumonia, has identified two distinct genotypes of CoV-HKU1 upon S and N gene sequence analysis, with seven strains belonging to genotype A and tv belonging to genotype B (13). However, one of the two "genotype B" strains l the S and N sequences (from patient 8) was clustered with the other sev "genotype A" strains. Furthermore, the same phenomenon was also observed

CoV-HKU1 strains detected from our subsequent prospective study on CoV-HKU1 associated respiratory tract infections (14). Based on these observations, an additional CoV-HKU1 genotype was suspected to have arisen from recombination between genotypes A and B of CoV-HKU1. In their subsequent study on 22 genomes of CoV-HKU1, they documented for the first time evidence for natural recombination in coronavirus associated with human infection (E). Major recombination was detected among the three CoV-HKU1 genotypes, with the generation of genotype C from recombination between genotypes A and B. Using phylogenetic analysis, bootscan analysis and multiple alignments of the nucleotide sequences, the site of major recombination was localized to a stretch of 29 nucleotides in nsp16, just upstream to the stop codon of ORF1ab. In addition to this recombination site in nsp16, another potential recombination was identified at the end of nsp6 between genotypes B and C, giving rise to genotype A. The study also revealed potential recombination events that might have occurred in ORF1ab upstream to nsp5.

It has been well known that recombination is an important mechanism f the generation and evolution of virus genotypes (30, 31, 32). In the present stud it was shown that there were at least three distinct clusters of HCoV-OC43 upc S and N gene analysis. One cluster, clade A, was formed by the ATCC and Par

strains. Another cluster, clade B, was formed by five 2004 strains and the BE03 strains; while the third, clade C, was formed by 18 strains from years 2004-2006 and the BE04 strains. Although RdRp gene analysis also revealed the presence of three clusters, one (HK04-12) of the clade C strains by S and N gene and the BE04 strain were not clustered with clade C, but fell into the cluster formed by the BE03 and clade B strains by S and N gene. The results suggested the presence of at least three genotypes of HCoV-OC43, with strains potentially belonging to an additional genotype probably arising from recombination between strains from clade B and C, a situation similar to that reported for CoV-HKU1. Further studies on the complete genome sequences of these HCoV-OC43 strains are required to confirm this suspicion and localize the potential sites of recombination. Analysis of more HCoV-OC43 strains from other countries will also reveal the relative prevalence of the different genotypes in different localities and the presence of additional genotypes arising from other recombination events.

From the results of the present study, no association was observed betwee the genotypes and clinical characteristics of the patients. However, similar results from the previous study from Belgium, the different genotypes

HCoV-OC43 appeared to exhibit different temporal patterns. In the Belgium study, based on S gene analysis, they showed the presence of three phylogenetic clusters, the ATCC cluster and two clusters containing four 2003 (BE03) strains and three 2004 (BE04) strains respectively. In the present study, five of our 2004 strains were clustered with their BE03 strains forming clade B while 18 strains from year 2004-2006 were clustered with their BE04 strains forming clade C upon S and N gene analysis. This suggests that new genotypes of HCoV-OC43 have evolved over time, which could have been the result of immune selection, with the ATCC related strains being the most ancient and BE04 related strains the most recently evolved genotype detected so far. Therefore, it is expected that new genotypes of HCoV-OC43 will be observed in the future, including those that could have arisen from recombination.

Natural recombination in coronaviruses associated with human infection may have clinical, laboratory and epidemiology implication. A previous study on CoV-HKU1 proved that this phenomenon occurred which generated a new genotype. The generation of novel coronavirus genotypes by recombination analogous to reassortment in influenza virus, may allow the virus to escape immune protection, enhance pathogenicity or overcome inter-species barrier. As

for the labpratory implication, the diagnosis of human coronavirus is mainly done by RT-PCR. However, the sensitivity of this technique depends on the choice of primers. If recombination takes place at the site where original primers are designed, the primers may not be able to anneal and therefore may give false negative RT-PCR results. The implication of recombination events suggested that amplification and sequencing of a single gene is not sufficient to define the genotype. It would require amplification and sequencing of at least two gene loci on opposite side of the recombination site. Determining the genotypes of coronavirus is important for better understanding of its epidemiology.

Continuous studies are warranted to detect new genotypes and better understand the molecular evolution of HCoV-OC43 and other human coronaviruses. The data and results from previous studies on CoV-HKU1 also suggested that amplification and sequencing of a single gene may not be sufficient to define the genotypes of HCoV-OC43 and CoV-HKU1 (and probably other coronaviruses) as recombination events are not uncommon between th different genotypes. Instead, amplification and sequencing of at least two gen loci, probably one from ORF1ab e.g RdRp or helicase) and one from HE to I (e.g S or N), should be performed.

