

## Review

# Epidemiology, Genetic Recombination, and Pathogenesis of Coronaviruses

Shuo Su,<sup>1,\*</sup> Gary Wong,<sup>2,3,4</sup> Weifeng Shi,<sup>5</sup> Jun Liu,<sup>4,6</sup>  
Alexander C.K. Lai,<sup>7</sup> Jiyong Zhou,<sup>1</sup> Wenjun Liu,<sup>3,4</sup>  
Yuhai Bi,<sup>2,3,4,\*</sup> and George F. Gao<sup>2,3,4,6,8,9,\*</sup>

**Human coronaviruses (HCoVs) were first described in the 1960s for patients with the common cold. Since then, more HCoVs have been discovered, including those that cause severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), two pathogens that, upon infection, can cause fatal respiratory disease in humans. It was recently discovered that dromedary camels in Saudi Arabia harbor three different HCoV species, including a dominant MERS HCoV lineage that was responsible for the outbreaks in the Middle East and South Korea during 2015. In this review we aim to compare and contrast the different HCoVs with regard to epidemiology and pathogenesis, in addition to the virus evolution and recombination events which have, on occasion, resulted in outbreaks amongst humans.**

### A Brief Introduction to Human and Animal Coronaviruses

Coronaviruses (CoVs), of the family Coronaviridae, are enveloped viruses with a single-strand, positive-sense RNA genome approximately 26–32 kilobases in size, which is the largest known genome for an RNA virus [1]. The term 'coronavirus' refers to the appearance of CoV virions when observed under electron microscopy, in which spike projections from the virus membrane give the semblance of a crown, or *corona* in Latin [2,3]. All coronaviruses share similarities in the organization and expression of their genome, in which 16 nonstructural proteins (nsp1 through nsp16), encoded by open reading frame (ORF) 1a/b at the 5' end, are followed by the structural proteins spike (S), envelope (E), membrane (M), and nucleocapsid (N), which are encoded by other ORFs at the 3' end. CoVs are separated into four genera based on phylogeny: alpha-CoV (group 1), beta-CoV (group 2), gamma-CoV (group 3) and delta-CoV (group 4) ([http://ictvonline.org/proposals/2008.085-122\\_V.v4.Coronaviridae.pdf](http://ictvonline.org/proposals/2008.085-122_V.v4.Coronaviridae.pdf)). Within the beta-CoV genus, four lineages (A, B, C, and D) are recognized. Distinct from other beta-CoV lineages, lineage A viruses also encode a smaller protein called hemagglutinin esterase (HE), which is functionally similar to the S protein [4].

Since the early 1970s, a variety of pathological conditions in domestic animals have been attributed to CoV infections [5]. With the exception of infectious bronchitis virus (IBV), which causes avian infectious bronchitis in chickens [6], canine respiratory coronavirus (CRCoV), which causes respiratory disease in dogs [7], and mouse hepatitis virus (MHV), which can cause a progressive demyelinating encephalitis in mice [3,8], other CoV infections typically result in

### Trends

Six coronaviruses (CoVs) are known to infect humans: 229E, OC43, SARS-CoV, NL63, HKU1, and MERS-CoV.

Many CoVs are simultaneously maintained in nature, allowing for genetic recombination, resulting in novel viruses.

Recombination of CoV in camels has resulted in a dominant MERS lineage that caused human outbreaks in 2015.

<sup>1</sup>Engineering Laboratory of Animal Immunity of Jiangsu Province, Institute of Immunology and College of Veterinary Medicine, Nanjing Agricultural University, Nanjing, China

<sup>2</sup>Shenzhen Key Laboratory of Pathogen and Immunity, Shenzhen Third People's Hospital, Shenzhen, China

<sup>3</sup>CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, Beijing, China

<sup>4</sup>CAS Center for Influenza Research and Early-Warning (CASICRE), Chinese Academy of Sciences, Beijing, China

<sup>5</sup>Institute of Pathogen Biology, Taishan Medical College, Tai'an, China

<sup>6</sup>National Institute for Viral Disease Control and Prevention, Chinese Center for Disease Control and Prevention (China CDC), Beijing, China

<sup>7</sup>Kentucky State University, Frankfort, Kentucky, USA

<sup>8</sup>Collaborative Innovation Center for

gastrointestinal symptoms. For instance, transmissible gastroenteritis coronavirus (TGEV) [9,10], bovine coronavirus (BCV) [11], feline coronavirus (FCoV) [12], canine coronavirus (CCoV) [13], and turkey coronavirus (TCV) [14] are known to cause enteritis in their respective hosts [1].

In humans, CoV infections primarily involve the upper respiratory tract and the gastrointestinal tract, and vary from mild, self-limiting disease, such as the common cold, to more severe manifestations, such as bronchitis and pneumonia with renal involvement [15]. The first human coronavirus (HCoV) was isolated during 1965 from the nasal discharge of patients with the common cold and termed B814 [16]. Currently, six different CoV strains are known to infect humans. These include: HCoV-229E (229E), HCoV-OC43 (OC43), severe acute respiratory syndrome coronavirus (SARS-CoV), HCoV-NL63 (NL63), HCoV-HKU1 (HKU1), and Middle East respiratory syndrome coronavirus (MERS-CoV) [17]. 229E and OC43 are the prototype viruses from the two main HCoV lineages (Alpha and Beta, respectively) that cause 15–29% of all common colds, and are the best characterized [18]. SARS-CoV is the aetiological agent that was behind an outbreak of severe respiratory disease through China during 2002–2003 [19], and MERS-CoV is the pathogen responsible for an ongoing outbreak of severe respiratory disease centered in the Middle East since 2012 [20]. In this review, we first compare and contrast the epidemiology and pathogenesis of the known CoVs infecting humans, in which CoVs that cause mild to severe disease in humans are newly emerged from a zoonotic source. Second, we describe the ecology of CoVs, and highlight evidence for viral recombination between the different CoVs within animal populations, which may result in the generation of novel CoVs that are transmissible and lethal to humans. Finally, we summarize the phylogeny, evolution, and genetic recombination of CoVs in detail.

## Epidemiology and Pathogenesis of HCoV

The epidemiology and pathogenesis of each HCoV are discussed within the following sections and summarized in [Table 1](#) and [Figure 1](#), respectively. In general, 229E, OC43, and NL63 are distributed globally ([Figure 1A](#)) and tend to be transmitted predominantly during the winter season in temperate-climate countries [21], while NL63 showed a spring–summer peak of activity from a study in Hong Kong [22].

### Human Coronavirus 229E (alphacoronavirus)

Strain 229E was discovered during 1966, when researchers were characterizing five novel agents that were isolated from the respiratory tract of humans who had contracted the common cold [23]. 229E was adapted to grow in WI-38 lung fibroblast cell lines, and was later shown to be morphologically identical to IBV and MHV [24]. Symptoms of 229E infection include general malaise, headache, nasal discharge, sneezing, and a sore throat [25]. A small portion of patients (10–20%) will also exhibit fever and cough. The incubation time is approximately 2–5 days, followed by illness lasting between 2 and 18 days, and clinically indistinguishable from respiratory tract infections caused by other pathogens, such as rhinovirus and influenza A [18]. 229E is distributed globally ([Figure 1A](#)).

