The New Age of Virus Discovery: Genomic Analysis of a Novel Human Betacoronavirus Isolated from a Fatal Case of Pneumonia

Kathryn V. Holmes[a] and Samuel R. Dominguez[a,b,c]

Departments of Microbiology[a] and Pediatrics,[b] University of Colorado School of Medicine, Aurora, Colorado, USA, and Childrens' Hospital Colorado, Aurora, Colorado, USA[c]

ABSTRACT A new human betacoronavirus in lineage c, tentatively called HCoV-EMC, was isolated from a patient from the Kingdom of Saudi Arabia who died from acute severe pneumonia and renal failure. The viral RNA has been detected in eight additional cases. Sequencing and bioinformatic analysis of the viral genomic RNA showed that it is a novel virus not previously detected in any other species and that its closest relatives are two Asian bat coronaviruses. HCoV-EMC may represent a sporadic spillover to humans from an unknown animal reservoir. In a recent article, van Boheemen et al. demonstrated how state-of-the-art sequencing technology and bioinformatic analysis can quickly provide critical insight into the viral genome sequence, phylogeny, replication strategy, and potential drug and vaccine targets and generate tools to evaluate the possible epidemic risk associated with this novel human virus.

During the past decade, at least 9 novel human respiratory viral pathogens have been discovered, primarily by using highly sensitive nucleotide sequencing and new virus detection technologies. These human viruses include human metapneumovirus, rhinoviruses in clade C, bocavirus, polyomaviruses WU and KI 2009 pandemic H1N1 influenza virus, and 3 new human coronaviruses (HCoV). Until 2003, only two coronaviruses, HCoV-OC43 and HCoV-229E, were known to cause human disease, primarily upper respiratory tract infections. The discovery of severe acute respiratory syndrome-CoV (SARS-CoV) as the cause of the SARS pandemic of 2002 to 2003 demonstrated the epidemic potential of this large family of RNA viruses and emphasized their importance in human respiratory diseases. After the SARS pandemic, two additional human coronaviruses, NL63 and HKU1, were identified and found to cause both upper and lower respiratory tract disease. Although these coronaviruses were only recently discovered, they have probably been circulating in the human population worldwide for a long time. HCoV-OC43 apparently jumped from a bovine host into humans more than 100 years ago and has become endemic worldwide. In contrast, the SARS pandemic was caused by a novel human virus that had very recently emerged into the human population from its zoonotic reservoirs, Chinese horseshoe bats (suborder *Microchiroptera*, family *Rhinolophidae*, genus *Rhinolophus*) (1, 2). Indeed, extensive phylogenetic studies suggest that all alpha- and beta coronaviruses may be derived from bat coronaviruses (3).

In June 2012, a novel coronavirus, tentatively named HCoV-EMC (for Erasmus Medical Center where the initial genome sequence was done), was isolated in a monkey kidney cell line from a sputum specimen from a 60-year-old man from the Kingdom of Saudi Arabia who died of acute severe pneumonia and renal failure. Results of standard diagnostic tests of his clinical samples for known respiratory viruses were all negative, which prompted testing of infected cell culture supernatants using a pan-coronavirus reverse transcription-PCR (RT-PCR) assay to amplify a short, highly conserved region of the RNA-dependent RNA polymerase gene. Sequencing the amplicon from the patient's specimen revealed the novel human coronavirus (4). The discovery of a new coronavirus in a patient who died of severe respiratory disease was promptly reported to the World Health Organization and posted on ProMed to alert physicians and diagnostic laboratories to look for additional cases. Eight additional patients from the Middle East with HCoV-EMC infection and severe respiratory disease have been identified, bringing the total of laboratory-confirmed cases up to 9, with 5 deaths (5). In a recent article, van Boheemen et al. described the characterization of the HCoV-EMC genome using unbiased high-throughput next-generation sequencing and state-of-the-art bioinformatics to mine a wealth of information about this new human virus (6).

What can analysis of its RNA genome tell us about this new human coronavirus? Analysis of the HCoV-EMC genome was able to distinguish whether HCoV-EMC was a known human coronavirus with mutations that increased its virulence or a product of recombination between two or more known coronavirus species or, like SARS-CoV, a spillover of a zoonotic coronavirus that was transmitted, either directly or indirectly, to humans from an animal reservoir. The genomic analysis showed that HCoV-EMC is a new coronavirus that had never before been detected in any other species.