References

1. **Leen Vijgen, Els Keyaerts, Elien Moës, Inge Thoelen, Elke Wollants, Philippe Lemey, Anne-Mieke Vandamme, and Marc Van Ranst** 2005. Complete genomic sequence of human coronavirus OC43: Molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. J. Virol. **79:**1595-1604

2. **Patrick C. Y. Woom Susanna K. P. Lau, Cyril C. Y. Yip, Yi Huang, Hoi-Wah Tsoi, Kwok-Hung Chan, and Kwok-Yuen Yuen** 2006. Comparative analysis of 22 coronavirus HKU1 genomes reveals a novel genotype and evidence of natural recombination in coronavirus HKU1. J. Virol. **80:**7136-7145

3. **Susanna K. P. Lau, Patrick C. Y. Woo, Kenneth S. M. Li, Yi Huang, Ming Wang, Carol S. F. Lam, Huifang Xu, Rongtong Guo, Kwok-hun Chan, Bo-jian Zheng, Kwok-yung Yuen** 2007. Complete genom sequence of bat coronavirus HKU2 from Chinese horseshoe bats revealed much smaller spike gene with a different evolutionary lineage from the re

of genome. Virol. [In Press]

4. **Patrick C. Y. Woo, Ming Wang, Susanna K. P. Lau, Huifang Xu, Rosana W. S. Poon, Rongtong Guo, Beatrice H. L. Wong, Kai Gao, Hoi-wah Tsoi, Yi Huang, Kenneth S. M. Li, Carol S. F. Lam, Kwok-hung Chan, Bo-jian Zheng, and Kwok-yung Yuen** 2007. Comparative analysis of twelve genomes of three novel group 2c and group 2d coronavirus reveals unique group and subgroup features. J. Virol. **81:**1574-1585

5. **Lai, M. M., and K. V. Holmes.** 2001. Coronaviridae: the viruses and their replication, p. 1163-1185. *In* D. M. Knipe, P. M. Howley, D. E. Griffin, R. A. Lamb, M. A. martin, B, Roizman, and S. E. Straus (ed.), Fields virology, Lippincott Williams & Wilkins, Philadelphia, Pa.

6. **Gorbalenya AE, Snijder EJ, Spaan WJ** 2004. Sever acute respirat(syndrome coronavirus phylogency: toward consensus. J. Vi **78:**7863-7866.

7. **Jeffrey S. Kahn** 2006. The widening scope of coronaviruses. Curr. Opin. Pediatr. **18:**42-47

8. **Jeffrey S. Kahn, Kenneth McIntosh** 2005. History and recent advances in coronavirus discovery. Pediatr. Infect. Dis. J. **24(11)** Supplement:S223-227

9. **Esper F, Weibel C, Ferguson D, Landry ML, Kahn JS**. Coronavirus HKU1 infection in the United States. Emerg. Infect. Dis. [serial on the Internet]. 2007 http://www.cdc.gov/ncidod/EID/vol12no05/05-1316.htm

10. **Astrid Vabret, Julia Dina, Stephanie Gourain, Joelle Petitjean, Sandrine Corbet, and Francois Freymuth**. 2006. Detection of the new human coronavirus HKU1: a report of 6 cases. Clinl. Infect. Dis. **42:**634-637

11. **Sloots TP, McErlean P, Speicher DJ, Arden KE, Nissen MD, Macka IM**. 2006. Evidence of human coronavirus HKU1 and human bocavirus i Australian children. J. Clin. Virol. **35:**00-102

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

13. **Woo PC, Lau SK, Tsoi HW, Huang Y, Poon RW, Chu CM, Lee RA, Luk WK, Wong GK, Wong BH, Cheng VC, Tang BS, Wu AK, Yung RW, Chen H, Huan Y, Chan KH, Yuen KY**. 2005. Clinical and molecular epidemiological features of coronavirus HKU1-associated community-acquired pneumonia. J. Infect. Dis. **192:** 1898-1907

14. **Susanna K. P. Lau, Patrick C. Y. Woo, Cyril C. Y. Yip, Herman Tse, Hoi-wah Tsoi, Vincent C. C. Cheng, Paul Lee, Bone S. F. Tang, Chris H. Y. Cheung, Rodney A. Lee, Lok-yee So, Yu-lung Lau, Kwok-hung Chan and Kwok-yung Yuen** 2006. Coronavirus HKU1 and other coronaviru infections in Hong Kong. J. Clin. Microbiol. **44:**2063-2071