### Human Coronavirus OC43 (betacoronavirus, lineage A)

Strain OC43 was discovered, during 1967, in the nasopharyngeal wash of a patient with the common cold, and is adapted to grow in organ cultures containing suckling mouse brains. Similar to 229E, OC43 is morphologically indistinguishable from IBV and MHV [24], and patients infected with OC43 present the same clinical symptoms as that of 229E [24]. However, there is no serological cross-reactivity between 229E and OC43 [24]. OC43 is also distributed globally ([Figure 1A](#)).

### SARS-CoV (betacoronavirus, lineage B)

Patients infected with SARS-CoV initially present with fever, myalgia, headache, malaise, and chills, followed by a nonproductive cough, dyspnea, and respiratory distress generally

Diagnosis and Treatment of Infectious Disease, Zhejiang University, Hangzhou, China

<sup>9</sup>University of Chinese Academy of Sciences Medical School, Chinese Academy of Sciences, Beijing, China

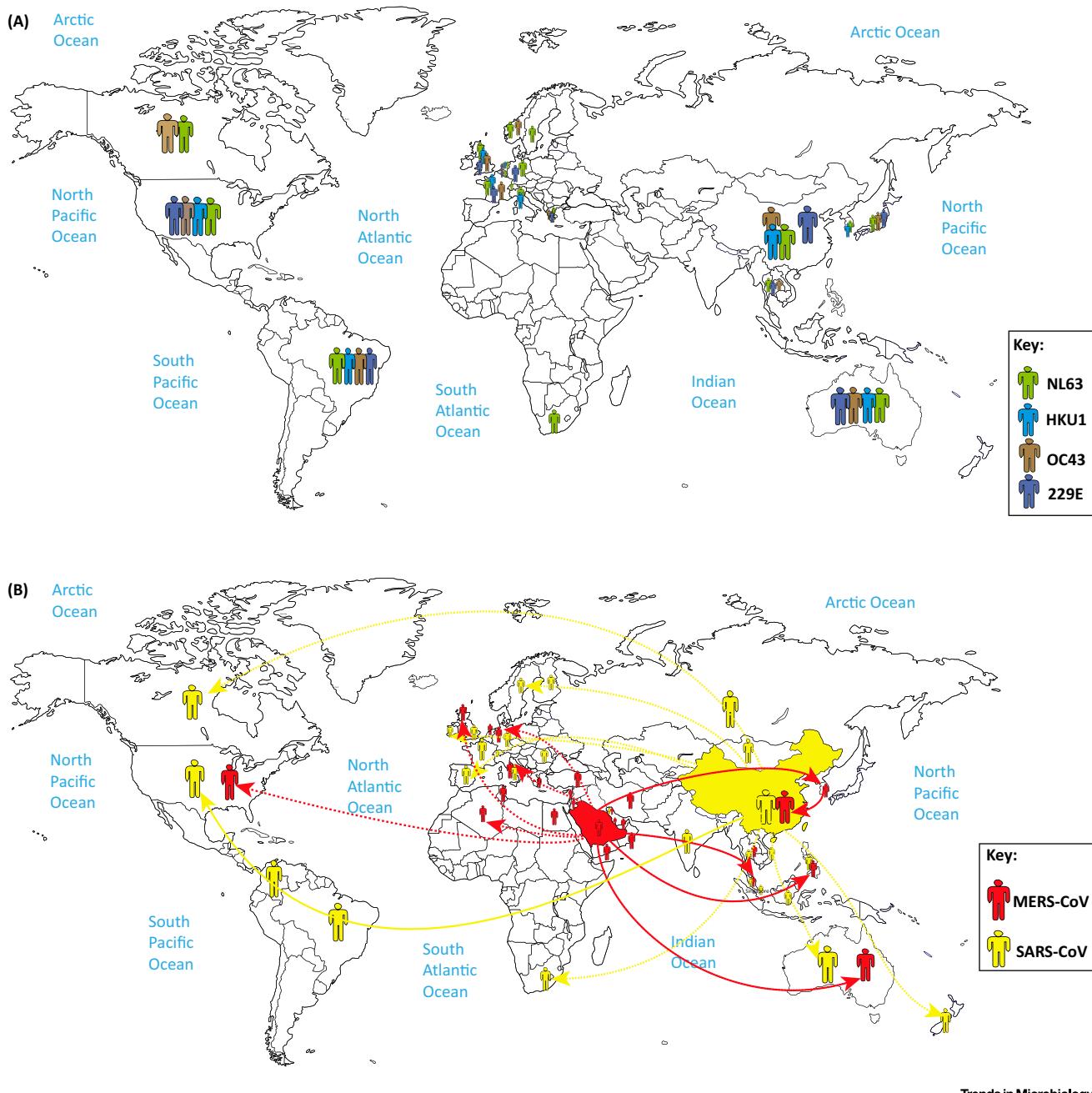
\*Correspondence: [shuosu@njau.edu.cn](mailto:shuosu@njau.edu.cn) (S. Su), [beeyh@im.ac.cn](mailto:beeyh@im.ac.cn) (Y. Bi), [gaof@im.ac.cn](mailto:gaof@im.ac.cn) (G.F. Gao).

Table 1. Comparison of Clinical Symptoms, Incubation Time, Median Time to Death, and Case Fatality Rates of Human Coronaviruses

HCoV	Clinical Symptoms	Case Fatality Rate	Incubation Period	Median Time to Death	Refs
229E	General malaise Headache Nasal discharge Sneezing Sore throat Fever and cough (10–20% of patients)	N/A <sup>a</sup>	2–5 days	—	[18,81,82]
OC43	General malaise Headache Nasal discharge Sneezing Sore throat Fever and cough (10–20% of patients)	N/A	2–5 days	—	[18,82–84]
SARS-CoV	Fever Myalgia Headache Malaise Chills Nonproductive cough Dyspnea Respiratory distress Diarrhea (30–40% of patients)	9%	2–11 days	23 days	[19,42,85–88]
NL63	Cough, Rhinorrhea Tachypnea Fever Hypoxia Obstructive laryngitis (croup)	N/A	2–4 days	—	[89–92]
HKU1	Fever Running nose Cough Dyspnea	N/A	2–4 days	—	[2,32–34,93,94]
MERS-CoV	Fever Cough Chills Sore throat Myalgia Arthralgia Dyspnea Pneumonia Diarrhea and vomiting (one-third of patients) Acute renal impairment	36%	2–13 days	14 days	[41–43,95,96]

<sup>a</sup>N/A, not applicable.

5 to 7 days later, which may result in death (<http://www.who.int/csr/sars/en/WHOconsensus.pdf?ua=1>). Other notable features in some cases include infection of the gastrointestinal tract, liver, kidney, and brain. Diffuse alveolar damage, epithelial cell proliferation, and an increase in macrophages is seen in SARS-CoV infection of the lung. Lymphopenia, hemophagocytosis in the lung, in addition to white-pulp atrophy of the spleen observed in SARS patients, are similar to fatal H5N1 influenza virus infections [1]. Diarrhea is observed in approximately 30–40% of SARS infections (<http://www.cdc.gov/sars/about/faq.html>).



**Figure 1. Global Distribution of Human Coronaviruses.** (A) Green, blue, brown, and purple represent the global distribution of the NL63, HKU1, OC43, and 229E human coronaviruses, respectively. (B) Red and yellow represent the global distribution of MERS-CoV and SARS-CoV, respectively.

