



Prevalence and Genetic Diversity Analysis of Human Coronavirus OC43 among Adult Patients with Acute Respiratory Infections in Beijing, 2012

Qin Hu^{1,2,3}, Roujian Lu^{2,3}, Kun Peng^{3,3}, Xijie Duan^{4,3}, Yanqun Wang², Yanjie Zhao^{1,2}, Wen Wang², Yongliang Lou^{1*}, Wenjie Tan^{1,2*}

1 Institute of Medical Virology, Wenzhou Medical University, Zhejiang, China, **2** Key Laboratory of Medical Virology, Ministry of Health, National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention, Beijing, China, **3** Beijing Sixth Hospital, Beijing, China, **4** Peking Union Medical College Hospital, CAMS, Beijing, China

Abstract

To determine the prevalence, epidemiology and genetic diversity of human coronavirus OC43 (HCoV-OC43) among adult patients with acute respiratory infections (ARI) in Beijing, five hundred and fifty-nine nasopharyngeal swab samples were collected from adult patients with ARI in Beijing. The prevalence of HCoV-OC43 infection among these patients was assessed using two different OneStep reverse transcription polymerase chain reaction (RT-PCR) assays. The epidemiological profiles of the patients with HCoV-OC43 infection were described. Partial S and N genes of HCoV-OC43 circulating strains were sequenced followed by phylogenetic analysis and amino acid alignment. Our results showed that the prevalence of HCoV-OC43 infection was 12.52% (95% CI: 9.78–15.26%), and the epidemic peak occurred in autumn. Fifty partial S and 40 partial N fragments were obtained from these patients. Phylogenetic analysis based on neighbour-joining method showed that at least three distinct clusters (A, B, C/D) of HCoV-OC43 strains were circulating among adult patients with ARI in Beijing. In addition, some novel unique clusters (UNT) of HCoV-OC43 were found in the S- and N-based phylogenetic trees. Furthermore, consensus amino acids substitutes for each cluster were also found after alignment of partial S or N sequence coding region in this study. In conclusion, we herein describe the prevalence of HCoV-OC43 among adult patients and provide substantial evidence for the genetic diversity of HCoV-OC43 circulating in Beijing.

Citation: Hu Q, Lu R, Peng K, Duan X, Wang Y, et al. (2014) Prevalence and Genetic Diversity Analysis of Human Coronavirus OC43 among Adult Patients with Acute Respiratory Infections in Beijing, 2012. PLoS ONE 9(7): e100781. doi:10.1371/journal.pone.0100781

Editor: Shan-Lu Liu, University of Missouri, United States of America

Received: April 2, 2014; **Accepted:** May 27, 2014; **Published:** July 2, 2014

Copyright: © 2014 Hu. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: This work was supported by grants from the State Megaproject for Infectious Disease Research of China (2011ZX10004-001, 2013ZX10004601) and National Science Funds of China (81100062). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: tanwj28@yahoo.cn (WT); louyongliang2013@163.com (YL)

† These authors contributed equally to this work.

Introduction

Coronaviruses are ubiquitous in the environment, and they infect various animal species causing lesions of varying severity in different organs [1,2]. The virions of coronaviruses are enveloped, roughly spherical particles that possess a linear, nonsegmented, positive-sense RNA genome of up to 32 kb [2,3], which is the largest genome among all known RNA viruses.

To date, six species of human coronavirus (HCoVs) have been described [3]. With the exception of SARS coronavirus (SARS-CoV) (a *Betacoronavirus*) and the recent Middle East respiratory syndrome coronavirus (MERS-CoV) (a *Betacoronavirus*) [4,5,6], which cause severe acute respiratory syndrome, the other four common and epidemic HCoVs are HCoV-229E (an *Alphacoronavirus*), HCoV-OC43 (a *Betacoronavirus*), HCoV-NL63 (an *Alphacoronavirus*), and HCoV-HKU1 (a *Betacoronavirus*) [7–10]. HCoV-OC43 was first found in a patient with a cold in 1967 [8]. Virologists paid little attention to the virus, however, until SARS-

CoV emerged in 2002–2003 in China [4]. Outbreaks of HCoV-OC43 occur frequently worldwide and are often associated with acute upper respiratory tract infections (URTI) [11–16]. Such infections might be dangerous to infants, elderly individuals, and immunocompromised patients [17,18]. HCoV-OC43 is also suspected of causing gastrointestinal and central nervous system diseases [19–21].

Phylogenetic methods have enabled the classification of HCoV-OC43 as part of the 2b subgroup of *Betacoronaviruses* [4]. The time of divergence was first extensively used for coronaviruses after the SARS epidemic for estimating the date of interspecies jumping of BCov or HCoV-OC43 ancestor from bovine to humans [22,23]. HCoV-OC43 has a high frequency of mutation, using both base substitution and homologous RNA recombination [26,27]. Among the structural viral proteins, the spike (S) protein is used for receptor binding and viral entry and has the most variable sequence within the coronavirus genome [24–27]. S protein is cleaved into two subunits, S1 and S2 [2]. The S1 gene has been

reported to show the dominant genetic diversity or heterogeneity among circulating HCoV strains [24,27]. The nucleocapsid phosphoprotein (N) gene is another common target for serological assays [28,29]. In view of the published data and the role of structural proteins, analysis of the S and N genes would be appropriate for determination of the genetic diversity of HCoV variants. In addition, the use of multiple primer sets targeting the S and N genes is recommended for diagnosis of all types of HCoV infection [30–36].

