

# Human Coronaviruses: General Features

Xin Li and Hayes KH Luk, The University of Hong Kong, Hong Kong, China

Susanna KP Lau and Patrick CY Woo, The University of Hong Kong, Hong Kong, China; Zhejiang University, Hangzhou, China

© 2019 Elsevier Inc. All rights reserved.

Introduction	1
Taxonomy and Classification	1
Virion Structure	1
Genome Organization	2
Replication	4
Recombination and Evolution	4
Epidemiology and Clinical Features of Human Coronaviruses	5
Emerging Human Coronaviruses	5
Acknowledgments	6
Further Reading	6

## Introduction

Coronaviruses are enveloped, non-segmented, positive-sense RNA viruses that are known to infect humans and a wide variety of animals, causing mainly respiratory diseases in humans, although other organ systems are also affected with varying severity. Two highly pathogenic coronaviruses have claimed thousands of lives globally in the past two decades, leading to renewed interest in the study of their evolution, transmission and pathogenicity. Severe acute respiratory syndrome coronavirus (SARS-CoV) emerged in southern China in 2002/2003 and spread rapidly to cause a global pandemic, leading to more than 8000 confirmed cases of SARS with a case fatality rate of nearly 10%. Almost 10 years after the SARS outbreak, Middle East respiratory syndrome coronavirus (MERS-CoV) emerged as a highly fatal human pathogen in the Arabian Peninsula in 2012 with even higher mortality approaching 40%. Before the SARS-CoV pandemic, only two human coronaviruses (HCoVs) were known, namely HCoV-229E and HCoV-OC43. After 2003, two more human coronaviruses HCoV-NL63 and HCoV-HKU1 were identified. These four HCoVs typically cause mild, self-limiting upper respiratory infections in humans, although occasional cases of severe lower respiratory infection have been reported. Over the past two decades, studies in searching of the zoonotic origin of SARS-CoV and MERS-CoV have seen major breakthroughs. As a result of inherently high mutation rates and high frequency of recombination, coronaviruses manifest rapid adaptation to new host receptors with the ability to overcome interspecies barrier. Therapies and preventive strategies such as vaccination are generally limited for these emerging zoonotic pathogens, leaving few treatment options for fatal human infections.

## Taxonomy and Classification

Coronaviruses form the largest group of viruses under the order *Nidovirales*, which includes the families *Coronaviridae*, *Arteriviridae*, *Roniviridae* and *Mesoviridae*, with highly conserved genomic organization and 3' nested subgenomic mRNAs (*nido*, Latin word for "nest"). The *Coronaviridae* family consists of two subfamilies: *Coronavirinae* and *Torovirinae*. The *Coronavirinae* are classified into four genera, *Alphacoronavirus*, *Betacoronavirus*, *Gammacoronavirus*, and *Deltacoronavirus*. The genus *Betacoronavirus* was previously further subdivided into lineages A, B, C, and D. Recently, these four lineages have been reclassified as subgenera of *Betacoronavirus*, and renamed *Embecovirus* (previous lineage A), *Sarbecovirus* (previous lineage B), *Merbecovirus* (previous lineage C) and *Nobecovirus* (previous lineage D). In addition, a fifth subgenus, *Hibecovirus*, was also included (Table 1).

## Virion Structure

The virion of coronaviruses is approximately 120 nm in diameter, with club-shaped protein spikes projecting from the surface, resembling a solar corona (Fig. 1). Coronaviruses have four canonical structural proteins: the large transmembrane spike protein (S; 1160–1400 amino acids), a small envelope protein (E; 74–109 amino acids, present in small amounts in viral envelope), an integral membrane glycoprotein (M; 250 amino acids, most abundant protein in viral envelope), and a heavily phosphorylated nucleocapsid protein (N; 500 amino acids, the only protein present in the nucleocapsid). In addition, viruses belonging to *Embecovirus* (previous lineage A of the *Betacoronavirus*) have an additional membrane-anchored hemagglutinin-esterase protein (HE; 430 amino acids). The HE protein is not essential for viral replication in vitro but may affect the production of infectious viral particles and viral tropism in vivo by reversible attachment to *O*-acetylated sialic acids.