An outbreak of disease caused by SARS-CoV, originating from Guangdong Province in southern China during November 2002, eventually spread to other countries in Asia, in addition to North America and Europe (37 countries/regions in total) over 9 months (<http://www.who.int/ith/diseases/sars/en/>) (Figure 1B). An eventual 8273 cases were reported, with 775 deaths for a case fatality rate (CFR) of 9%, and the majority of cases and deaths occurred in mainland China and in Hong Kong [19]. The elderly were more susceptible to SARS disease, with a mortality rate of over 50% (<http://www.cdc.gov/sars/surveillance/absence.html>).

### Human Coronavirus NL63 (alphacoronavirus)

NL63 is primarily associated with young children, the elderly, and immunocompromised patients with respiratory illnesses [26]. NL63 was a novel HCoV isolated from a 7-month-old child with coryza, conjunctivitis, fever, and bronchiolitis in the Netherlands during late 2004 [26], and an independent investigation described the isolation of virtually the same virus from a nasal sample collected from an 8-month-old boy suffering from pneumonia in the Netherlands [27]. NL63 was also detected in New Haven, USA, during 2005 amongst 79 of 895 children, and was initially called HCoV-NH (NH), but genetic and phylogenetic analyses showed that NH and HKU1 likely are the same species [26,28]. Infections with NL63 typically result in mild respiratory disease similar to the common cold, characterized by cough, rhinorrhea, tachypnea, fever, and hypoxia, and resolve on their own [29]. Obstructive laryngitis, also known as croup, is frequently observed with NL63 infections. A study estimated that NL63 accounts for an estimated 4.7% of common respiratory diseases [26]. NL63 is distributed globally (Figure 1A).

### Human Coronavirus HKU1 (betacoronavirus, lineage A)

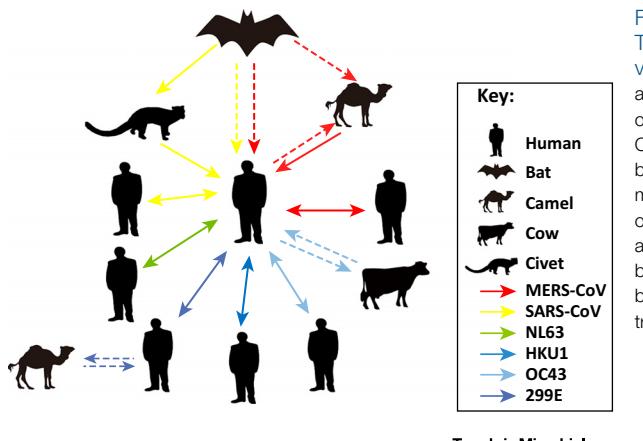
HKU1 was first discovered during January 2005 in a 71-year-old patient from Hong Kong who had been hospitalized with pneumonia and bronchiolitis [30]. The symptoms of HKU1 respiratory tract infections are not able to be separated from those caused by other respiratory viruses. Most patients present with fever, running nose, and cough for infections in the upper respiratory tract, whereas a fever, productive cough, and dyspnea are the common symptoms presenting for infections in the lower respiratory tract [31]. Most HKU1 infections are self-limiting, with only two deaths reported in patients with pneumonia due to HKU1 [32]. Although HKU1 is clinically relatively mild in children, HKU1 infection is associated with a high incidence of seizures, and was found in a patient with meningitis [33,34]. HKU1 cases outside Asia were detected in New Haven, USA, in 2 out of 851 children [35], and also in Australia [36], France [37], and Brazil [38], indicating a global distribution for the virus (Figure 1A).

### MERS-CoV (betacoronavirus, lineage C)

MERS-CoV was first isolated from the lungs of a 60-year-old patient who had died from a severe respiratory illness in Jeddah, Saudi Arabia, 2012 [39]. Clinical manifestations of MERS-CoV infection range from asymptomatic to severe pneumonia with acute respiratory distress, septic shock, and renal failure resulting in death [40]. A typical disease course begins with fever, cough, chills, sore throat, myalgia, and arthralgia, followed by dyspnea and rapid progression to pneumonia [40–43]. Approximately one-third of patients present with gastrointestinal symptoms, such as diarrhea and vomiting. Acute renal impairment was the most striking feature of disease caused by MERS-CoV, which is thus far unique for human CoV infections [40,44]. Seventy-five percent of patients with MERS disease also had at least one other comorbidity, and patients who died were more likely to have a pre-existing/underlying condition [40]. Countries around the Arabian Peninsula are known to be endemic for MERS-CoV, and Saudi Arabia has reported the most cases, but since its discovery in 2012, cases have been occasionally exported to other countries through travel, sometimes causing clusters of secondary outbreaks (Figure 1B) [45]. As of December 31, 2015, a total of 1621 laboratory-confirmed infections have been reported, with 584 deaths (CFR = 36.0%) over 26 countries ([http://www.who.int/csr/disease/coronavirus\\_infections/en](http://www.who.int/csr/disease/coronavirus_infections/en)), making MERS-CoV one of the most dangerous viruses known to humans.

### Ecology of Human Coronaviruses

Based on currently available evidence, 229E, OC43, NL63, and HKU1 are well adapted to humans, and the viruses widely circulate in the human population (Figure 2), with most cases causing mild disease in immunocompetent adults, and none of these viruses have been found to be maintained within an animal reservoir. However, in the case of SARS-CoV and MERS-CoV, the viruses are not as well adapted to being maintained in humans, and thus are likely spread



**Figure 2. Intra- and Inter-Species Transmission of Human Coronaviruses.** Red, yellow, green, blue, brown, and purple arrows represent transmission of MERS-CoV, SARS-CoV, NL63, HKU1, OC43, and 229E, respectively, between bats, camels, cows, humans, and masked palm civets (shown in a legend on the side of the figure). Unbroken arrows represent confirmed transmission between the two species in question, and broken arrows represent suspected transmission.

Trends in Microbiology

mainly in zoonotic reservoir(s), with occasional spillover into the susceptible human population, possibly via an intermediate host species.

Regarding SARS-CoV, epidemiology data implicated masked palm civets (*Paguma larvata*) from live animal markets (LAM) in Guangdong Province, China, as a route of exposure to SARS-CoV [46]. However, masked palm civets from the wild or from farms without LAM exposure were largely negative for SARS-CoV [47]. This suggests that palm civets were an intermediate host, but not a reservoir for SARS-CoV [48]. Subsequent studies have shown that wild horseshoe bats (Rhinolophidae family), which can also be found in LAM in China and served in some Chinese restaurants in Guangdong, China, have detectable levels of antibodies against SARS-CoV and a SARS-CoV-like virus (SARSr-Rh-Bat CoV) [49,50], suggesting a bat origin for SARS-CoV. An evolutionary relationship between coronaviruses and bats was proposed, in which the ancestor for SARS-CoV first spread to bats of the Hipposideridae family, then Rhinolophidae, then masked palm civets and eventually humans [51]. Recently, two new SARS-like CoVs were isolated in horseshoe bats, and these viruses showed the highest relationship to SARS-CoV from all known bat coronaviruses [52]. ORF8 analysis of the SARS-like CoVs in bats suggests that Chinese horseshoe bats are the natural reservoirs of SARS-CoV [52,53], and that intermediate hosts might not be needed for direct human infection, particularly for some bat SARS-CoV-like viruses [52].