15. **Niemann, H., Boschek, B., Evans, D., Rosing, M., Tamura, T. & Klenk, H. D** 1982. Posttranslational glycosylation of coronavirus glycoprotein E1: inhibition by monensin. EMBO Journal **2:**1499-1504

16. **Delmas, B., Gelfi, J., L'Haridon, R., Vogel, L. K., Sjőström, H., Norĕn, O. & Laude, H.** 1992. Aminopeptidase N is a major receptor for the enteropathogenic coronavirus TGEV. Nature **357:**417-420

17. **Samir Mounir and Pierre J. Talbot** 1993. Molecular characterization of the S protein gene of human coronavirus OC43. J. Gen. virol. **74:**1981-1987

18. **J. M. Darville and E. O. Caul** 2002. Coronavirus: Upper respiratory tract disease p.546-550. *In* Dacid Greenwood, Richard C. B. Slack, John F. Peutherer, Medical Microbiology (6th Edition), Churchill Livingstone

19. **John Holland & Esteban domingo** 1998. Origin and evolution of viruses Virus Genes **16:**13-21

20. **Julien R. St-Jean, Hélène jacomy, marc Desforges, Astrid Bavre**

Francois Freymuth, and Pierre J. Talbot 2004. Human respiratory coronavirus OC43: Genetic stability and neuroinvasion. J Virol. **78:**8824-8834

21. **Leen Vijen, Els Keyaerts, Philippe Lemey, Elien Moës, Sandra Li, Anne-Mieke Vandamme, Marc Van Ranst** 2005. Circulation of genetically distinct contemporary human coronavirus OC43 strains. J. Virol. **337:**85-92

22. **Banner, L. R. and M. M. Lai**. 1991. Random nature of coronavirus RNA recombination in the absence of selection pressure. Virology **185**:441-445

23. **Copper, P. D., A. Steiner-Pryor, P. D. Scotti, and D. Delong**. 1974. On the nature of poliovirus genetic recombinants. J. Gen. Virol. **23**:41-49

24. **Kirkegaard, K. and D. Baltimore**. 1986. The mechanism of RNA recombination in poliovirus. Cell **47**:433-443

25. **Lai, M. M., R. S. Baric, S. Makino, J. G. Keck, J. Egbert, J. L**

Leibowitz, and S. A. Stohlman. 1985. Recombination between nonsegmented RNA genomes of murine coronaviruses. J. Virol. **56**:449-456

26. **Makino, S., J. G. Keck, S. A. Stohlman, and M. M. Lai**. 1986. High-frequency RNA recombination of murine coronaviruses. J. Virol. **57**:729-737

27. **van der Most, R. G., L. Heijnen, W. J. Spaan, and R. J. de Groot**. 1992. Homologous RNA recombination allows efficient introduction of site-specific mutations into the genome of coronavirus MHV-A59 via synthetic co-replicating RNAs. Nucleic Acids Res. **20**:3375-3381

28. **Herrewegh, A. A., I. Smeenk, M. C. Horzinek, P. J. Rottier, and R. J. de Groot**. 1998. Feline coronavirus type II strains 79-1683 and 79-1146 originate from a double recombination between feline coronavirus type I and canine coronavirus. J. Virol. **72**:4508-4514

29. **Fu, K. and R. S. Baric**. 1992. Evidence for variable rates of recombinatio in the MHV genome. Virology **189**:88-102

30. **Kalinina, O., H. Norder, S. Mukomolov, and L. O. Magnius**. 2002. A natural intergenotypic recombinant of hepatitis C virus identified in St. Petersburg. J. Virol. **76**:4034-4043

31. **Simmonds, P. and S. Midgley**. 2005. Recombination in the genesis and evolution of hepatitis B virus genotypes. J. Virol. **79**:15467-15476

32. **Steininger, C., B. Schmied, M. Sarcletti, M. Geit, and E. Puchhammer-Stockl**. 2005. Cytomegalovirus genotypes present in cerebrospinal fluid of HIV-infected patients. AIDS **19**:273-278