The prevalence of HCoV-OC43 among respiratory tract samples from patients with ARI has been reported in China using reverse transcription-polymerase chain reaction (RT-PCR) or real-time RT-PCR assays targeting a single viral gene sequence [34,35,36]. Molecular epidemiology and evolution studies of HCoV-OC43 have recently been conducted in some areas [22–25], and the data revealed at least three distinct clusters based on RdRp, S, and N genes and the emergence of a novel genotype, D, arising from natural recombination events between genotypes B and C [25]. However, the molecular epidemiology and genetic diversity of HCoV-OC43 variants among adult patients with ARI in Beijing are not understood fully. To address these issues, this report includes investigation of the prevalence of HCoV-OC43 among 559 respiratory samples collected in Beijing as assessed using two RT-PCR assays targeting the S and N genes, respectively, followed by direct sequencing of PCR products and genetic diversity analysis of partial S and N genes of HCoV-OC43 variants. The epidemiological and clinical characteristics of patients were also evaluated and their relationships with genetic diversity determined. This is the first report of the molecular prevalence and genetic diversity of HCoV-OC43 variants among adult patients with ARI in Beijing.

Materials and Methods

Clinical specimens

From December 2011 to December 2012, 559 nasopharyngeal swab samples were collected from adult patients with acute respiratory tract infections (ARI) who visited the fever clinics of Peking Union Medical College Hospital and Peking Sixth Hospital. All patients over 14 years of age were selected to enroll according to a set of criteria that included respiratory symptoms, body temperature above 37.5°C, and a normal or low leukocyte count ($\leq 10^{10}$ /L). Clinical information and laboratory results were recorded for each patient using a standardised form. Clinical samples were placed in viral transport medium and then stored at -80°C .

All aspects of the study were performed in accordance with national ethics regulations and were approved (2010IR0128) by the Institutional Review Boards of the Centre for Disease Control and Prevention of China, as well as the Ethics Committee of Wenzhou Medical University, Peking Union Medical College Hospital or Beijing Sixth Hospital as applicable. Participants provided written informed consent regarding the purpose of the study and their right to keep information confidential. We also obtained written informed consent from the next of kin or guardians on behalf of the minors enrolled in this study.

Nucleic acid detection and sequencing of the S and N genes of HCoV-OC43

RNA was isolated from nasopharyngeal swab specimens using a QIAamp viral RNA Mini kit (QIAGEN, Valencia, CA) according to the manufacturer's instructions. To design primers for RT-PCR to detect HCoV-OC43, S and N gene sequences were retrieved from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) and

analysed using the Primer Express software version 2.0 (Applied Biosystems, Foster city, CA, USA) to allocate target sites. Screening of extracted viral RNA was conducted by OneStep RT-PCR. A 714-bp fragment (24, 204–24, 918 nt) of the S1 gene was generated using the forward primer 5'-GAACATATGG-CATTTGGATACAGG-3' and the reverse primer 5'-AT-GACTGCAAATAGCCCAAATT-3'; A 749-bp fragment (29,967–30,716 nt) of the N gene was generated using the forward primer 5'-GTAAGAGAGGCCCTAATCAGAA-3' and the reverse primer 5'-CTTCATTCATTTACTAATTACTGG-3'. The 25- μL reaction volume contained 4- μL RNA extract, 5- μL 5 \times QIAGEN OneStep RT-PCR Buffer, 1- μL dNTP mix (for a final concentration of 400 μM of each dNTP), 1- μL each primer (final concentration, 10 μM each), 1- μL QIAGEN OneStep RT-PCR Enzyme Mix (which contained Omniscript and Sensiscript reverse transcriptases and HotStarTaq DNA Polymerase), and RNase- and DNase-free water. The reactions were carried out with an initial 30-min reverse transcription step at 50°C , followed by a 15-min PCR activation step at 95°C , 40 cycles of amplification (30 s at 94°C , 30 s at 55°C , 1 min at 72°C), and a final 10-min extension step at 72°C . All PCR products were run on agarose gels, stained with ethidium bromide, and then visualised under UV light. Most amplicons were purified using the QIAquick Gel Extraction kit (QIAGEN) and sequenced by Cycle Sequencing Kits (Applied Biosystems) on a DNA analyser.

Specimens positive for HCoV-OC43 were tested for coinfection with other respiratory viruses, including HCoV-229E, HCoV-NL63, HCoV-HKU1, Flu A and B (influenza A and B viruses, respectively), PIV 1, 2, 3 (parainfluenza virus types 1, 2 and 3, respectively), RSV (respiratory syncytial virus), hMPV (human metapneumovirus), adenovirus and picornaviruses (including enterovirus and rhinovirus), as described previously [30,35].

