Genetic relatedness of the novel human group C betacoronavirus to *Tylonycteris* bat coronavirus HKU4 and *Pipistrellus* bat coronavirus HKU5

Patrick CY Woo1,2,3,4,*, Susanna KP Lau1,2,3,4,*, Kenneth SM Li1, Alan KL Tsang1 and Kwok-Yung Yuen1,2,3,4

Emerging Microbes and Infections (2012) 1, e35; doi:10.1038/emi.2012.45; published online 7 November 2012

The recent outbreak of severe respiratory infections associated with a novel group C betacoronavirus (HCoV-EMC) from Saudi Arabia has drawn global attention to another highly probable “SARS-like” animal-to-human interspecies jumping event in coronavirus (CoV). The genome of HCoV-EMC is most closely related to *Tylonycteris* bat coronavirus HKU4 (Ty-BatCoV HKU4) and *Pipistrellus* bat coronavirus HKU5 (Pi-BatCoV HKU5) we discovered in 2006. Phylogenetically, HCoV-EMC is clustered with Ty-BatCoV HKU4/Pi-BatCoV HKU5 with high bootstrap supports, indicating that HCoV-EMC is a group C betacoronavirus. The major difference between HCoV-EMC and Ty-BatCoV HKU4/Pi-BatCoV HKU5 is in the region between S and E, where HCoV-EMC possesses five ORFs (NS3a-NS3e) instead of four, with low (31%–62%) amino acid identities to other Subgenus Betacoronavirus. The source of the virus remains obscure. In this article, this novel human group C betacoronavirus originating from Saudi Arabia has drawn global attention to another highly probable “SARS-like” interspecies jumping event of coronavirus (CoV) from animal to human. In June 2012, a novel CoV was isolated using Vero cells from the lung tissue of a 60-year-old resident of Saudi Arabia with fatal acute pneumonia and renal failure. In September 2012, another 49-year-old male resident of Qatar with severe acute pneumonia and renal failure and recent travel history to Saudi Arabia was admitted to an intensive care unit in Qatar. RT-PCR and sequencing of a short fragment of RNA-dependent RNA polymerase (RdRp) confirmed the presence of the same CoV as detected in the first Saudi Arabian case. Complete genome sequencing of the virus isolated from the first patient was performed by Fouchier et al. at the Erasmus University Medical Centre, the Netherlands, and the sequence was released on September 28, 2012 (GenBank accession NO JX869059 and named as human betacoronavirus 2c EMC(2012)). So far, there is no evidence of human-to-human transmission. The source of the virus remains obscure. In this article, this novel human group C betacoronavirus is abbreviated as HCoV-EMC.