Table 1 Clinical characteristics of the 23 patients with HCoV-OC43 infections

Month / Year	Strain	Sex	Age	Diagnosis	Clade
Nov / 2004	HK04-11	F	9	Upper respiratory infection	C
Nov / 2004	HK04-12	M	35m	Upper respiratory infection	C[b]
Nov / 2004	HK04-13	F	1	Upper respiratory infection, febrile convulsion	C
Nov / 2004	HK04-14	M	24m	Upper respiratory infection, febrile convulsion	B
Nov / 2004	HK04-16	M	36	Pneumonia	C
Nov / 2004	HK04-18	F	82	COAD exacerbation	C
Nov / 2004	HK04-19	M	8m	NA[a]	C
Nov / 2004	HK04-20	F	24d	Upper respiratory infection	B
Nov / 2004	HK04-21	F	5	Upper respiratory infection, pneumonia	B
Nov / 2004	HK04-22	M	1	Upper respiratory infection, febrile convulsion	C
Nov / 2004	HK04-23	M	2	Upper respiratory infection	B
Nov / 2004	HK04-24	M	85	Upper respiratory infection, urinary tract infection	C
Nov / 2004	HK04-25	M	1m	viral infection	C
Nov / 2004	HK04-26	M	76	Upper respiratory infection	C
Nov / 2004	HK04-27	M	85	NA[a]	C
Dec / 2004	HK04-28	F	32	Upper respiratory infection	C
Dec / 2004	HK04-30	M	72	Upper respiratory infection	C
Dec / 2004	HK04-31	M	1	Pneumonia	C
Dec / 2004	HK04-32	F	46	Pneumonia	B
Dec / 2004	HK04-33	M	68	COAD exacerbation	C
Jan / 2005	HK05-01	M	11m	Upper respiratory infection, NSAID induced angioedema	C
Dec / 2005	HK05-02	M	88	Pneumonia	C
Jan / 2006	HK06-01	F	84	Upper respiratory infection	C

[a]NA, not available

[b]Strain HCoV-OC43-HK04-12 belongs to clade C upon S and N gene analysis, but to clade B upon RdRp gene analysis

Table 2 Primers used for sequencing of HCoV-OC43 complete RdRp, S and N genes

Primer name	Primer direction, sequence (5′ - 3′)	Gene target
LPW 1114	Forward, CYGTTTGTATATATTGCCGC	RdRp
LPW 1118	Reverse, TGATTATCAAGTGTTAAAAC	RdRp
LPW 5188	Forward, GCCCCTAGTTAGCGCTACTGAGTT	RdRp
LPW 3709	Reverse, ACTTAGGATAATCCCAACCCAT	RdRp
LPW 3064	Forward, CTGGGATGATATGTTACGCCG	RdRp
LPW 2579	Reverse, GTGTGTTGTGAACARAAYTCRTG	RdRp
LPW 3064	Forward, CTGGGATGATATGTTACGCCG	RdRp
LPW 2579	Reverse, GTGTGTTGTGAACARAAYTCRTG	RdRp
LPW 1223	Forward, TAAGTGCCTTTCAACAGGT	RdRp
LPW 1127	Reverse, KGCCTTTTGCGTTTCTGC	RdRp
LPW 1162	Forward, CCYRTTTGTRTGTATGATCC	S
LPW 1166	Reverse, YGCATAAAAAGTACCACC	S
LPW 1261	Forward, CTRCTATARYTATAGGTAGT	S
LPW 2094	Reverse, GCCCAAATTACCCAATTGTAGG	S
LPW 2095	Forward, TGATGCTGCTAAGATATATGG	S
LPW 2098	Reverse, ATTCCGARATAGCAATGCTGG	S
LPW 1839	Forward, ATCTTTTGTATGATTCTAATGG	S
LPW 1178	Reverse, GACACCAAGMCCATTAAT	S
LPW 1177	Forward, CWGCAGGTGTRCCATTTT	S
LPW 1183	Reverse, CCACAYTTCTTRAAACAAC	S
LPW 1275	Forward, TRAAATGGCCTTGGTATGT	S
LPW 1189	Reverse, TKWMWAGGAACTCTACAATA	S
LPW 6547	Forward, CTTCAAAGAACTATGGCATT	S
LPW 6548	Reverse, GACTGCAAATAGCCCAAATT	S
LPW 1192	Forward, AACCCMGAAACAAACAAC	N
LPW 1045	Reverse, GCAAGAATGGGGAACTGTGG	N
LPW 1195	Forward, GAGAGGCCCTAATCAGAA	N
LPW 1198	Reverse, TYAACTTCATTCATTTACTA	N