What is the relationship between HCoV-EMC and known coronaviruses? According to the current classification by the International Committee on Taxonomy of Viruses, coronaviruses are in the order *Nidovirales*, family *Coronaviridae*, and subfamily *Coronavirinae*. There are at least 4 genera, alpha-, beta-, gamma-, and deltacoronaviruses, and the many virus species within each genus are grouped in multiple lineages called a, b, c, etc. The genome sequence revealed that HCoV-EMC is a betacoronavirus in lineage c, significantly different from SARS-CoV, a betacoronavirus in lineage b, and human betacoronaviruses OC43 and HKU1, which are placed in lineage a. The closest known relatives of HCoV-EMC in lineage c are two betacoronaviruses of Asian bats,

Address correspondence to Kathryn V. Holmes, Kathryn.Holmes@ucdenver.edu.

but because HCoV-EMC and the bat viruses have only 75% to 77% amino acid identity in the 7 key conserved replicase domains used for phylogenetic analysis, HCoV-EMC is classified as a novel species within betacoronavirus lineage c (6). The genomic analysis suggests that HCoV-EMC may be a zoonotic virus that spilled over to infect the 9 laboratory confirmed patients, either directly or indirectly, from an unknown reservoir, possibly a bat. This discovery justifies the recent worldwide surveillance of wildlife for coronavirus infection and the extensive genome sequencing and phylogenetic analysis that have been done since the SARS pandemic to identify potential risk to humans from zoonotic coronavirus infections.

What features of the genome and its translational strategy are similar to or different from those of other coronaviruses? Like all coronaviruses, HCoV-EMC has a single-stranded, plus-sense, polyadenylated RNA genome about 30 kb in length. As expected, the 20-kb sequence at the 5′ end of the genome is translated to yield a huge polyprotein that is cotranslationally cleaved in *cis* by two viral proteases into 16 functional nonstructural proteins that cooperatively form the complex machinery for viral RNA synthesis and RNA recombination. The 10-kb sequence in the 3′ region of coronavirus genomes uses a different translational strategy. This region encodes 4 structural proteins with features common to all CoVs as well as several so-called accessory proteins that are different for each coronavirus and whose origins and functions are unknown. A nested series of 3′ coterminal polyadenylated subgenomic mRNAs is generated in the cytoplasm, and only the gene at the 5′ end of each of these mRNAs is translated. The genome reveals conservation in HCoV-EMC of several potential targets for drugs and vaccines being developed for other CoVs, including the viral spike glycoprotein (S), virus-encoded proteases, and essential enzymatic functions such as the RNA-dependent RNA polymerase and helicase.

How does the genomic analysis of HCoV-EMC expedite further research on this new human virus? As soon as partial nucleotide sequences of HCoV-EMC were available, sensitive RT-PCR tests were developed to specifically detect RNA of this virus in tissues and body fluids of humans and animals. These tests are being used to screen for HCoV-EMC RNA in patients with severe respiratory disease of unknown etiology and in wildlife surveillance. Based on the predicted amino acid sequences of the HCoV-EMC proteins, plasmids can be engineered to express the proteins for structural, antigenic, and functional studies. Recombinant viral proteins are being used in enzyme-linked immunosorbent assays (ELISAs) to detect HCoV-EMC-specific antibodies in sera. HCoV-EMC spike protein in retrovirus pseudotypes can be used to identify the virus receptor and study virus entry. Antisera raised against the recombinant HCoV-EMC proteins can detect viral antigens in infected cell cultures and infected tissues of humans or animals for studies on virus tissue tropism and pathogenesis. The viral genome can now be reconstructed by synthetic biology to create a manipulable cDNA copy that can be mutated for analysis of virus replication, pathogenesis, virulence factors, host range, and vaccines. When additional virus isolates from other patients or animal reservoirs become available, genomic analysis can be used to analyze variants to detect amino acid substitutions in the spike or other proteins that are associated with adaptation to cell culture or changes in antigenicity, host range, and virulence.