After the SARS epidemic, we started to focus on CoV biodiversity, genomics and phylogeny and built up an evolutionary map of CoV evolution. Before 2003, there were less than 10 CoVs with complete genome sequences available, which include two human CoVs, human coronavirus 229E (HCoV-229E) and human coronavirus OC43 (HCoV-OC43). By September 2012, the number of CoVs with complete genomes sequenced had tripled. It includes two additional human CoVs, human coronavirus NL63 (HCoV-NL63) and human coronavirus HKU1 (HCoV-HKU1). Traditionally, CoVs were classified into groups 1, 2 and 3. In 2011, the Coronavirus Study Group of the International Committee for Taxonomy of Viruses has re-classified these three groups of CoVs as three genera, *Alphacoronavirus*, *Betacoronavirus* and *Gammacoronavirus* and we have discovered a fourth genus of CoV, *Deltacoronavirus*, which includes at least nine avian CoVs and a porcine coronavirus HKU15. Within the *betacoronavirus* group, they are further subclassified into group A, including HCoV-HKU1, HCoV-OC43, bovine coronavirus (BCoV), sable antelope coronavirus, giraffe coronavirus, equine coronavirus, porcine hemagglutinating encephalomyelitis virus, murine hepatitis virus, rat coronavirus and rabbit coronavirus HKU14 (RbCoV HKU14); group B, including the human and civet SARS-related CoVs (SARSr-CoV) and SARS-related Rhinolophus bat coronavirus (SARS-Rh-BatCoV); group C, including *Tylonycteris* bat coronavirus HKU4 (Ty-BatCoV HKU4) and *Pipistrellus* bat coronavirus HKU5 (Pi-BatCoV HKU5) we discovered in 2006 and group D, including *Roussettus* bat coronavirus HKU9 (Ro-BatCoV HKU9). In addition to Ty-BatCoV HKU4 and Pi-BatCoV HKU5, other group C bat betacoronaviruses should also be present, but their complete genome sequences are not available. Based on the CoVs discovered, we have constructed a model of CoV evolution, with evidence supporting that bat CoVs are the gene source of alphaCoVs and betacovs and avian CoVs are the gene source of gammaCoVs and deltaCoVs. All these works have laid down an evolutionary map for rapid phylogenetic and bioinformatics analyses of HCoV-EMC. The diversity of CoVs is a result.
of the infidelity of RdRp which make CoV genomes especially plastic, a high frequency of homologous RNA recombination due to their unique random template switching during RNA replication, and their large genomes. In addition to biodiversity, a number of natural recombination and possible interspecies jumping events has also been documented in betaCoVs.\textsuperscript{6,11,14–18} For group A betaCoVs, molecular clock analysis has shown that HCoV-OC43 is a relatively recent zoonotic virus of bovine origin that emerged in around 1890 likely from bovine-to-human transmission.\textsuperscript{17} We have also recently discovered RhCoV HKU14, closely related to other members of the species Betacoronavirus 1 including HCoV-OC43 and BCoV, with recombination events that may have played a role in interspecies transmission of these HCoV-OC43-related viruses between human, cattle, rabbits, swine and horses.\textsuperscript{8} Despite having circulated in humans for more than a century, HCoV-OC43 is also found to be continuously evolving, with the recent emergence of a novel genotype due to natural recombination.\textsuperscript{15} For group B betaCoVs, SARSr-CoV is believed to be transmitted from civet to humans, although it is the horseshoe bat that was likely the primary host.\textsuperscript{7,8} Civet SARSr-CoV was also likely a recombinant virus arising from different strains of SARSr-Rh-BatCoV from different geographical locations in China.\textsuperscript{14,16} Although no interspecies transmission events have been documented in group D betaCoVs, we have also identified recombination events between different Rh-BatCoV HKU9 strains from different bat individual, which may have allowed for the generation of different genotypes.\textsuperscript{11} While these findings supported that betaCoVs have the propensity to recombine and cause interspecies transmission, such events were unknown in group C betaCoVs. As HCoV-EMC is most closely related to Ty-BatCoV HKU4 and Pi-BatCoV HKU5, it would be important to study their genetic relatedness, which may provide clues on whether bats are the possible origin as in SARSr-CoV.

The genome characteristics and organization of HCoV-EMC are similar to those of Ty-BatCoV HKU4 and Pi-BatCoV HKU5. Ty-BatCoV HKU4 was discovered from lesser bamboo bats (Tylonycteris pachypus) and Pi-BatCoV HKU5 was discovered from Japanese pipistrelles (Pipistrellus abramus) in Hong Kong.\textsuperscript{9} Both lesser bamboo bats and Japanese pipistrelles are insectivorous microbats found in China and some other parts of Asia. The size of the genome of HCoV-EMC is 30,106 bases, slightly smaller than those of Ty-Bat-CoV HKU4 (30,286 to 30,316 bases) and Pi-Bat-CoV HKU5 (30,482 to 30,488 bases); and the G+C content is 41%, in between those of Ty-Bat-CoV HKU4 (38%) and Pi-Bat-CoV HKU5 (43%). The replicase ORF1ab occupies 21.5 kb of the genome. This ORF encodes 16 putative non-structural proteins. These proteins are produced by the replicase polyprotein by PLpro and 3CLpro at specific sites which are conserved with those in Ty-Bat-CoV HKU4 and/or Pi-Bat-CoV HKU5 (Table 1).