Phylogenetic analysis and amino acid alignment

Fifty partial S gene and forty partial N gene sequences from positive samples were sequenced and submitted to the GenBank sequence database under the accession numbers KF512574 to KF512663 (**Table S1 in File S1**).

To determine the diversity of HCoV-OC43, two multiple nucleotide sequence alignments were prepared using Clustal X version 2.1 and BioEdit Sequence Alignment Editor [37]. Using MEGA 5.05 (<http://www.megasoftware.net/>) [38], two neighbour-joining phylogenetic trees (bootstrap values: 1000, Kimura 2-parameter model) were constructed based on alignment of the S and N nucleotide sequences [39], respectively, from the HCoV-OC43 isolates in Beijing and some HCoV-OC43 reference strains in GenBank (**Table S1 in File S1**).

Amino acid alignments of partial S and N protein sequences were performed using the BioEdit Sequence Alignment Editor and GeneDoc software (<http://www.nrbsc.org/gfx/genedoc>).

Statistical analysis

Demographic data and clinical features were compared between groups of patients using the χ^2 - or Fisher's exact test for categorical variables in SAS programme version 9.2. Logistic regression analysis was performed using 95% confidential intervals (CI) to identify any relationship between HCoV-OC43 infection and clinical symptoms. *P* values < 0.05 were considered to indicate a significant difference.

Results

HCoV-OC43 prevalence and coinfection

Patients over 14 years old (range, 14–89 years; average, 33.43 years) were enrolled in this study (**Table 1**), comprising 47.76% (267/559) males and 52.23% (292/559) females. Among the 559 nasopharyngeal swab specimens from adult patients with ARI collected from December 2011 to December 2012, 70 (12.52%; 95% CI: 9.78–15.26%) were positive for HCoV-OC43 by either the S-target or N-target RT-PCR assay, which was significantly higher than the number determined by either assay alone, as 50 (8.94%) were positive by targeting only the S gene and 40 (7.16%) were positive by targeting only the N gene. However, statistical analysis showed no significant difference in the detection rates between the S- and N-targeting assays (8.94% vs. 7.16%; $p > 0.05$). We also found that 20 (3.58%) samples were positive by targeting both S and N genes (**Table 1**).

Coinfection with HCoV-OC43 and another respiratory virus (HCoV-HKU1, influenza [Flu] A and B viruses, parainfluenza virus [PIV] type 2, respiratory syncytial virus [RSV], human metapneumovirus [hMPV], rhinovirus, Coxsackie virus, echovirus, or poliovirus) was found in 25 patients (25/70; 35.71%; 95% CI: 24.49–46.93%), as shown in **Table 1**. The predominant viruses found in coinfections with HCoV-OC43 were rhinovirus and Flu A.

Based on analysis of clinical records, the major clinical signs/symptoms associated with presentation with HCoV-OC43 infection included fever (average temperature, 38.36°C; 97.14%; 95% CI: 93.24–101.04%), sore throat (77.14%; 95% CI: 68.96–88.18%), headache (75.71%; 95% CI: 67.30–86.98%), cough (60%; 95% CI: 48.52–71.48%), nasal stuffiness (47.14%; 95% CI: 35.45–58.83%), nasal discharge (42.86%; 95% CI: 31.27–54.45%), chills (31.43%; 95% CI: 20.55–42.31%), sputum production (18.57%; 95% CI: 9.46–27.68%), and gastrointestinal symptoms (11.43%; 95% CI: 3.98–18.88%), such as diarrhoea and

vomiting (**Table S2 in File S1**). Logistic regression analysis showed that HCoV-OC43 infection was not specifically associated with gastrointestinal symptoms. Only nasal stuffiness differed significantly between the HCoV-OC43 positive and negative groups ($p = 0.005$).

Seasonal and age-group distribution of HCoV-OC43 infection

During the study period (December 2011 to December 2012), autumn (August through October 2012) was the peak season of HCoV-OC43 infection (**Figure S1 in File S2**; $p < 0.0001$). Among the 70 adult patients who were infected with HCoV-OC43, the positivity rate between males (38/267) and females (32/292) showed no significant difference. Furthermore, no significant difference in the HCoV-OC43 positivity rate was found among age groups (young, 14–40 years; middle-aged, 41–65 years; old, >65 years; **Figure S1 in File S2**).

Phylogenetic analysis, amino acid alignment, and recombination event analysis

The partial S and N genes of HCoV-OC43 from the nasopharyngeal swabs of patients with ARI were amplified and sequenced. The partial S gene from 50 amplicons from Beijing patients were compared with those from porcine haemagglutinating encephalomyelitis virus (PHEV), bovine coronavirus (BCoV), and 19 reference strains of HCoV-OC43 found in GenBank (**Table S1 in File S1**). The resulting phylogenetic tree revealed the existence of various virus clusters (left in **Figure 1**). Seven HCoV-OC43 strains from Beijing (LY1, LY82, LY84, XH1316, XH1327, XH1400, and XH1444) along with five reference strains from genotype A belonged to Cluster A. The OC43 (DQ355405) and OC43-BE03 (AY903459) strains of genotype B belonged to Cluster B. Genotypes C (HK04-01) and D (BE04 and HK04-02) were difficult to separate based on partial S gene sequences. Nineteen clinical isolates from Beijing patients formed a novel

Table 1. Frequency of HCoV-OC43 infection, as determined by two molecular assays, and coinfection.