**Table 1** Classification of coronavirus

Genera	Subgenera	Species
Alphacoronavirus	Colacovirus	Bat coronavirus CDPHE15
		Bat coronavirus HKU10
	Decacovirus	Rhinolophus ferrumequinum alphacoronavirus HuB-2013
		Human coronavirus 229E
	Duvinacovirus	Lucheng Rn rat coronavirus
	Luchacovirus	Ferret coronavirus
	Minacovirus	Mink coronavirus 1
	Minunacovirus	Miniopterus bat coronavirus 1
		Miniopterus bat coronavirus HKU8
	Myotacovirus	Myotis ricketti alphacoronavirus Sax-2011
	Nyctacovirus	Nyctalus velutinus alphacoronavirus SC-2013
	Pedacovirus	Porcine epidemic diarrhea virus
	Rhinacovirus	Scotophilus bat coronavirus 512
		Rhinolophus bat coronavirus HKU2
	Setracovirus	Human coronavirus NL63
NL63-related bat coronavirus strain BtKYNL63-9b		
Betacoronavirus	Tegacovirus	Alphacoronavirus 1 <sup>a</sup>
		Betacoronavirus 1
	Embecovirus (lineage A)	China Rattus coronavirus HKU24
		Human coronavirus HKU1
	Sarbecovirus (lineage B)	Murine coronavirus
		Severe acute respiratory syndrome-related coronavirus
	Merbecovirus (lineage C)	Hedgehog coronavirus 1
		Middle East respiratory syndrome-related coronavirus
	Nobecovirus (lineage D)	Pipistrellus bat coronavirus HKU5
		Tylonycteris bat coronavirus HKU4
	Hibecovirus	Rousettus bat coronavirus GCCDC1
		Rousettus bat coronavirus HKU9
	Cegacovirus	Bat Hp-betacoronavirus Zhejiang 2013
		Beluga whale coronavirus SW1
	Igcacovirus	Avian coronavirus
Wigeon coronavirus HKU20		
Andecovirus	Bulbul coronavirus HKU11	
	Buldecovirus	Coronavirus HKU15
Herdacovirus	Munia coronavirus HKU13	
	White-eye coronavirus HKU16	
Moordecovirus	Night heron coronavirus HKU19	
	Common moorhen coronavirus HKU21	

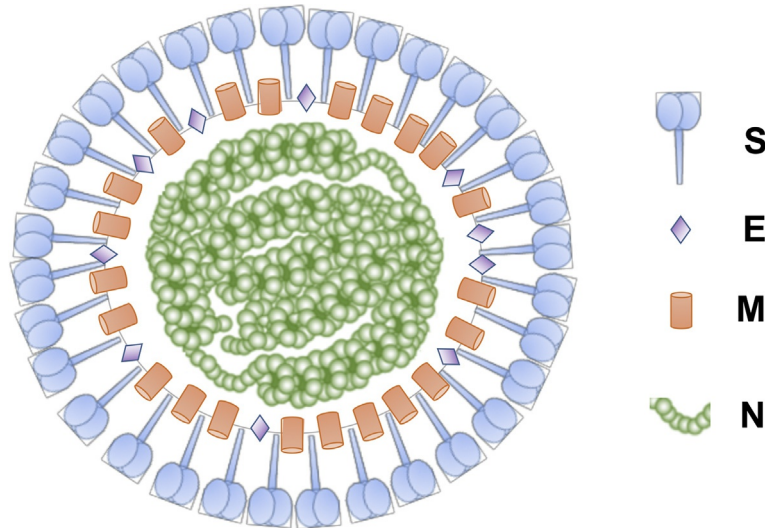
<sup>a</sup>Alphacoronavirus 1 contains viruses with the historical names: *Transmissible gastroenteritis virus of swine*, *Porcine transmissible gastroenteritis virus*, *Feline infectious peritonitis virus*, *Canine coronavirus*, and *Feline coronavirus*.

Based on International Committee on Taxonomy of Viruses <https://talk.ictvonline.org/taxonomy/>.