HCoV-EMC has the same basic genome structure as Ty-Bat-CoV HKU4 and Pi-Bat-CoV HKU5 (Figure 1). It also possesses the same putative transcription regulatory sequence (TRS) motif, 5'–ACGAAC–3', as Ty-Bat-CoV HKU4 and Pi-Bat-CoV HKU5 at the 3' end of the leader sequence and precedes each ORF except NS3c, NS3e and N. This TRS has also been shown to be the TRS for other group B, C and D betaCoVs. The TRS for N is 5'–ACGAUA–3'. Similar to other group B, C and D betaCoVs, the genome of HCoV-EMC has a putative PLpro, which is homologous to PL2\textsuperscript{pro} of alphaCoVs and group A betaCoVs and PL\textsuperscript{pro} of gammaCoVs and deltaCoVs. Similar to Ty-Bat-CoV HKU4 and Pi-Bat-CoV HKU5, no proteolytic cleavage site is present in S of HCoV-EMC. All cysteine residues in S of HCoV-EMC, Ty-Bat-CoV HKU4 and Pi-Bat-CoV HKU5 are conserved. In contrast to the genomes of Ty-Bat-CoV HKU4 and Pi-Bat-CoV HKU5 which contain four ORFs that encode putative non-structural proteins (NS3a, NS3b, NS3c and NS3d) between S and E, this region of HCoV-EMC contains five ORFs that encode putative non-structural proteins (NS3a, NS3b, NS3c, NS3d and NS3e) (Figure 1). This is the region of HCoV-EMC that possesses the lowest amino acid identities to those in Ty-Bat-CoV HKU4 and Pi-Bat-CoV HKU5. NS3a, NS3b and NS3c of HCoV-EMC possess 42%–43%, 41%–47% and 31% amino acid identities to NS3a, NS3b and NS3c of Ty-Bat-CoV HKU4 and Pi-Bat-CoV HKU5, respectively. NS3d of HCoV-EMC is homologous to amino acids 1 to 110/210/310 of NS3d of Ty-Bat-CoV HKU4 and/or Pi-Bat-CoV HKU5.
103 of NS3d in Ty-BatCoV HKU4 and Pi-BatCoV HKU5 (35%–49% amino acid identities), with a stop codon UAG present at nucleotide position 27,160, leading to premature termination. NS3e of HCoV-EMC is homologous to amino acids 116/122 to 223/227 of NS3d in Ty-BatCoV HKU4 and Pi-BatCoV HKU5 (60%–62% amino acid identities). NS3c and NS3e do not possess any TRS or internal ribosomal entry site. BLAST search revealed no amino acid similarities between these putative non-structural proteins and other known proteins and no functional domains were identified by PFAM and InterProScan. TMHMM and TMpred analyses show one and two putative transmembrane domains in NS3a (residues 9 to 29) and NS3d (residues 36 to 56 and 71 to 91), respectively. Similar to Ty-BatCoV HKU4 and Pi-BatCoV HKU5, the 3' untranslated region of the genome of HCoV-EMC contains predicted bulged stem-loop structures 16 to 76 nucleotides downstream of the N genes. Downstream to the bulged stem-loop structure, 97 to 121 nucleotides downstream of the N genes, a pseudoknot structure is present. Bootscan analysis did not show any recombination between HCoV-EMC, Ty-BatCoV HKU4 and Pi-BatCoV HKU5.

The phylogenetic trees constructed using the amino acid sequences of the 3CLpro, RdRp, Hel, S and N of HCoV-EMC and other CoVs are shown in Figure 2. For all the five genes, HCoV-EMC is clustered with Ty-BatCoV HKU4 and Pi-BatCoV HKU5, with high bootstrap supports in all cases, indicating that HCoV-EMC is a group C betaCoV. Although it seems that HCoV-EMC is clustered with Pi-BatCoV HKU5 in the phylogenetic trees constructed using RdRp and Hel, the bootstrap supports were only 652 and 588, respectively, suggesting that there is no obvious difference between the relatedness of HCoV-EMC to Ty-BatCoV HKU4 and Pi-BatCoV HKU5, with high bootstrap supports in all cases, indicating that HCoV-EMC is a group C betaCoV (Figure 2). Although it seems that HCoV-EMC is clustered with Pi-BatCoV HKU5 in the phylogenetic trees constructed using RdRp and Hel, the bootstrap supports were only 652 and 588, respectively, suggesting that there is no obvious difference between the relatedness of HCoV-EMC to Ty-BatCoV HKU4 and Pi-BatCoV HKU5.

Comparison of the amino acid identities of the seven conserved replicase domains for species demarcation (ADRP, nsp5 (3CLpro), nsp12 (RdRp), nsp13 (Hel), nsp14 (ExoN), nsp15 (NendoU) and nsp16 (2'-O-MT)) between HCoV-EMC, Ty-BatCoV HKU4 and Pi-BatCoV HKU5 showed that there is less than 90% identity in four of the seven domains (ADRP 68%–69% identity, nsp5 81%–83% identity, nsp15 76%–80% identity and nsp16 84%–85% identity), indicating that HCoV-EMC is a novel CoV species. For nsp12, nsp13 and nsp14, there are 90%–92%, 92%–94% and 86%–92% amino acid identities between HCoV-EMC and Ty-BatCoV HKU4/Pi-BatCoV HKU5.