HCoV-OC43 detection status	Positive Number (%)
Different targeting assays (from total 559 samples)	
Either S or N target detected	70 (12.52)
Only S target detected	50 (8.94)
Only N target detected	40 (7.16)
Both S and N targets detected	20 (3.58)
Total Coinfection	
Coinfection with (from 70 samples)	25 (35.71)
Rhinovirus	6 (8.57)
Flu A	4 (5.71)
Coxsackie virus	3 (4.29)
HCoV-HKU1	2 (2.86)
RSV	2 (2.86)
hMPV	2 (2.86)
Echovirus	2 (2.86)
Flu B	1 (1.43)
PIV 2	1 (1.43)
Poliovirus	1 (1.43)
Rhinovirus+Flu A	1 (1.43)

doi:10.1371/journal.pone.0100781.t001

unique cluster and showed high similarity with isolates from Japan (AB695079) and France (DQ355408), which were named as UNT (C/D). Interestingly, two ATCC prototype strains (Z32768 and Z32769) and three English isolates (HC734573, CQ772300, and DD266155) of HCoV-OC43, which showed high similarity with BCoV (AF391542), formed a unique untyped cluster, which we termed untyped cluster I (UNT1). In addition, 20 HCoV-OC43 clinical isolates from Beijing ARI patients were phylogenetically distinct from the other major clusters and formed another unique cluster, which we termed as UNT (B).

Forty partial N-based amplicons from Beijing ARI patients combined with those from eight reference strains of HCoV-OC43 generated four major phylogenetic clusters (right in **Figure 1**). Cluster A comprised only one clinical strain from Beijing patient (LY373) and four reference strains from genotype A. Two clinical strains from Beijing (LY254, LY321) and the reference strain of genotype B (BE03) formed Cluster B. Seventeen clinical strains from Beijing (LY241, 249, 262–267, 270–274, 277, 278, 308, 326), which formed unique subcluster (UNT) and showed significant divergence from the reference strain of genotype C (HK04-

01) and D (BE04, HK04-02). The remaining reference or clinical strains from Beijing formed the major cluster, named as C/D.

Multiple alignments of amino acid sequences based on partial S or N protein sequences are performed and the consensus substitutes for each cluster are summarized in **Table 2 and Table 3**. Some cluster-specific markers of amino acid changes were found among partial S protein sequences of Cluster A and UNT1, with the exception of strain ATCC-VR759 (AY391777), showed a characteristic four-amino-acid deletion at the same location (amino acids 263–266). Compared to ATCC-VR759, each other cluster also showed unique amino acid substitutions at different sites, which may be considered cluster-specific markers (**Table 2**). Similar cluster-specific markers were found among the partial N protein sequences of HCoV-OC43 isolates when amino acids were aligned (**Table 3**), which showed less diversity than the analysis of S1 protein sequences.

Both partial S1 and N sequence data were obtained from the same set of samples from 20 Beijing ARI patients. Our results showed that only few of 20 clinical variants belonged to the same phylogenetic clusters based on either the S or N gene (**Table S3 in File S1**). Lots of clinical strains were discrepant. These