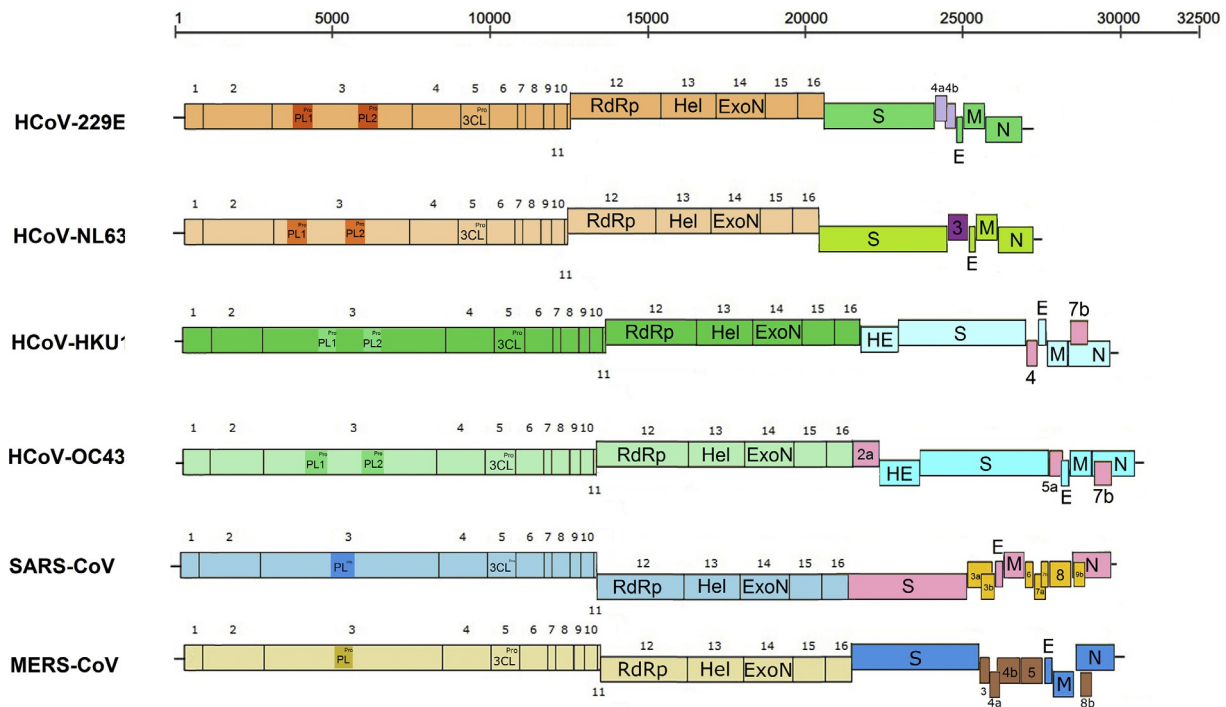
Trimers of S protein form 18–23 nm-long spikes on the surface of coronavirus which give its characteristic morphology. The heavily N-glycosylated S protein comprises two functionally distinct subunits—the N-terminal S1 and C-terminal S2 domains which are involved in receptor binding and membrane fusion, respectively. Like other class I viral fusion proteins, the S protein undergoes a series of events upon receptor recognition, including receptor engagement, proteolytic cleavage to shed the S1 subunit and conformational changes in S2 domain that lead to fusion of the viral and host membranes. The receptor-binding domain (RBD) within S1 is implicated in host receptor recognition, the site of which varies depending on the virus species. For example, the RBD of SARS-CoV locates at residues 318 to 510 of S1; while the RBD of HCoV-229E is at residues 417 and 547 of S1, more downstream along the peptide chain. Substitutions in the RBD confer adaptability to new or orthologous entry receptors, ultimately affecting tissue tropism. The S protein is also the major target of neutralizing antibodies. Variations in S1 determine the susceptibility to protective immune responses.

## Genome Organization

In general, coronaviruses have the largest known RNA genomes with a length of 28–32 kb (Fig. 2). The four HCoVs, HCoV-229E, HCoV-NL63, HCoV-OC43, and HCoV-HKU1, have genome sizes of around 27.3, 27.5, 30.5, and 29.9 kb respectively. SARS-CoV



**Fig. 1** Diagrammatic representation of coronavirus virion. S, spike protein; E, envelope protein; M, membrane protein; N, nucleocapsid protein encapsidating the RNA genome.



**Fig. 2** Schematic diagram of human coronavirus genomes. ORF1a/b, occupying the 5' two thirds of the genome, translates into two replicase polyproteins pp1a and pp1ab as a result of ribosomal frameshifting, which are further cleaved to form 16 non-structural proteins nsp 1–16. This is followed by the subgenomic RNA fragments encoding the structural and accessory proteins. The genome structure of coronaviruses follows a characteristic order as shown in this diagram. *ORF*, open reading frame; *PLpro*, papain-like protease; *3CLpro*, chymotrypsin-like protease; *RdRp*, RNA-dependent RNA polymerase; *Hel*, RNA helicase; *RFS*, ribosomal frameshift region; *S*, spike; *E*, envelope; *M*, membrane; *N*, nucleocapsid.