In the case of MERS-CoV, studies in Oman, Saudi Arabia, Qatar, United Arab Emirates, and Jordan have shown that dromedary camels are seropositive for neutralizing antibodies against MERS-CoV [54], as well as camels of Middle East origin in Africa, including Egypt, Kenya, Nigeria, Ethiopia, Tunisia, Somalia, and Sudan [55]. Subsequent studies on dromedary camels in Saudi Arabia [56], Qatar [57], and Egypt [58] show that live MERS-CoV can be isolated primarily from the nasal swabs of camels, showing that camels are potential source of MERS-CoV infection. However, many confirmed cases lack contact history with camels [59], suggesting direct human-to-human transmission, or through contact with a yet-to-be-identified animal species that is maintaining MERS-CoV. Studies on HKU4, a coronavirus of bat origin and the most phylogenetically closely related to MERS-CoV, had shown that HKU4 is able to utilize the CD26 receptor for virus entry [60]. As CD26 is a known receptor for MERS-CoV [61], the similarity in receptor specificity of these two CoVs supports the hypothesis that MERS-CoVs is also a bat-originated CoV. However, live MERS-CoV has not been isolated from wild bats, nor have viral sequences been detected.

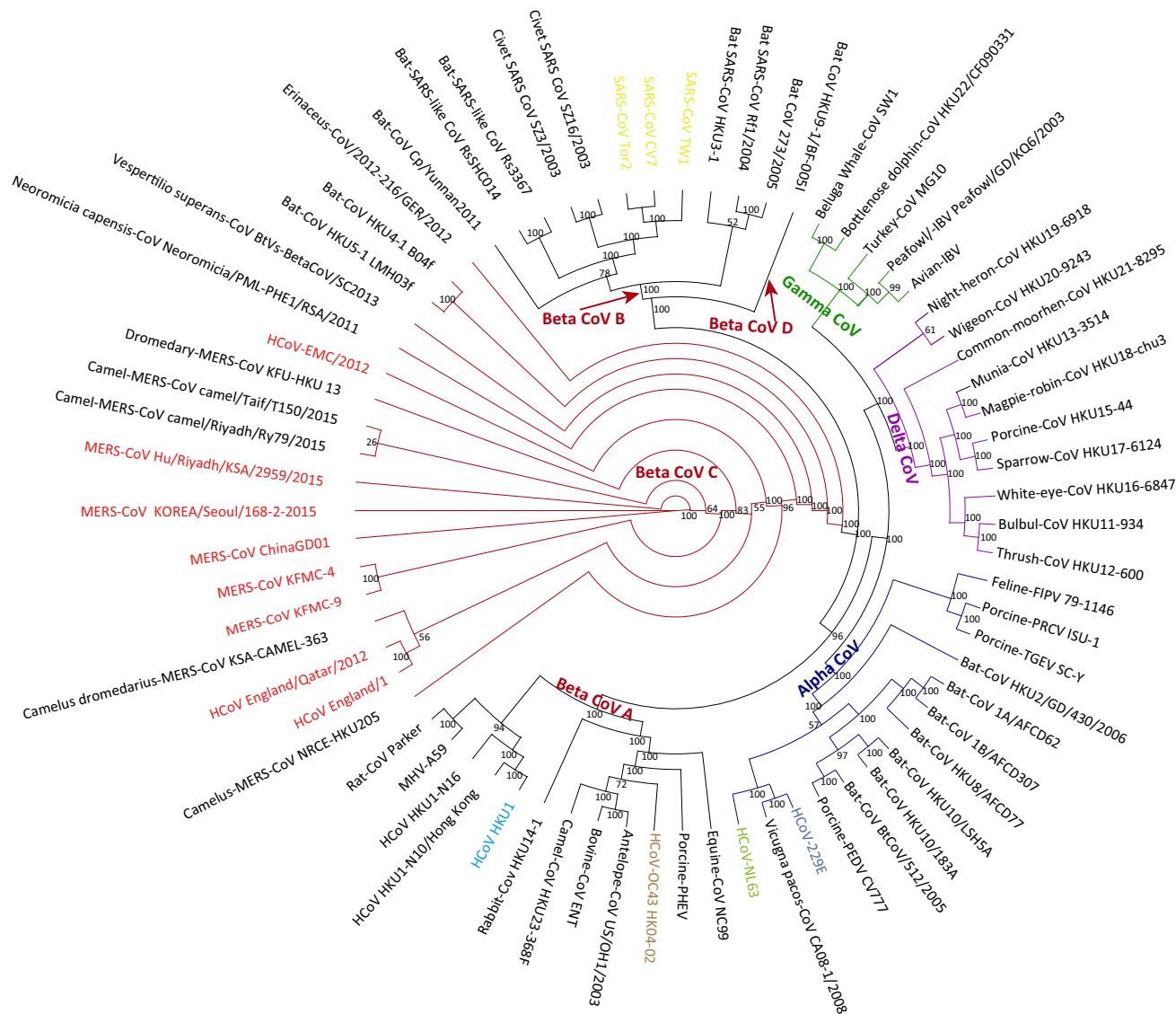
A recent investigation discovered that multiple HCoV species, including MERS-CoV, beta-CoV group A, and a 229E-like virus, circulate amongst dromedary camels in Saudi Arabia [62]. Furthermore, multiple lineages of circulating MERS-CoV in these camels had resulted in a dominant,

recombinant MERS-CoV strain that was responsible for the MERS-CoV outbreaks during 2015 [62]. These results show that various CoVs of human and animal origin are currently circulating in the wild, providing ample opportunity for CoVs to undergo evolution and genetic recombination, thereby resulting in recombinant CoVs that may potentially be more deadly to humans. In the following sections, we summarize the evolution and genetic recombination of the current HCoVs.

## Phylogeny, Evolution and Genetic Recombination in Human Coronaviruses

### Phylogenetic Analysis

The phylogenetic tree for the CoVs of zoonotic and human origin was constructed based on the full genome. It can be observed that 229E and NL63 belong to the alpha-CoV genus and are



**Figure 3. Phylogenetic Tree of Known Animal and Human Coronaviruses.** Phylogenetic analysis of the full-length genome sequences of coronaviruses publicly available in GenBank was performed using RAxML, with 1000 bootstrap replicates. Lines with different colors represent different coronavirus species, Alpha-coronaviruses (blue lines), Beta-coronaviruses (red lines for Beta-coronavirus C, and black lines for Beta-coronavirus A, B, and D), Gamma-coronaviruses (green lines), and Delta-coronaviruses (purple lines). Human isolates are highlighted with different colors, whereas strains from other hosts are shown in black. Numbers at the branches represent bootstrap values obtained in the phylogenetic analysis.

grouped together with porcine epidemic diarrhea virus (PEDV), TGEV, porcine respiratory coronavirus (PRCV), feline infectious peritonitis virus (FIPV), and some bat-derived coronaviruses (Figure 3). The topology of the phylogeny is similar to the tree constructed based on the RNA-dependent RNA polymerase gene [55].