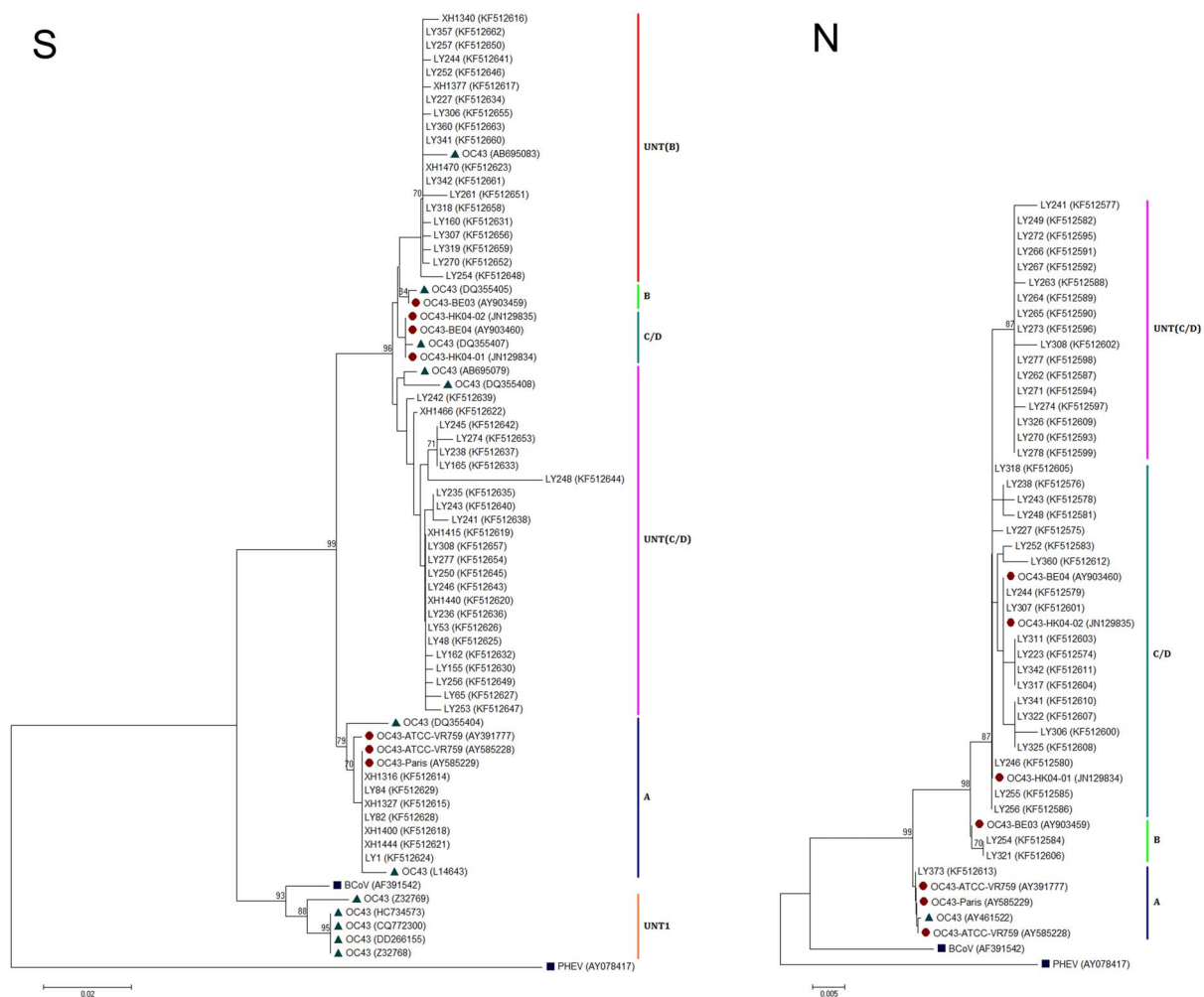


Figure 1. Phylogenetic analysis of partial S (left) and partial N (right) gene sequences of HCoV-OC43 isolates with respect to reference strains was performed with MEGA 5.05 using neighbour-joining. Bootstrap values were calculated from 1,000 trees. The outgroup comprised BCoV (AF391542) and PHEV (AY078417). Bootstrap values over 80% are shown. A, Cluster A; B, Cluster B; C/D, Cluster C/D; UNT, untyped Cluster (1, B and C/D); ●, the sequences of reference strains cut from complete genomes found in GenBank; ▲, the sequences of reference strains cut from the partial S or N gene of HCoV-OC43 in GenBank. doi:10.1371/journal.pone.0100781.g001

Table 2. Summary on consensus amino acid changes for each cluster among S coding region of HCoV-OC43.

Cluster	Location of amino acid in S																											
	212	234	246	247	259	260	261	262	263	264	265	266	277	278	289	297	309	314	335	336	337	340						
ATCC-VR759 (AT391777)	D	V	M	A	N	S	K	V	K	N	G	F	S	R	I	M	Q	P	K	L	N	N						
Other A	:	:	:	:	:	:	:	L	-	-	-	-	:	:	:	:	:	:	:	P	:	:						
UNT1	:	:	T	V	:	A	M	-	-	-	-	-	:	K	V	K	L	S	I	P	:	D						
B	T	F	:	:	I	R	R	D	I	:	:	:	:	:	:	:	:	:	:	P	D	:						
C/D	T	F	:	:	I	R	L	D	I	:	:	:	:	:	:	:	:	:	:	P	D	:						
UNT(B)	T	F	:	:	I	R	L	D	I	:	:	:	P	:	:	:	:	:	:	P	D	:						
UNT(C/D)	T	:	:	:	I	A	R	R	N	I	:	:	:	:	:	:	:	:	:	P	D	:						

doi:10.1371/journal.pone.0100781.t002

discrepancies between S- and N-based phylogenetic trees of the same clinical isolates may indicate the emergence of natural recombination events in the HCoV-OC43 population among Beijing patients with ARI. However, no demographic or clinical association with specific clusters was found (data not shown).