and MERS-CoV have genome sizes of around 29.7 and 30.1 kb respectively. Similar to all other coronaviruses, the genome structures of all HCoVs follow a characteristic order: the 5' two thirds of the genome comprises a leader sequence and open reading frame 1a/b (ORF1a/b) encoding two replicase polyproteins, while the 3' one third consists of a nested set of subgenomic RNAs encoding structural proteins in the order of (HE)-S-E-M-N, as well as several accessory proteins. Both the 5' and 3' ends of the coronavirus genome contain short untranslated regions (UTRs). Additionally, at the beginning of each structural and accessory protein gene is a common sequence called transcriptional regulatory sequence (TRS). The translational product of ORF1a/b is expressed as two

co-terminal polyproteins pp1a and pp1ab, which are further cleaved by self-encoded proteases to generate 15–16 non-structural proteins (nsp) including major enzymes such as papain-like protease(s) (PL<sup>Pro</sup>, corresponding to nsp3), chymotrypsin-like protease (3CL<sup>Pro</sup>, corresponding to nsp5), RNA-dependent RNA polymerase (RdRp, corresponding to nsp12), RNA helicase (Hel, corresponding to nsp13) and exoribonuclease (ExoN, corresponding to nsp14) which ensures RdRp fidelity. These proteins associate to form the replicase complex. In order to express the two polyproteins pp1a and pp1ab, a slippery sequence (5'-UUUAAAC-3') followed by an RNA pseudoknot is utilized to achieve ribosomal frameshifting, which creates two different sets of reading frames from a single starting sequence. During the translation of pp1a, the ribosome will need to unwind the pseudoknot to continue elongation till the stop codon of ORF1a. However, occasionally the ribosome is stalled at the slippery sequence due to the presence of the pseudoknot structure, slips one nucleotide backward, then melts the pseudoknot and continues with the translation for pp1ab, this time in a –1 frame (ORF1b) compared with ORF1a. Thus, pp1ab is essentially a C-terminal extension of pp1a as a result of this programmed –1 frameshift. The frequency of ribosomal frameshifting varies among different virus species, which can be as high as 20%–30% in vitro. The various accessory proteins encoded by the subgenomic mRNAs interspersed among the structural protein genes are mostly non-essential for viral replication in vitro, but purportedly serve modulatory functions in DNA synthesis, unfolded protein response (UPR), cellular apoptosis and interaction with innate immunity.

## Replication

The interaction between virion and host cells is initiated by the attachment of S protein to specific receptors, the availability of which determines host tissue susceptibility. The receptor profiles of several coronavirus species have been well established, for example angiotensin-converting enzyme 2 (ACE2) for SARS-CoV and HCoV-NL63, dipeptidyl peptidase-4 (DPP4) for MERS-CoV, and aminopeptidase-N (APN) for HCoV-229E and many other alphacoronaviruses. Following receptor binding and fusion of viral and cellular membranes, the coronavirus genomic RNA is released into the cytoplasm of host cells for translation of the replicase polyproteins pp1a and pp1ab and subsequent cleavage into 15–16 nsps as detailed above. In addition to forming the replicase complex, different nsps may serve additional non-replicative functions such as blockage of innate immune responses and cell cycle modulation. Following the translation and assembly of the viral replicase complex, negative-sense copies of both genomic and subgenomic RNAs are synthesized, which serve as templates for the synthesis of positive-sense genomic RNA and subgenomic mRNAs. The subgenomic mRNAs are generated by a process named discontinuous transcription, and provide templates for the production of the various structural and accessory proteins. The genomic RNA replicates are encapsidated by the N protein, the resulting nucleocapsid then incorporate into the membranes of endoplasmic reticulum-Golgi intermediate compartment (ERGIC) where other structural proteins are being processed. The M protein directs the interaction with both N protein and C-terminal part of the S protein. Interestingly, the E protein, despite being in small quantities, is essential for virus-like particle formation, which cannot be formed by M protein alone. It is postulated that the E protein induces membrane curvature and prevents M protein aggregation. The release of newly assembled virions usually starts 3–4 h after initial infection.