OC43, HKU1, SARS-CoV, and MERS-CoV belong to the beta-CoV genus. Lineage A includes OC43 and HKU1, MHV, porcine hemagglutinating encephalomyelitis virus (PHEV), in addition to equine, rabbit, camel, bovine, antelope-derived animal coronaviruses. Lineage B includes SARS-CoV, SARS-like viruses of bat and palm civet origin, and some bat-derived CoVs. Lineage C includes MERS-CoV and some bat-derived viruses. Lineage D contains only bat-derived coronavirus (Figure 3). In contrast to alpha- and beta-CoVs, gamma-CoVs consist mainly of avian coronaviruses (such as IBV), as well as CoVs isolated from aquatic animals, whales, and dolphins. CoVs of wild-bird origin are clustered into the delta-CoV, which also contains some swine-derived CoVs (Figure 3).

In general, alpha- and beta-CoVs primarily include CoVs from mammals, whereas gamma- and delta-CoVs mostly include CoVs of avian origin. Notably, all currently known HCoVs belong to either the alpha- or betacoronavirus genera, which also contain CoVs of mainly bat origin.

#### Mutation and Genetic Recombination in Coronaviruses

The estimated mutation rates in CoV are moderate to high compared to other single-stranded RNA (ssRNA) viruses, and the average substitution rate for CoVs was  $\sim 10^{-4}$  substitutions per year per site [63]. The nucleotide mutation rate of IBV for the hypervariable region in the S gene was estimated to be  $0.3\text{--}0.6 \times 10^{-2}$  per site per year [64]. The S gene of OC43 possessed an average rate of  $6.410.58 \times 10^{-4}$  substitution rates per site per year [65]. Additionally, the S gene of 229E possessed a rate of  $\sim 3 \times 10^{-4}$  substitutions per site per year [63]. For SARS-CoV, the mutation rate in the whole genome was estimated to be  $0.80\text{--}2.38 \times 10^{-3}$  nucleotide substitutions per site per year, and the nonsynonymous and synonymous substitution rates were estimated to be  $1.16\text{--}3.30 \times 10^{-3}$  and  $1.67\text{--}4.67 \times 10^{-3}$  per site per year, respectively, which is similar to other RNA viruses [66]. The evolutionary rate for MERS-CoV genomes was estimated as  $1.12 \times 10^{-3}$  substitutions per site per year (95% credible interval [95% CI],  $8.76 \times 10^{-4}$ ;  $1.37 \times 10^{-3}$ ) [67]. Furthermore, the large RNA genome in CoV allows for extra plasticity in genome modification by mutations and recombinations, thereby increasing the probability for intraspecies variability, interspecies ‘host jump’, and novel CoVs to emerge under the right conditions [68–70].

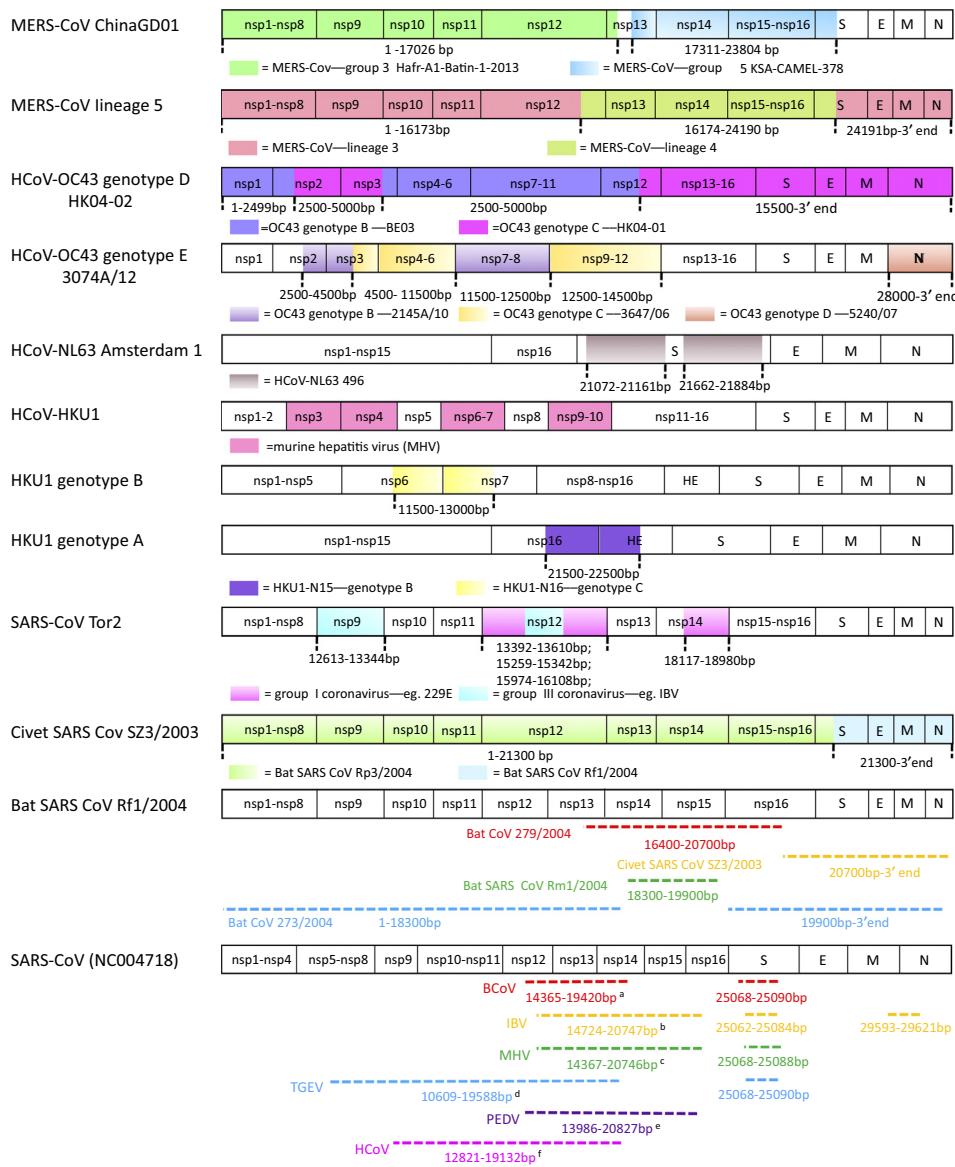
The major reason for recombination may be at the replication step in the virus life cycle. During replication, a set of subgenomic RNAs is generated, increasing the homologous recombination rate among closely related genes from different lineages of CoVs or other viruses by template switching [33]. Concurrently, circulating CoVs in multiple host species likely contribute to increases in the rate of recombination events. However, the exact mechanism of genetic recombination in CoVs remains unclear: recombination site ‘breakpoints’ in the viral genome – the crossing point of the recombinant genes between two distinct viral strains or genotypes – appear to be random, as different recombinant strains have different breakpoints.

Genetic recombination has been previously documented for animal CoVs, for instance: MHV [71], TGEV [72], as well as feline and canine coronaviruses [73,74]. Genetic recombination for other HCoVs including OC43, NL63, HKU1, SARS-CoV [51], and MERS-CoV have also been documented and are discussed in detail below.