Discussion

Although the human coronavirus OC43 (HCoV-OC43) was identified in 1967 [8], little information on its molecular epidemiology or phylogenetic features among adult ARI patients in Beijing is available. This study is the first report of the molecular epidemiology and genetic diversity of HCoV-OC43 infection in adult patients with ARI from Beijing. We analysed 559 nasopharyngeal swab samples, 70 (12.52%; 95% CI: 9.78–15.26%) of which tested positive for detection of HCoV-OC43, which is consistent with a previous report [34–36,40]. In total, 50 amplicons from the S gene and 40 amplicons from the N gene of HCoV-OC43 isolates were sequenced. Phylogenetic tree analysis based on partial S or N gene sequences in this study showed the presence of at least three clusters (A, B, and C/D) among HCoV-OC43 strains from Beijing adult patients with ARI. Interestingly, we are the first to describe three novel HCoV-OC43 clusters or subclusters based on phylogenetic analysis of partial S genes. The UNTI cluster comprised of 4 isolates from Germany and England showed high similarity with BCoV (AF391542); however, no UNTI strain was found in our patients from Beijing. Some cluster-specific markers of amino acid changes were noted among partial S and N protein sequences from the HCoV-OC43 isolates.

Though viral isolation is the “gold standard” for laboratory diagnosis of virus infection, the great difficulty of HCoV isolation and the time-consuming nature of its culture led numerous researchers to prefer molecular detection methods, such as RT-PCR, to screen for HCoV-OC43 infection [13–16,30–36,40]. The analytic sensitivity of primers specific for different targets has been described previously [33,35]. In the current study, we designed two sets of primers to amplify partial S and N gene sequences of HCoV-OC43 based on the bioinformatics literature. The detection limit for each assay was 10 copies/reaction and rate was not significantly different for the two RT-PCR assays (8.94 vs. 7.16%; $p > 0.05$), but the HCoV-OC43 detection rate was enhanced significantly when the results from the two assays were combined (12.52%; 95% CI: 9.78–15.26%). This indicates that combining RT-PCR assays with different targets is advantageous for detection of HCoV-OC43 infection with increased sensitivity.

Previous reports have shown that infection with HCoV-OC43 was more common amongst elderly persons [18,34,40]. However, the HCoV-OC43 detection rates were similar among the different age groups of adults enrolled in our study. A few reports have suggested that the peak of infection occurs in winter and early spring [39,40,41]. Our study determined that HCoV-OC43 infection among adult patients occurred mainly in autumn of 2010 in Beijing, which is consistent with another study in Beijing that found that HCoV-OC43 infection was most prevalent in summer and early autumn [34]. No association was found between specific clusters of HCoV-OC43 strains and clinical signs or symptoms in this study.

The most commonly used genes used for phylogenetic studies include chymotrypsin-like protease, Pol, helicase, S, and N, because these are present in all coronavirus genomes and are of considerable length [3,22–25]. Two studies from Europe analysing the S1 genes of various HCoV-OC43 strains showed high genetic diversity [26,27]. In the current study, we constructed phylogenetic trees using the S1 sequence, which encodes the N-terminus

Table 3. Summary on consensus amino acid changes for each cluster among N coding region of HCoV-OC43.

Cluster	Location of amino acid in N				
	374	387	401	424	441
ATCC-VR759 (AT391777)	N	M	N	R	Y
Other A
B	S	I	..
C/D	S	..N	..	I	F
UNT(C/D)	S	..	T	I	F

doi:10.1371/journal.pone.0100781.t003

domain (spanning amino acids 188–425) of HCoV-OC43 S and contains a hypervariable region (**Figure S2 in File S2**). Our data showed that there were at least five distinct clusters of HCoV-OC43 strains based on analysis of this partial S gene. We also constructed phylogenetic trees using the partial N sequence (spanning amino acids 288–448 of N), which represents the major diagnostic antigen and more conserved transcripts of the HCoV infection. We found that some of HCoV-OC43 isolates from our patients with ARI showed discrepancies between the S- and N-based phylogenetic trees, which might indicate the emergence of natural recombination events, as suggested in previous reports [25–27], since one unique feature of coronavirus replication is the high frequency of RNA recombination [25,42–44]. RNA recombination is an important mechanism for the natural evolution of coronaviruses. The mean evolutionary rate of HCoV-OC43 was calculated to be 6.7×10^{-4} and 3.6×10^{-4} nucleotide substitutions per site per year for the S and N genes, respectively [24,25]. Hence, it is easy to speculate that the appearance of a novel cluster of HCoV-OC43 strains in the current study was due to natural recombination events within the circulating HCoV-OC43 population in Beijing. To better understand the recombination and evolution of HCoV-OC43, future studies should take into account the complete genome sequences of clinical strains. That being said, the data presented here provide a better understanding of the evolution of HCoV-OC43 in different geographic areas and the role of selective pressures at different regions of the genome.

In conclusion, the prevalence of HCoV-OC43 infection among 559 adult patients with ARI in Beijing was first investigated in this study using two RT-PCR assays targeting the S and N genes. To our knowledge, these data present the genetic analysis of the largest number (70 strains) of HCoV-OC43 clinical isolates from Beijing based on partial S and N coding gene sequences. Through this analysis, we identified several novel clusters of HCoV-OC43.