Figure 1 Mutations among different strains in RdRp gene

Nucleotide positions

Clade / Strain	36	185	192	425	467	578	611	720	878	881	884	930	934	1025	1026	1123	1202	1379	1445	1508	1589	1625	1643	1664	1814
A	A	C	A	T	C	G	T	A	C	T	T	A	T	C	C	C	A	G	G	C	T	T	T	T	T
B	G	C	A/G	T	T	G	T/A	A/G	C/T	G	T/C	A	T	C	C/A	C/G	A/G	G/A	T	C	T	T/C	T/C	T	C
C	G	C/T	A	T	C	G/A	T	A	C	T	T	G	T/C	C/T	C	C	A	G	G	T	T/C	T	T	C	T
HCoV-OC43-HK04-12	G	C	A	T	T	G	T	A	C	G	T	A	T	C	C	C	A	G	T	C	T	C	C	T	C
HCoV-OC43-BE04-AY903460	G	C	A	T	T	G	T	A	C	G	T	A	T	C	C	C	A	G	T	C	T	T	T	T	C

Nucleotide positions

Clade / Strain	1872	1925	1964	2042	2195	2303	2423	2732	2777
A	C	T	C	A	G	C	C	C	T
B	C/A	T	C	A	G	C/T	T	C	C
C	C	C	C/T	A/G	A	C	T	C/T	C
HCoV-OC43-HK04-12	A	T	C	A	G	T	T	C	C
HCoV-OC43-BE04-AY903460	A	T	C	A	G	C	T	C	C

The nucleotide positions correspond to those of strain HCoV-OC43-ATCC (AY391777)

Figure 2 Mutations among different strains in S gene

Nucleotide positions

Clade / Strain	60	64	68	73	74	75	76	77	78	79	80	-	-	-	-	-	-	-	-	-	87	88	92	103	109
A	G	A	C	A	C	T/-	A/-	G/-	T/-	T/-	A	-	-	-	-	-	-	-	-	-	T	G	A	C	C
B	G	C	T	A/-	C/-	T/-	A	G	T	C	T/G	A/C	A	G	G	T	A	G	C	T	T	A	G	C/T	T
C	T	C	T	A	C	T/-	A	G	G	C	T	A/C	A	G	G	T	A	G	C	T	T/C	A	G	C	T
HCoV-OC43-HK04-12	T	C	T	A	C	T	A	G	G	C	T	A	A	G	G	T	A	G	C	T	T	A	G	C	T
HCoV-OC43-BE04-AY903460	T	C	T	A	C	T	A	G	G	C	T	A	A	G	G	T	A	G	C	T	T	A	G	C	T

Nucleotide positions

Clade / Strain	122	183	186	196	213	254	257	258	259	260	267	269	271	343	352	420	441	446	454	460	461	513	525	534	543
A	A	T	T	C	T	G	T	A	C	T	C	G	C	A	C	G	C	C	G	T/G	A	C	A	T	C
B	A/C	T	T/A	C/A	T/C	C	A	C	A	A	C/T	C	T	T	G	G/A	T	C	G	G	T	C/T	T	T	A
C	A	T	T	C	C	C	A	C	A	A	C	C	T	T	G	G	T	C	G/A	G	T	T	T	T/C	A
HCoV-OC43-HK04-12	A	C	T	C	C	C	A	C	A	A	C	C	T	T	G	G	T	T	G	G	T	T	T	T	A
HCoV-OC43-BE04-AY903460	A	T	T	C	C	C	A	C	A	A	C	C	T	T	G	G	T	C	G	G	T	T	T	T	A

Figure 2 Mutations among different strains in S gene (Con't)

Nucleotide positions

Clade / Strain	544	552	555	556	557	574	588	634	635	672	687	700	741	753	766	776	782	784	785	786	787	788	789	790	791
A	C	T	T	C	G	T	T	G	A	T	T	G	G	T	C	A	A	G/-	T/-	T/-	A/-	A/-	G/-	A/-	A/-
B	C	T/C	C	T	T	T/A	T	A	C	T/C	T/C	G/T	A	T	C	T	G	C	G	T	G	A	T	A	T
C	T	C	C	T	T	T	T/C	A	C	T	C	T	A	T/C	C/T	T	G	C	T	T	G	A	T	A	T
HCoV-OC43-HK04-12	T	C	C	T	T	T	T	A	C	T	C	T	A	T	C	T	G	C	T	T	G	A	T	A	T
HCoV-OC43-BE04-AY903460	T	C	C	T	T	T	T	A	C	T	C	T	A	T	C	T	G	C	T	T	G	A	T	A	T