OC43 has diverged into five distinct genotypes (A to E). Genotype A is typified by the prototype VR759 strain that was isolated in 1967 [75]. Genotypes B and C are two naturally circulating viruses. Genotype D emerged via recombination between genotypes B and C viruses at the

**Key Figure**

## Genetic Recombination in Human Coronaviruses



BCoV: bovine coronavirus; IBV: avian infectious bronchitis virus; MHV: murine hepatitis

TGEV: transmissible gastroenteritis; PEDV: porcine epidemic diarrhea virus

HCoV: human coronavirus

a:14365-14412, 14490-14523, 15210-15238, 19206-19253, 19396-19420

b:14724-14746, 15210-15253, 15417-15457, 17088-17125, 19102-19132, 20712-20747

c:14367-14395, 14986-15055, 15417-15457, 17088-17125, 19102-19132, 20712-20747

d:10609-10641, 17688-17714, 19548-19588

e:13986-14011, 15210-15232, 18558-18577, 20797-20827

f:12821-12854, 14808-14835, 15062-15093, 15624-15670, 17783-17809, 19113-19132

Trends in Microbiology

**Figure 4.** The colored regions of each human coronavirus indicate genetic recombination at the genomic sites with coronaviruses of zoonotic origin. Specific nucleotide locations where recombination occurred are also shown in broken lines.

breakpoint of 2500–5000 bp in the NSP2–NSP3 gene and 15500 bp-3'end in NSP12-N genes [75] (Figure 4, Key Figure). Genotype E, identified in 2014, emerged by recombination amongst the genotype B, C, and D viruses with different breakpoints in the genes compared to the generation of genotype D [76] (Figure 4). A study in France, between 2001 and 2013, demonstrated that OC43 strains have more recombination patterns in Lower Normandy, and circulating OC43 viruses have a high genetic diversity [17].

NL63 also has the potential for genetic recombination, and one such example is between the Amsterdam 1 and 496 strains. Recombination site breakpoints were identified in two parts of the S gene [63] (Figure 4).

Recombination events have been observed for different genotypes of HKU1, in which recombination breakpoints were found on genotypes B and C in the nsp6-nsp7 and nsp16-HE genes, respectively [70] (Figure 4). HKU1 was also shown to be able to recombine with other animal beta-CoVs, such as MHV, at the nsp3, nsp4, nsp6, nsp7, nsp9, and nsp10 genes [77] (Figure 4).

SARS-CoV has a possible recombinant history with lineages of alpha- and gamma-CoV [78]. Many specific breakpoints and a large number of smaller recombinant regions had been identified in the RNA-dependent RNA polymerase (RDRP, nsp12 gene) at the following locations: nucleotides 13 392–13 610, 15 259–15 342, and 15 974–16 108 based on the sequence of the SARS-CoV TOR2 strain [78]. Recombination locations were also identified in the nsp9 and most of nsp10 gene located at 12 613–13 344, and parts of nsp14 (18 117–18 980) [78]. Another study analyzing the genome of the SARS-CoV CV7 strain had shown that seven putative recombination locations were found between SARS-CoV and six other coronaviruses: PEDV, TGEV, BCV, 229E, MHV, and IBV (Figure 4). This further suggests that SARS-CoV emerged via serial horizontal transmission and by genetic recombination, enhancing the adaptation process to its new host [79].

Recombination was also detected between the SARS-related bat-associated CoV (SARSr-Bat CoVs), in which recombination between the Rp3 strain from Guangxi Province and Rf1 strain from Hubei Province, China, generated a new strain (SARSr-Civet CoV SZ3) in civets, with a breakpoint at the nsp16/spike and S2 region [51] (Figure 4). Furthermore, the SARS-CoV Rf1/2004 strain is itself a recombinant from the bat CoV 279/2004, bat CoV Rm1/2004 and civet CoV SZ3/2003 [51] (Figure 4).

A recent study showed that the epidemic MERS-CoV experienced recombination events between the different lineages, which occurred in dromedary camels in Saudi Arabia [62]. Lineages 3 and 5 viruses were predominant during July and December 2014; however, lineage 5 became dominant within the camel population by 2015. Lineage 5, which is closely associated with the MERS-CoV causing the outbreak in South Korea in 2015, as well as recent human infections in Riyadh, Saudi Arabia, is a recombinant virus between lineages 3 and 4, or groups 3 and 5 of clade B (Figure 4) [62,80].

### Concluding Remarks

The result from a high frequency of recombination events in CoVs is the generation of novel viruses with a high genetic diversity, with unpredictable changes in virulence during human infections. With multiple species of CoVs circulating in the wild amongst different animal species that may constantly interact with one another, it is likely not a matter of if, but when, the next recombinant CoV will emerge and cause another outbreak in the human population. As such, some crucial future areas of investigation include: (i) the prevalence of HCoVs already circulating within the animal population, (ii) the commonality of coronavirus recombination in animals, (iii) animals which may potentially serve as mixing vessels for the generation of novel recombinant

### Outstanding Questions

How common is coronavirus recombination in animals?

Can coronavirus undergo recombination in animal hosts other than camels?

Is there a way to predict the potential emergence of a novel recombinant coronavirus in the human population?

CoVs, and (iv) a surveillance network to monitor and predict the potential emergence of a highly virulent, recombinant CoV from animals (see Outstanding Questions). Furthermore, lessons from the SARS-CoV and MERS-CoV outbreaks must be urgently learned in advance to effectively prepare for the next CoV outbreak.

### Author Contributions

Shuo Su and Gary Wong contributed equally to this manuscript.

### Acknowledgments

This work was supported by the National Natural Science Foundation of China (81461168030), the China National Grand S&T Special Project (2014ZX10004-001-006), and the National Key Program for Infectious Disease of China (2016ZX10004222). G.F.G. is a leading principal investigator of the NSFC Innovative Research Group (grant number 81321063). J.Z and S.S are supported by the Priority Academic Program Development of Jiangsu Higher Education Institution. G.W. is the recipient of a Banting Postdoctoral Fellowship from the Canadian Institutes of Health Research (CIHR), and a President's International Fellowship Initiative from the Chinese Academy of Sciences (CAS). All authors declare no competing interests.