References

- van der Hoek L (2007) Human coronaviruses: what do they cause? *Antivir Ther* 12(4 Pt B): 651–8.
- Woo PC, Huang Y, Lau SK, Yuen KY (2010) Coronavirus genomics and bioinformatics analysis. *Viruses* 2(8): 1804–20.
- (2012) Virus taxonomy, classification and nomenclature of viruses: ninth report of the International Committee of Taxonomy of Viruses. San Diego, CA: Academic Press.
- Peiris JS, Yuen KY, Osterhaus AD, Stöhr K (2003) The severe acute respiratory syndrome. *N Engl J Med* 349(25): 2431–2441.
- Zaki AM, van Boheemen S, Bestebroer TM, Osterhaus AD, Fouchier RA (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med* 367(19): 1814–20.
- van Boheemen S, de Graaf M, Lauber C, Bestebroer TM, Raj VS, et al (2012) Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. *MBio* pii: e00473-12.
- Tyrrell DA, Bynoe ML (1966) Cultivation of viruses from a high proportion of patients with colds. *Lancet* 1: 76–77.
- McIntosh K, Becker WB, Chanock RM (1967) Growth in suckling mouse brain of “IBV-like” viruses from patients with upper respiratory tract disease. *Proc Natl Acad Sci* 58: 2268–2273.
- van der Hoek L, Pyrc K, Jebbink MF, Vermeulen-Oost W, Berkhout RJ, et al (2004) Identification of a new human coronavirus. *Nat Med* 10: 368–373.
- Woo PCY, Lau SKP, Chu CM, Chan KH, Tsui HW, et al (2005) Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J Virol* 79: 884–895.
- McIntosh K, Chao RK, Krause HE, Wasil R, Moegea HE, et al (1974) Coronavirus infection in acute lower respiratory tract disease of infants. *J Infect Dis* 130: 502–507.
- Gagneur A, Sizun J, Vallet S, Legr MC, Picard B, et al (2002) Coronavirus-related nosocomial viral respiratory infections in a neonatal and paediatric intensive care unit: a prospective study. *J Hosp Infect* 51: 59–64.
- Jean A, Quach C, Yung A, Semret M (2013) Severity and outcome associated with human coronavirus OC43 infections among children. *Pediatr Infect Dis J* 32(4): 325–9.

The most important findings of this study are that we provide substantial evidence for the genetic diversity of HCoV-OC43. At least three distinct clusters (A, B, C/D) of HCoV-OC43 strains are circulating among adults with ARI in Beijing in 2010 and some novel unique clusters (UNT) of HCoV-OC43 was found. Our data also suggest that cluster-specific genetic markers and natural recombinant events in the HCoV-OC43 genome need further investigation.

Supporting Information

File S1 This file contains Table S1, Table S2, and Table S3. Table S1, Reference strains of HCoV-OC43 used in this study. Table S2, Comparison of clinical signs of those infected with HCoV-OC43 versus non-HCoV-OC43. Table S3, Information on 20 HCoV-OC43 detected using both S and N as targets. (DOC)

File S2 This file contains Figure S1 and Figure S2. Figure S1, Seasonal ($p < 0.0001$) and age ($p = 0.68$) distribution of HCoV-OC43 infection from December 2011 to December 2012. Figure S2, SimPlot analysis of complete genome sequence data of HCoV-OC43 strains in reference to HCoV-OC43 ATCC-VR759 (AY391777) and the location of PCR targets (S or N) in this study. The highest variability is found in the S coding region of the genome sequence. (DOCX)

Author Contributions

Conceived and designed the experiments: QH RL KP YL WT. Performed the experiments: QH RL XD WW YZ. Analyzed the data: QH RL KP YW WT. Contributed reagents/materials/analysis tools: KP. Contributed to the writing of the manuscript: QH RL YW WT.