Nucleotide positions

Clade / Strain	792	793	794	795	796	822	963	984	1002	1007	1009	1056	1228	1254	1277	1278	1344	1386	1399	1411	1412	1418	1422	1423	1432
A	T/-	G/-	G/-	T/-	T/C	T	C	C	T	T/C	A	G	G	T	T	T	T	T	G	A	A	G	T	G	C
B	T	G	G	T	T	C	C/T	T	C	C	G	G	G	T	T	T	T	T/C	A	A	A	A	T	G	C/T
C	T	G	G	T	T	C	C	C	C	C	G	G/A	C	C	C	C	C	T	G	G	T	A	T/C	A	T
HCoV-OC43-HK04-12	T	G	G	T	T	C	C	C	C	C	G	G	C	C	C	C	C	T	G	G	T	A	T	A	T
HCoV-OC43-BE04-AY903460	T	G	G	T	T	C	C	C	C	C	G	G	C	C	C	C	C	T	G	G	T	A	T	A	T

Figure 2 Mutations among different strains in S gene (Con't)

Nucleotide positions

Clade / Strain	1437	1440	1443	1444	1445	1458	1464	1470	1476	1479	1488	1502	1506	1507	1512				1514	1515	1516	1517	1518	1519	1520
A	T	T	T	G	A	A	C	C	T	T	C	T	T	G	G	-	-	-	G	T	G	T	A	G	G
B	C	T	T	G	A	C	T	T	T	T/C	T	T	C	G/A	T	T	G	T	G	T	G	T	A	G	G
C	T	T/A	C	A	G	A	C	C	C	T	C/T	T/C	T	G	C	-	-	-	-	-	-	-	-	-	-
HCoV-OC43-HK04-12	T	T	C	A	G	A	C	C	C	T	C	T	T	G	C	-	-	-	-	-	-	-	-	-	-
HCoV-OC43-BE04-AY903460	T	T	C	A	G	A	C	C	C	T	C	T	T	G	C	-	-	-	-	-	-	-	-	-	-

Nucleotide positions

Clade / Strain	1437	1440	1443	1444	1445	1458	1464	1470	1476	1479	1488	1502	1506	1507	1512				1514	1515	1516	1517	1518	1519	1520
A	T	T	T	G	A	A	C	C	T	T	C	T	T	G	G	-	-	-	G	T	G	T	A	G	G
B	C	T	T	G	A	C	T	T	T	T/C	T	T	C	G/A	T	T	G	T	G	T	G	T	A	G	G
C	T	T/A	C	A	G	A	C	C	C	T	C/T	T/C	T	G	C	-	-	-	-	-	-	-	-	-	-
HCoV-OC43-HK04-12	T	T	C	A	G	A	C	C	C	T	C	T	T	G	C	-	-	-	-	-	-	-	-	-	-
HCoV-OC43-BE04-AY903460	T	T	C	A	G	A	C	C	C	T	C	T	T	G	C	-	-	-	-	-	-	-	-	-	-

Figure 2 Mutations among different strains in S gene (Con't)

Nucleotide positions

Clade / Strain	1521	1522	1523	1524	1525	1573	1580	1585	1587	1593	1596	1598	1599	1601	1619	1620	1621	1622	1623	1624	1629	1630	1648	1665	1678
A	T	A	G	T	G	T	C	G	T	G	C	C	T	C	A/-	A/-	G/-	C/-	T/-	A/-	T	A	A	C	T
B	T	A	G	T	G	T/C	C	C	C	G	C	A	T	C	-	-	-	-	-	-	T	C	A	T	T/C
C	-	-	-	-	-	T	C/A	G	T	A	T	C	G/T	T	A	A	G	C	T	C	T/C	A	A/T	T	T
HCoV-OC43-HK04-12	-	-	-	-	-	T	C	G	T	A	T	C	G	T	A	A	G	C	T	C	T	A	A	T	T
HCoV-OC43-BE04-AY903460	-	-	-	-	-	T	C	G	T	A	T	C	G	T	A	A	G	C	T	C	T	A	A	T	T

Nucleotide positions

Clade / Strain	1683	1714	1715	1720	1725	1733	1779	1806	1810	1812	1826	1869	1873	1874	1875	1894	1902	1913	1930	1992	1998	2022	2072	2076	2079
A	G	G	G	T	T	G	A	T	T	G	G	C	A	T	T	T	C	T	G	T	C	C	T	C	T
B	A	G	G	C	C	A	G	A	T/C	G	G	C	A	T/A	T/A	T/C	T	C	G	T	T	C	T	C	T
C	G	A	A	T	T	A	A	C	T	G/A	A	A	G	A	A	T	C	C	G/T	C	C	C/T	A	T	C
HCoV-OC43-HK04-12	G	A	A	T	T	A	A	C	T	G	A	A	G	A	A	T	C	C	G	C	C	C	A	T	C
HCoV-OC43-BE04-AY903460	G	A	A	T	T	A	A	C	T	G	A	A	G	A	A	T	C	C	G	C	C	C	A	T	C