### References

- Weiss, S.R. and Navas-Martin, S. (2005) Coronavirus pathogenesis and the emerging pathogen severe acute respiratory syndrome coronavirus. *Microbiol. Mol. Biol. Rev.* 69, 635–664
- Lai, M. et al. (2007) Coronaviridae. In *Fields Virology* (5th edn) (Knipe, D.M. and Howley, P.M., eds), pp. 1305–1335, Lippincott Williams & Wilkins
- Lai, M.M. and Cavanagh, D. (1997) The molecular biology of coronaviruses. *Adv. Vir. Res.* 48, 1–100
- Langereis, M.A. et al. (2010) Attachment of mouse hepatitis virus to O-acetylated sialic acid is mediated by hemagglutinin-esterase and not by the spike protein. *J. Virol.* 84, 8970–8974
- Durham, P.J. et al. (1979) Rotavirus and coronavirus associated diarrhoea in domestic animals. *N. Zealand Vet. J.* 27, 30–32
- Cavanagh, D. (2007) Coronavirus avian infectious bronchitis virus. *Vet. Res.* 38, 281–297
- Erles, K. et al. (2003) Detection of a group 2 coronavirus in dogs with canine infectious respiratory disease. *Virology* 310, 216–223
- Weiner, L.P. (1973) Pathogenesis of demyelination induced by a mouse hepatitis. *Arch. Neurol.* 28, 298–303
- Pensaert, M. et al. (1970) Transmissible gastroenteritis of swine: virus-intestinal cell interactions. II. Electron microscopy of the epithelium in isolated jejunal loops. *Arch. Gesamte Virusforsch.* 31, 335–351
- Pensaert, M. et al. (1970) Transmissible gastroenteritis of swine: virus-intestinal cell interactions. I. Immunofluorescence, histopathology and virus production in the small intestine through the course of infection. *Arch. Gesamte Virusforsch.* 31, 321–334
- Brider, J.C. et al. (1978) Replication of an enteric bovine coronavirus in intestinal organ cultures. *Arch. Virol.* 57, 43–51
- Pedersen, N. et al. (1984) Pathogenicity studies of feline coronavirus isolates 79-1146 and 79-1683. *Am. J. Vet. Res.* 45, 2580–2585
- Binn, L.N. et al. (1974) Recovery and Characterization of a Coronavirus from Military Dogs with Diarrhea. *Proc. Annu. Meet. U.S. Anim. Health Assoc.* 359–366
- Ismail, M.M. et al. (2003) Pathogenicity of turkey coronavirus in turkeys and chickens. *Avian Dis.* 47, 515–522
- Wevers, B.A. and van der Hoek, L. (2009) Recently discovered human coronaviruses. *Clin. Lab. Med.* 29, 715–724
- Tyrrell, D.A. and Bynoe, M.L. (1965) Cultivation of a novel type of common-cold virus in organ cultures. *Br. Med. J.* 1, 1467–1470
- Kin, N. et al. (2015) Genomic analysis of 15 human coronaviruses OC43 (HCoV-OC43s) circulating in France from 2001 to 2013 reveals a high intra-specific diversity with new recombinant genotypes. *Viruses* 7, 2358–2377
- Monto, A.S. (1974) Medical reviews. Coronaviruses. *Yale J. Biol. Med.* 47, 234–251
- Peiris, J.S.M. et al. (2003) Coronavirus as a possible cause of severe acute respiratory syndrome. *Lancet* 361, 1319–1325
- Raj, V.S. et al. (2014) MERS: emergence of a novel human coronavirus. *Curr. Opin. Virol.* 5, 58–62
- Hendley, J.O. et al. (1972) Coronavirus infections in working adults. Eight-year study with 229 E and OC 43. *Am. Rev. Respir. Dis.* 105, 805–811
- Chiu, S.S. et al. (2005) Human coronavirus NL63 infection and other coronavirus infections in children hospitalized with acute respiratory disease in Hong Kong. *China. Clin. Infect. Dis.* 40, 1721–1729
- Hamre, D. and Procknow, J.J. (1966) A new virus isolated from the human respiratory tract. *Proc. Soc. Exp. Biol. Med.* 121, 190–193
- McIntosh, K. et al. (1967) Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. *Proc. Natl. Acad. Sci. U.S.A.* 57, 933–940
- Tyrrell, D.A. et al. (1993) Signs and symptoms in common colds. *Epidemiol. Infect.* 111, 143–156
- van der Hoek, L. et al. (2006) Human coronavirus NL63, a new respiratory virus. *FEMS Microbiol. Rev.* 30, 760–773
- Fouchier, R.A. et al. (2004) A previously undescribed coronavirus associated with respiratory disease in humans. *Proc. Natl. Acad. Sci. U.S.A.* 101, 6212–6216
- Esper, F. et al. (2005) Evidence of a novel human coronavirus that is associated with respiratory tract disease in infants and young children. *J. Infect. Dis.* 191, 492–498
- Abdul-Rasool, S. and Fielding, B.C. (2010) Understanding human coronavirus HCoV-NL63. *Open Virol. J.* 4, 76–84
- Woo, P.C. et al. (2005) Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. *J. Virol.* 79, 884–895
- Woo, P.C. et al. (2009) More and more coronaviruses: human coronavirus HKU1. *Viruses* 1, 57–71
- Woo, P.C. et al. (2005) Clinical and molecular epidemiological features of coronavirus HKU1-associated community-acquired pneumonia. *J. Infect. Dis.* 192, 1898–1907
- Lau, S.K.P. et al. (2006) Coronavirus HKU1 and other coronavirus infections in Hong Kong. *J. Clin. Microbiol.* 44, 2063–2071
- Gaunt, E.R. et al. (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, 2940–2947
- Esper, F. et al. (2006) Coronavirus HKU1 infection in the United States. *Emerg. Infect. Dis.* 12, 775–779
- Sloots, T.P. et al. (2006) Evidence of human coronavirus HKU1 and human bocavirus in Australian children. *J. Clin. Virol.* 35, 99–102