14. 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(8): 985–9.
15. Cabeça TK, Granato C, Bellei N (2013) Epidemiological and clinical features of human coronavirus infections among different subsets of patients. *Influenza Other Respir Viruses* 7(6): 1040–7.
16. Lau SK, Woo PC, Yip CC, Tse H, Tsoi HW, et al (2006) Coronavirus HKU1 and other coronavirus infections in Hong Kong. *J Clin Microbiol* 44(6): 2063–71.
17. Gerna G, Campanini G, Rovida F, Percivalle E, Sarasini A, et al (2006) Genetic variability of human coronavirus OC43-, 229E-, and NL63-like strains and their association with lower respiratory tract infections of hospitalized infants and immunocompromised patients. *J Med Virol* 78(7): 938–49.
18. Birch CJ, Clothier HJ, Seccull A, Tran T, Catton MC, et al (2005) Human coronavirus OC43 causes influenza-like illness in residents and staff of aged-care facilities in Melbourne, Australia. *Epidemiol Infect* 133(2): 273–7.
19. Arbour N, Day R, Newcombe J, Talbot PJ (2000) Neuroinvasion by human respiratory coronaviruses. *J Virol* 74(19): 8913–21.
20. Gerna G, Passarini N, Battaglia M, Rondanelli EG (1985) Human enteric coronaviruses: antigenic relatedness to human coronavirus OC43 and possible etiologic role in viral gastroenteritis. *J Infect Dis* 151: 796–803.
21. Frank E, Zhen O, Huang YT (2010) Human coronaviruses are uncommon in patients with gastrointestinal illness. *J Clin Virol* 48(2): 131–133.
22. Vijgen L, Keyaerts E, Lemey P, Maes P, Van Reeth K, et al (2006) Evolutionary history of the closely related group 2 coronaviruses: porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, and human coronavirus OC43. *J Virol* 80(14): 7270–4.
23. Vijgen L, Keyaerts E, Moës E, Thoelen I, Wollants E, et al (2005) Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. *J Virol* 79(3): 1595–604.
24. Bidokhti MR, Trávén M, Krishna NK, Munir M, Belák S, et al (2013) Evolutionary dynamics of bovine coronaviruses: natural selection pattern of the spike gene implies adaptive evolution of the strains. *J Gen Virol* 94(Pt 9): 2036–49.
25. Lau SK, Lee P, Tsang AK, Yip CC, Tse H, et al (2011) Molecular epidemiology of human coronavirus OC43 reveals evolution of different genotypes over time and recent emergence of a novel genotype due to natural recombination. *J Virol* 85(21): 11325–37.
26. Vabret A, Dina J, Mourez T, Gouarin S, Petitjean J, et al (2006) Inter- and intra-variant genetic heterogeneity of human coronavirus OC43 strains in France. *J Gen Virol* 87(Pt 11): 3349–53.
27. Vijgen L, Keyaerts E, Lemey P, Moës E, Li S, et al (2005) Circulation of genetically distinct contemporary human coronavirus OC43 strains. *Virology* 337(1): 85–92.
28. Blanchard EG, Miao C, Haupt TE, Anderson IJ, Haynes LM (2011) Development of a recombinant truncated nucleocapsid protein based immunoassay for detection of antibodies against human coronavirus OC43. *J Virol Methods* 177(1): 100–6.
29. Liang FY, Lin LC, Ying TH, Yao CW, Tang TK, et al (2013) Immunoreactivity characterisation of the three structural regions of the human coronavirus OC43 nucleocapsid protein by Western blot: implications for the diagnosis of coronavirus infection. *J Virol Methods* 187(2): 413–20.
30. Druce J, Tran T, Kelly H, Kaye M, Chibo D, et al (2005) Laboratory diagnosis and surveillance of human respiratory viruses by PCR in Victoria, Australia, 2002–2003. *J Med Virol* 75(1): 122–9.
31. Gaunt ER, Hardie A, Claas EC, Simmonds P, Templeton KE (2010) Epidemiology and clinical presentations of the four human coronaviruses 229E, HKU1, NL63, and OC43 detected over 3 years using a novel multiplex real-time PCR method. *J Clin Microbiol* 48(8): 2940–7.
32. Vijgen L, Keyaerts E, Moës E, Maes P, Duson G, et al (2005) Development of one-step, real-time, quantitative reverse transcriptase PCR assays for absolute quantitation of human coronaviruses OC43 and 229E. *J Clin Microbiol* 43(11): 5452–6.
33. Vabret A, Mouton F, Mourez T, Gouarin S, Petitjean J, et al (2001) Direct diagnosis of human respiratory coronaviruses 229E and OC43 by the polymerase chain reaction. *Journal of Virological Methods* 97: 59–66.
34. Ren L, Gonzalez R, Xu J, Xiao Y, Li Y, et al (2011) Prevalence of human coronaviruses in adults with acute respiratory tract infections in Beijing, China. *J Med Virol* 83(2): 291–7.
35. Lu R, Yu X, Wang W, Duan X, Zhang L, et al (2012) Characterization of human coronavirus etiology in Chinese adults with acute upper respiratory tract infection by real-time rt-PCR assays. *PLoS One* 7(6): e38638.
36. Yu X, Lu R, Wang Z, Zhu N, Wang W, et al (2012) Etiology and Clinical Characterization of Respiratory Virus Infections in Adult Patients Attending an Emergency Department in Beijing. *Plos One* 7(2): e32174.
37. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22: 4673–4680.
38. Tamura K, Peterson D, Peterson N, Stecher G, Nei M (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28: 2731–2739.
39. Saitou N, Nei M (1987) The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4: 406–425.
40. Greenberg SB (2011) Update on rhinovirus and coronavirus infections. *Semin Respir Crit Care Med* 32(4): 433–46.
41. Larson HE, Reed SE, Tyrell DAJ (1980) Isolation of rhinoviruses and coronaviruses from 38 colds in adults. *J. Med. Virol* 5: 221–229.
42. Lai MMC (1992) RNA recombination in animal and plant viruses. *Microbiol Rev* 56: 61–79.
43. Vijaykrishna D, Smith GJ, Zhang JX, Peiris JS, Chen H, et al (2007) Evolutionary insights into the ecology of coronaviruses. *J. Virol* 81: 4012–4020.
44. St-Jean JR, Jacomy H, Desforges M, Vabret A, Freymuth F, et al (2004) Human respiratory coronavirus OC43: genetic stability and neuroinvasion. *J Virol* 78(16): 8824–34.