## Recombination and Evolution

The high frequency of homologous RNA recombination is one of the defining features of coronavirus, which is attributed to the strand switching ability of RdRp. This results in a highly plastic genome readily adaptable to new host species. The role recombination events play in viral evolution is most elaborated in SARS-CoV. Soon after the identification of SARS-CoV as the causative agent of the 2003 SARS pandemic, SARS-related CoVs (SARSr-CoVs) were found in palm civets from live animal markets in Guangdong. Hence civets were initially believed to be the animal reservoir of SARS-CoV. However, SARSr-CoVs were only detected in civets from the market, but not those in the wild or civet breeding farms, suggesting a later mixing event during transportation and trading. In addition, there was high ratio of nonsynonymous to synonymous mutation rates of the *S*, *orf3a* and *nsp3* genes in civet SARSr-CoVs collected in 2003–04, indicating that the viruses were undergoing rapid genetic adaptation in civets during the specific period of time, which is further corroborated by the findings that the S protein of civet SARSr-CoVs had less efficient usage of ACE2 receptor compared with human SARS-CoV during the 2003 outbreak and that civet SARSr-CoVs showed resistance to inhibitory antibodies. Subsequently, studies by different groups of scientists have proven that bats are likely to be the ultimate reservoir of SARS-CoV, as SARSr-CoVs were isolated in a large number of horseshoe bats which also demonstrated serologic evidence of past SARSr-CoV infection. Till November 2018, 313 SARSr-CoV genomes have been sequenced, including 274 from human, 18 from civets and 47 from bats [mostly from Chinese horseshoe bats (*Rhinolophus sinicus*),  $n = 30$ ; and greater horseshoe bats (*Rhinolophus ferrumequinum*),  $n = 9$ ]. It is evident from gene alignment and phylogenetic analyses that certain genes of a particular bat SARSr-CoV possess higher degree of nucleotide identity to the corresponding genes of SARS-CoV from human/civets than other genes in the same virus, resulting in shift of phylogenetic position when trees are constructed for alternative genes. In addition, genome fragments with higher and higher nucleotide identity to the SARS-CoV isolated in humans were continuously being found in bats, mostly the horseshoe bat species, including viral strains that can directly utilize human ACE2 for viral entry without further adaption. It is concluded from latest genome analyses that the SARS-CoV causing the 2003 outbreak was generated through a series of recombination events from a number of SARSr-CoV ancestors from horseshoe bats. Overall, it is observed that

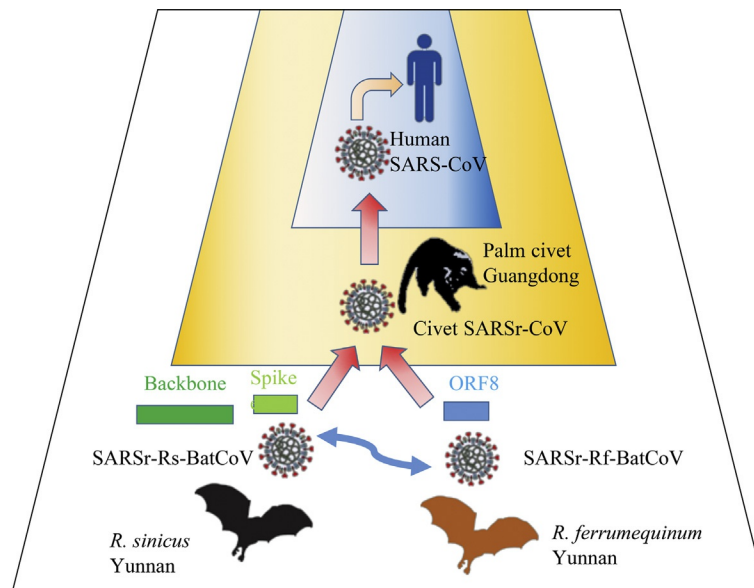
the Yunnan Province of China possesses the greatest biodiversity of bats and that the SARSr-CoVs from bats in Yunnan demonstrate highest nucleotide identity to human/civet SARS-CoVs along the entire length of genome (Fig. 3).