Figure 2 Mutations among different strains in S gene (Con't)

Nucleotide positions

Clade / Strain	2101	2151				2234	2296	2304	2358	2367	2388	2429	2448	2454	2495	2510	2583	2602	2625	2629	2652	2679	2686	2787	2790
A	C	A	-	-	-	C	G	T	C	A	T	A/T	C	T	A	C	C	G	T	C	C	T	C	T	T
B	C	A	-	-	-	C	A	T	T	A	C	A	C/T	T	A/T	C	C	G	C	C	C	T	T	C	T
C	C/T	A/T	C	G	T	C/A	A	C	T	G	T	A	C	C	A	T	T	G/T	T/C	C/T	T	T/C	C	C	T/C
HCoV-OC43-HK04-12	C	A	C	G	T	C	A	C	T	G	T	A	C	C	A	T	T	G	T	C	T	T	C	C	T
HCoV-OC43-BE04-AY903460	C	A	C	G	T	C	A	C	T	G	T	A	C	C	A	T	T	G	T	C	T	T	C	C	T

Nucleotide positions

Clade / Strain	2792	2811	2859	2874	2878	2896	2907	2922	2934	2942	2994	3010	3051	3070	3082	3199	3205	3231	3279	3303	3375	3393	3417	3473	3503
A	A	A	C	T	C	A/T	C	C	A	T/C	T	A	T	T	G	G	T	T	T	T	C	G	T	G	G
B	C	T	T	T	C/T	A	C	T	G	C	T	A/C	T	G/C	G/C	G/T	T/C	T	T/C	T	C	G/A	T/C	G/A	A
C	A	A	C	C	C	A	C/T	C	A	C	T/C	A/T	T/G	C	C	G	T	T/C	C	T/C	C/T	G	T	A	A
HCoV-OC43-HK04-12	A	A	C	C	C	A	T	C	A	C	T	T	T	C	C	G	T	T	C	T	C	G	T	A	A
HCoV-OC43-BE04-AY903460	A	A	C	C	C	A	T	C	A	C	T	A	T	C	C	G	T	T	C	T	C	G	T	A	A

Figure 2 Mutations among different strains in S gene (Con't)

	Nucleotide positions													
Clade / Strain	3532	3536	3639	3654	3699	3725	3738	3747	3798	3852	3861	3987	3988	4083
A	G	G	T	C	T	A	T	T	T	C	G	T	A	C
B	G/A	G	T/C	C/A	T	A/G	T/C	T/C	T	C/T	G/T	T	A	C
C	A	G/A	T	C	T/C	A	C	C	T/C	C	T	T/C	A/T	C/T
HCoV-OC43-HK04-12	A	G	T	C	C	A	C	C	C	C	T	T	A	C
HCoV-OC43-BE04-AY903460	A	G	T	C	T	A	C	C	T	C	T	T	A	C

The nucleotide positions correspond to those of strain HCoV-OC43-ATCC (AY391777)

Figure 3 Mutations among different strains in N gene

Nucleotide positions

Clade / Strain	54	150	230	242	345	397	411	414	427	433	521	530	551	552	642	671	672	676	701	721	747	807	893	983	1032
A	T	C	A	T/C	C	C	G	C	G	G	C	C	A	G	T	C	T	T	A	C	T	G	G	G	C
B	G	T	C	C	C	C	G	C/T	G	G/A	C	C/T	A	G	T	C/T	T	T	A	C/T	T	A	G	G	T
C	G	T	A	C	C/T	C/T	G/A	C	A	G	C/T	C	A/C	G/A	T/C	C	T/C	T/A/C	A/G	C	C	A	G/T	A	T
HCoV-OC43-HK04-12	G	T	A	C	C	C	G	C	A	G	C	C	A	G	T	C	T	T	A	C	C	A	G	A	T
HCoV-OC43-BE04-AY903460	G	T	A	C	C	C	G	C	A	G	C	C	A	G	C	C	C	C	A	C	C	A	G	A	T