What questions must be answered to show whether HCoV-

EMC is likely to cause an epidemic or adapt to become endemic in humans? First, it is necessary to prove whether the HCoV-EMC cultured from the first patient actually caused his fatal disease (7). Several bacterial pathogens were also cultured from his respiratory tract, although this is not uncommon in severely ill patients on respirators. Development of an animal model for HCoV-induced respiratory disease would fulfill Koch's postulates and be useful for testing candidate coronavirus drugs and vaccine strategies. Sera from humans in the Middle East and elsewhere are being tested by ELISA for HCoV-EMC-specific antibodies to determine the prevalence of infection of humans in different regions and estimate the percentage of clinically apparent HCoV-EMC infections and the case/fatality ratio. Surveillance of bats and other animals in the Middle East and elsewhere for the presence of RNA from betacoronavirus lineage c viruses or antibodies to these viruses will help to identify reservoir hosts and possible intermediate hosts that could transmit the virus to humans. So far, no epidemiological connections have been found between animals and the HCoV-EMC patients. This virus is apparently much less easily transmitted from human to human than SARS-CoV, although three of the cases in Saudi Arabia were within one family and several health care workers who cared for two of the confirmed cases in Jordan also developed pneumonia and are now considered probable cases (5). It is possible that as the virus replicates in sporadic cases, mutations, particularly in the genes encoding the spike protein and any immunomodulatory proteins, might be selected that increase human-to-human transmissibility.

Sporadic HCoV-EMC infections in epidemiologically unlinked individuals in the Middle East suggest that it is a zoonotic virus. Where encounters between humans and wildlife have become common, sporadic zoonotic infections may occur without human-to-human transmission. This so-called "virus chatter" was obvious in retrospective studies of SARS-CoV infections in China which showed that isolated cases occurred well before the virus began to spread from human to human during the SARS pandemic (8–10). Mutations in the SARS-CoV spike protein and other viral proteins were associated with the transition from "chatter" to epidemic. It is possible that HCoV-EMC is not yet a human-adapted virus and has simply been detected in rare cases of "chatter" by increasingly sophisticated surveillance. The work by van Boheemen et al. demonstrates the ability of state-of-the-art technology and bioinformatics to obtain full-genome sequence data within days and provide critical insight into the potential risks associated with novel viral pathogens (6). Over the past decade, other advances, including the development of enhanced virus detection technologies, surveillance programs, increased global awareness and communication through Internet-based technologies such as ProMed, real-time sharing of genome sequences in GenBank, and Internet-based publications and the posting of publications in advance of printing, as well as increasing collaborations between researchers and medical and veterinary practitioners such as the One Health Initiative, have greatly facilitated rapid and effective responses to threats of new emerging infectious diseases.

REFERENCES

1. **Li W, et al.** 2005. Bats are natural reservoirs of SARS-like coronaviruses. Science **310**:676–679.
2. **Lau SK, et al.** 2005. Severe acute respiratory syndrome coronavirus-like

virus in Chinese horseshoe bats. Proc. Natl. Acad. Sci. U. S. A. **102**: 14040–14045.

3. **Woo PC, et al.** 2012. Discovery of seven novel mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. J. Virol. **86**:3995–4008.

4. **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**:1814–1820.

5. **World Health Organization Global Alert and Response.** Background and summary of novel coronavirus infection—as of 21 December, 2012. http://www.who.int/csr/disease/coronavirus_infections/update _20121221/en/index.html.

6. **van Boheemen S, et al.** 2012. Genomic characterization of a newly discovered coronavirus associated with acute respiratory distress syndrome in humans. mBio **3**:e00473-12.

7. **Lipkin WI.** 2010. Microbe hunting. Microbiol. Mol. Biol. Rev. **74**: 363–377.

8. **Song HD, et al.** 2005. Cross-host evolution of severe acute respiratory syndrome coronavirus in palm civet and human. Proc. Natl. Acad. Sci. U. S. A. **102**:2430–2435.

9. **Parrish CR, et al.** 2008. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol. Mol. Biol. Rev. **72**: 457–470.

10. **Wolfe ND, Daszak P, Kilpatrick AM, Burke DS.** 2005. Bushmeat hunting, deforestation, and prediction of zoonoses emergence. Emerg. Infect. Dis. **11**:1822–1827.