Using the sequences available at the moment and Yule process speciation under a relaxed clock model with an uncorrelated lognormal distribution, the mean evolutionary rate of betaCoVs was estimated at $2.37 \times 10^{-4}$ nucleotide substitutions per site per year for the RdRp gene. Molecular clock analysis using the RdRp gene showed that HCoV-EMC diverged from the most recent common ancestor of group C betaCoVs at around 941 (HPDs, 529 BC to 1878). Compared to the human and civet SARS-CoV and SARSr-Rh-BatCoV cluster, the human/civet SARS-CoV diverged from the
Novel human group C betacoronavirus

PCY Woo et al
Figure 2. Phylogenetic analysis of HCoV-EMC. The trees were constructed by the neighbor-joining method using Kimura correction and bootstrap values calculated from 1000 trees. 318, 951, 600, 1491 and 510 amino acid positions in chymotrypsin-like protease (3CLpro), RNA-dependent RNA polymerase (RdRp), helicase (Hel), spike (S) and nucleocapsid (N) respectively were included in the analysis. For 3CLpro, S and N, the scale bars indicate the estimated number of substitutions per 20 amino acids. For RdRp and Hel, the scale bars indicate the estimated number of substitutions per 50 amino acids. PEDV, porcine epidemic diarrhea virus (NC_016994); WiCoV HKU20, wigeon coronavirus HKU20 (NC_016995); CMCoV HKU21, common moorhen coronavirus HKU21 (NC_016996).}

ACKNOWLEDGEMENTS

This work is partly supported by a Research Grants Council Grant HKU 78070/M and the HSARK Research Fund for the Control of Infectious Diseases of the Health, Welfare and Food Bureau.

most recent common ancestor of the human/ civet SARSr-CoV and SARSr-Rh-BatCoV at ~year 1653 (HPDs, 1150 to 1668). By definition, the human and civet SARSr-CoV and SARSr-Rh-BatCoV are the same CoV species. These observations suggest that there should be one or more intermediate hosts between Ty-BatCoV HKU4, Pi-BatCoV HKU5 and HCoV-EMC. Sequencing more strains of Ty-BatCoV HKU4, Pi-BatCoV HKU5 and HCoV-EMC, as well as other group C betacovs collected at different time points, should be performed to achieve a more accurate estimation of the divergence time.

In the last decade, we have already witnessed the discovery of two novel human CoVs and an animal-to-human CoV interspecies jumping event on SARSr-CoVs. In contrast to HCoV-229E, HCoV-OC43, HCoV-NL63 and HCoV-HKU1, which are notoriously difficult to culture, HCoV-EMC and human SARS-CoV are both readily cultivable using primate cell lines. This may suggest a possible correlation between cultivability and virulence/recent interspecies jumping. Sequencing more genomes and performing evolutionary analysis will help us understand whether HCoV-EMC represent another recent interspecies jumping event from animal to human or another human CoV that has stably infected human. Our most recent findings showed that CoVs can be transmitted between two bat species of different suborders, suggesting that different degrees of interspecies jumping can occur in nature. More intensive surveillance studies for group C betaCoVs in bats and other animals may reveal the natural host of this novel group human CoV betaCoV. As coronaviruses are prone to recombination and mutation and it has been documented that different levels of interspecies jumping can indeed occur in nature, we should not underestimate the potential of coronaviruses being the cause of another major “SARS-like” pandemic.

1 Zaki AM, van Boheemen S, Bestebroer TM et al. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med 2012; in press.
2 van der Hoek L, Pyrc K, Jebbink MF et al. Identification of a new human coronavirus. Nat Med 2004; 10: 368–373.
3 Woo PC, Lau SK, Chu CM et al. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J Virol 2005; 79: 11668–11676.
4 Woo PC, Lau SK, Lam CS et al. Comparative analysis of complete genome sequences of three avian coronaviruses reveals a novel group 3c coronavirus. J Virol 2009; 83: 908–917.
5 Lau SK, Leung HC, Peiris J et al. Complete genomic sequence of the novel coronavirus HKU9. J Virol 2008; 82: 1819–1826.
6 Lau SK, Lee P, Tsang AK et al. 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 2011; 85: 11325–11337.
7 Lau SK, Li KS, Huang Y et al. Complete genome comparison of SARS-related Rhinophous bat coronavirus 512 reveals bats as reservoir for acute, self-limiting infection that allows recombination events. J Virol 2010; 84: 2808–2819.
8 Vignon JL, Keyaerts E, Moës E et al. Complete genomic sequence of the human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. J Virol 2005; 79: 1595–1604.
9 Woo PC, Lau SK, Yip CC et al. Comparative analysis of 22 coronavirus HKU1 genomes reveals a novel genotype and evidence of natural recombination in human coronavirus HKU1. J Virol 2006; 80: 7136–7145.
10 Lau SK, Li KS, Tsang AK et al. Transmission of a novel alphacoronavirus, bat coronavirus HKU10, from Leschenault’s rousette to Pomona leaf-nosed bats: first evidence of interspecies transmission of coronaviruses between bats of different suborders. J Virol 2012; 86: 11906–11918.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivative Works 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-nd/3.0/