Nucleotide positions

Clade / Strain	1056	1062	1095	1121	1134	1159	1169	1271	1322
A	C	C	T	A	T	A	G	G	A
B	C/T	C/T	C	G	T	A	G	T	A
C	C	C	C	G	T/C	A/G	G/T	T	T
HCoV-OC43-HK04-12	C	C	C	G	T	A	G	T	T
HCoV-OC43-BE04-AY903460	C	C	C	G	T	G	G	T	T

The nucleotide positions correspond to those of strain HCoV-OC43-ATCC (AY39177

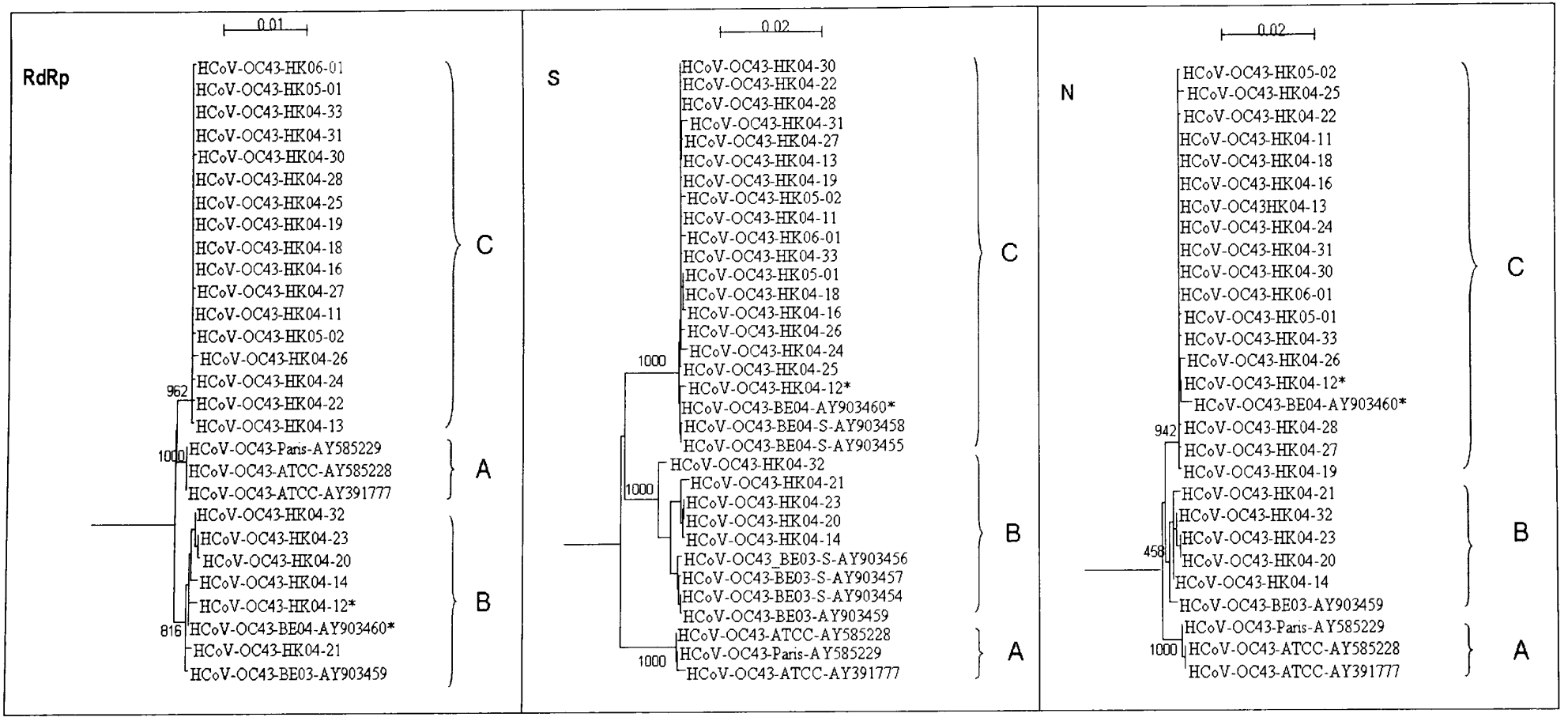


Figure 4 Phylogenetic trees of complete RdRp, S and N genes. The trees were constructed by the neighbor-joining method using Jukes-Cantor correction and bootstrap values calculated from 1,000 trees. 2,783, 4,086 and 1,347 nucleotide positions in RdRp, S and N respectively, were included in the analysis. * HCoV-OC43-HK04-12 and HCoV-OC43-BE04-AY903460 belong to clade C upon S and N gene analysis, but to clade B upon RdRp gene analysis