37. Vabret, A. et al. (2006) Detection of the new human coronavirus HKU1: a report of 6 cases. *Clin. Infect. Dis.* 42, 634–639
38. Goes, L.G. et al. (2011) Coronavirus HKU1 in children, Brazil, 1995. *Emerg. Infect. Dis.* 17, 1147–1148
39. Zaki, A.M. et al. (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N. Engl. J. Med.* 367, 1814–1820
40. Zumla, A. et al. (2015) Middle East respiratory syndrome. *Lancet* 386, 995–1007
41. Al-Tawfiq, J.A. et al. (2014) Middle East respiratory syndrome coronavirus: a case-control study of hospitalized patients. *Clin. Infect. Dis.* 59, 160–165
42. Assiri, A. et al. (2013) Epidemiological, demographic, and clinical characteristics of 47 cases of Middle East respiratory syndrome coronavirus disease from Saudi Arabia: a descriptive study. *Lancet Infect. Dis.* 13, 752–761
43. Arabi, Y.M. et al. (2014) Clinical course and outcomes of critically ill patients with Middle East respiratory syndrome coronavirus infection. *Ann. Intern. Med.* 160, 389–397
44. Drosten, C. et al. (2013) Clinical features and virological analysis of a case of Middle East respiratory syndrome coronavirus infection. *Lancet Infect. Dis.* 13, 745–751
45. Su, S. et al. (2015) MERS in South Korea and China: a potential outbreak threat? *Lancet* 385, 2349–2350
46. Guan, Y. et al. (2003) Isolation and characterization of viruses related to the SARS coronavirus from animals in southern China. *Science* 302, 276–278
47. Poon, L.L. et al. (2005) Identification of a novel coronavirus in bats. *J. Virol.* 79, 2001–2009
48. Perlman, S. and Netland, J. (2009) Coronaviruses post-SARS: update on replication and pathogenesis. *Nat. Rev. Microbiol.* 7, 439–450
49. Tu, C. et al. (2004) Antibodies to SARS coronavirus in civets. *Emerg. Infect. Dis.* 10, 2244–2248
50. Kan, B. et al. (2005) Molecular evolution analysis and geographic investigation of severe acute respiratory syndrome coronavirus-like virus in palm civets at an animal market and on farms. *J. Virol.* 79, 11892–11900
51. Lau, S.K. et al. (2010) Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related Rhinolophus bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. *J. Virol.* 84, 2808–2819
52. Ge, X.Y. et al. (2013) Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* 503, 535–538
53. Wu, Z. et al. (2016) ORF8-related genetic evidence for Chinese horseshoe bats as the source of human severe acute respiratory syndrome coronavirus. *J. Infect. Dis.* 213, 579–583
54. Reusken, C.B. et al. (2013) Middle East respiratory syndrome coronavirus neutralising serum antibodies in dromedary camels: a comparative serological study. *Lancet Infect. Dis.* 13, 859–866
55. Chan, J.F.W. et al. (2015) Middle East respiratory syndrome coronavirus: another zoonotic betacoronavirus causing SARS-like disease. *Clin. Microbiol. Rev.* 28, 465–522
56. Hemida, M.G. et al. (2014) MERS coronavirus in dromedary camel herd, Saudi Arabia. *Emerg. Infect. Dis.* 20, 1231–1234
57. Raj, V.S. et al. (2014) Isolation of MERS coronavirus from a dromedary camel, Qatar, 2014. *Emerg. Infect. Dis.* 20, 1339–1342
58. Chu, D.K. et al. (2014) MERS coronaviruses in dromedary camels, Egypt. *Emerg. Infect. Dis.* 20, 1049–1053
59. Samara, E.M. and Abdoun, K.A. (2014) Concerns about misinterpretation of recent scientific data implicating dromedary camels in epidemiology of Middle East respiratory syndrome (MERS). *Mbio* 5, e01430–e1514
60. Wang, Q. et al. (2014) Bat origins of MERS-CoV supported by bat coronavirus HKU4 usage of human receptor CD26. *Cell Host Microbe* 16, 328–337
61. Lu, G. et al. (2013) Molecular basis of binding between novel human coronavirus MERS-CoV and its receptor CD26. *Nature* 500, 227–231
62. Sabir, J.S. et al. (2016) Co-circulation of three camel coronavirus species and recombination of MERS-CoVs in Saudi Arabia. *Science* 351, 81–84
63. Pyrc, K. et al. (2006) Mosaic structure of human coronavirus NL63, one thousand years of evolution. *J. Mol. Biol.* 364, 964–973
64. Cavanagh, D. et al. (1998) Does IBV change slowly despite the capacity of the spike protein to vary greatly? *Adv. Exp. Med. Biol.* 440, 729–734
65. Ren, L. et al. (2015) Genetic drift of human coronavirus OC43 spike gene during adaptive evolution. *Sci. Rep.* 5, 11451
66. Zhao, Z. et al. (2004) Moderate mutation rate in the SARS coronavirus genome and its implications. *BMC Evol. Biol.* 4, 21
67. Cotten, M. et al. (2014) Spread, circulation, and evolution of the Middle East respiratory syndrome coronavirus. *Mbio* 5, e01062–e1113
68. Woo, P.C. et al. (2009) Coronavirus diversity, phylogeny and interspecies jumping. *Exp. Biol. Med. (Maywood)* 234, 1117–1127
69. Woo, P.C. et al. (2006) Infectious diseases emerging from Chinese wet-markets: zoonotic origins of severe respiratory viral infections. *Curr. Opin. Infect. Dis.* 19, 401–407
70. Woo, P.C. et al. (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
71. Lai, M.M. et al. (1985) Recombination between nonsegmented RNA genomes of murine coronaviruses. *J. Virol.* 56, 449–456
72. Tian, P.F. et al. (2014) Evidence of recombinant strains of porcine epidemic diarrhea virus, United States, 2013. *Emerg. Infect. Dis.* 20, 1735–1738
73. Terada, Y. et al. (2014) Emergence of pathogenic coronaviruses in cats by homologous recombination between feline and canine coronaviruses. *PLoS ONE* 9, e106534
74. Decaro, N. et al. (2009) Recombinant canine coronaviruses related to transmissible gastroenteritis virus of Swine are circulating in dogs. *J. Virol.* 83, 1532–1537
75. Lau, S.K.P. 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, 11325–11337
76. Zhang, Y. et al. (2015) Genotype shift in human coronavirus OC43 and emergence of a novel genotype by natural recombination. *J. Infect.* 70, 641–650
77. Woo, P.C. et al. (2005) Phylogenetic and recombination analysis of coronavirus HKU1, a novel coronavirus from patients with pneumonia. *Arch. Virol.* 150, 2299–2311
78. Stanshope, M.J. et al. (2004) Evidence from the evolutionary analysis of nucleotide sequences for a recombinant history of SARS-CoV. *Infect. Genet. Evol.* 4, 15–19
79. Zhang, X.W. et al. (2005) Testing the hypothesis of a recombinant origin of the SARS-associated coronavirus. *Arch. Virol.* 150, 1–20
80. Wang, Y. et al. (2015) Origin and possible genetic recombination of the Middle East respiratory syndrome coronavirus from the first imported case in China: phylogenetics and coalescence analysis. *Mbio* 6, e01280–e1215
81. Pene, F. et al. (2003) Coronavirus 229E-related pneumonia in immunocompromised patients. *Clin. Infect. Dis.* 37, 929–932
82. Papa, A. et al. (2007) Coronaviruses in children, Greece. *Emerg. Infect. Dis.* 13, 947–949
83. Cabeca, T.K. et al. (2013) Epidemiological and clinical features of human coronavirus infections among different subsets of patients. *Influenza Other Respir. Virus* 7, 1040–1047
84. Dijkman, R. et al. (2013) Isolation and characterization of current human coronavirus strains in primary human epithelial cell cultures reveal differences in target cell tropism. *J. Virol.* 87, 6081–6090
85. Cleri, D.J. et al. (2010) Severe Acute Respiratory Syndrome (SARS). *Infect. Dis. Clin. N. Am.* 24, 175–202
86. Gu, J. et al. (2005) Multiple organ infection and the pathogenesis of SARS. *J. Exp. Med.* 202, 415–424
87. Tsang, K.W. et al. (2003) A cluster of cases of severe acute respiratory syndrome in Hong Kong. *N. Engl. J. Med.* 348, 1977–1985
88. Peiris, J.S.M. et al. (2003) Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: a prospective study. *Lancet* 361, 1767–1772

89. Moes, E. *et al.* (2005) A novel pan-coronavirus RT-PCR assay: frequent detection of human coronavirus NL63 in children hospitalized with respiratory tract infections in Belgium. *BMC Infect. Dis.* 5, 6
90. Bastien, N. *et al.* (2005) Human coronavirus NL63 infection in Canada. *J. Infect. Dis.* 191, 503–506
91. van der Hoek, L. *et al.* (2005) Croup is associated with the novel coronavirus NL63. *PLoS Med.* 2, 764–770
92. Chiu, S.S. *et al.* (2005) Human coronavirus NL63 infection and other coronavirus infections in children hospitalized with acute respiratory disease in Hong Kong, China. *Clin. Infect. Dis.* 40, 1721–1729
93. Wat, D. (2004) The common cold: a review of the literature. *Eur. J. Intern. Med.* 15, 79–88
94. Chung, J.Y. *et al.* (2007) Detection of viruses identified recently in children with acute wheezing. *J. Med. Virol.* 79, 1238–1243
95. WHO Mers-CoV Research Group (2013) State of knowledge and data gaps of Middle East respiratory syndrome coronavirus (MERS-CoV) in humans. *PLoS Curr. Outbreaks* Published online November 12, 2013. <http://dx.doi.org/10.1371/currents.outbreaks.0bf719e352e7478f8ad85fa30127ddb8>
96. Corman, V.M. *et al.* (2016) Viral shedding and antibody response in 37 patients with Middle East respiratory syndrome coronavirus infection. *Clin. Infect. Dis.* 62, 477–483