### Epidemiology and Clinical Features of Human Coronaviruses

SARS-CoV and MERS-CoV are two highly pathogenic coronaviruses that have claimed thousands of lives globally in the past two decades. However, prior to the SARS-CoV pandemic in 2003, human coronaviruses were only believed to cause symptoms of common cold—mild, self-limiting upper respiratory tract infection. Four human coronaviruses have been identified so far, including two alphacoronaviruses HCoV-229E and HCoV-NL63, and two betacoronaviruses HCoV-OC43 and HCoV-HKU1. HCoV-229E and HCoV-OC43 have been known for more than 50 years, while HCoV-NL63 and HCoV-HKU1 were isolated after the SARS-CoV outbreak. They are characterized by direct human-to-human spread and, in contrary to SARS-CoV and MERS-CoV, are not known to be transmitted from animal reservoir. Nevertheless, significant differences in genetic variability have been observed among the human coronaviruses. HCoV-OC43 isolates from the same location but in different years demonstrated significant sequence variability, while HCoV-229E isolates from around the world displayed minimal genetic divergence. The higher degree of genetic variability likely explains the fact that HCoV-OC43 is able to cross genetic barrier to infect mice neural tissue. These human coronaviruses are globally distributed, although the frequency of isolation of the four viruses varies at different times in different places of the world, typically comprising 10%–40% of respiratory samples testing positive for any respiratory viruses. Human coronaviruses generally demonstrate winter seasonality between the months of December and April, similar to the pattern observed in influenza virus. They are associated with a range of respiratory outcomes, most commonly upper respiratory tract infection of moderate clinical concern, but occasionally also lower respiratory tract involvement including bronchiolitis and pneumonia leading to fatality, especially in neonates, the elderly and individuals with compromised immunity. The mainstay of laboratory diagnosis is via quantitative real-time polymerase chain reaction (qRT-PCR), either in-house developed and validated or commercially available multiplex PCR platforms. The utilization of serologic assays is limited to epidemiologic studies and in cases where the viral RNA is difficult to isolate. There are no directed antiviral agents for human coronaviruses to date, so treatment is largely supportive.

### Emerging Human Coronaviruses

Soon after its emergence in the Guangdong Province of China in 2002/2003, SARS-CoV quickly spread to 27 countries, mainly concentrated around Southeast Asia, except Canada where 251 confirmed cases were reported. There have been a total of 8096 laboratory confirmed cases of SARS globally, leading to 774 mortalities (9.6%) in 11 countries (<https://www.who.int/csr/sars/>)



**Fig. 3** Overview of evolution of human SARS-CoV from bat SARSr-CoV. Overall speaking, bat SARSr-CoVs isolated from Chinese horseshoe bats (*Rhinolophus sinicus*) in the Yunnan Province of China demonstrated highest genome identity to human/civet SARS-CoV. The highest nucleotide identity was observed in the spike protein gene. Several strains have been shown to be able to utilize human AEC2 for viral entry and infect well-differentiated human airway epithelial cell lines in vivo. The ORF8 gene of human/civet SARS-CoV, on the other hand, was predominantly closer to the ORF8 gene of bat SARSr-CoV found in greater horseshoe bats (*Rhinolophus ferrumequinum*), suggesting a separate origin. Thus, a hypothesis was proposed that multiple recombination events joining genome fragments of different bat SARSr-CoV origins culminated in the evolution into human SARS-CoV.



country/en/). Due to the implementation of infection control measures and effective quarantine, the SARS pandemic stopped in 2003, and no human infections are reported ever since. Approximately 10 years later, MERS-CoV emerged as a highly pathogenic human respiratory virus in the Kingdom of Saudi Arabia. So far, all cases of MERS have been linked to residence in or traveling to the Arabian Peninsula, with subsequent nosocomial transmission. The largest outbreak outside the region occurred in the Republic of Korea in 2015, starting with an imported case from the Middle East. As of 27 November 2018, a total of 2266 laboratory-confirmed cases of MERS have been reported, including 804 fatalities (35.5%) (<https://www.who.int/emergencies/mers-cov/en/>). The combined effect of viral replication in the lower respiratory tract and excitation of hyperinflammatory immune response contributes to the high mortality, especially in elderly patients and those with predisposing comorbidities. Several treatment strategies including ribavirin, protease inhibitors, neutralizing antibodies and immune modulators have been proposed for SARS and MERS based on animal model and in vitro studies, but clinical efficacy data are lacking, and none have been substantiated by rigorous human trials.

As afore mentioned, SARS-CoV in human is thought to arise from bat SARSr-CoV, with palm civets being intermediate hosts. Direct transmission routes from bats to humans have also been postulated. MERS-CoV, on the other hand, is thought to be transmitted to humans from infected dromedary camels (*Camelus dromedarius*). Although the most common ancestor of MERS-CoV still remains unknown, phylogenetic analysis showed that MERS-CoV is closely related to two known bat coronaviruses under the subgenus *Merbecovirus*, namely Ty-BatCoV HKU4 found in *Tylonycteris pachypus* and Pi-BatCoV HKU5 found in *Pipistrellus abramus*. However, MERS-CoV shares lower nucleotide identity (65%–80%) with other members of *Merbecovirus*, in contrast to SARS-CoV where some of the genome sequences of bat SARSr-CoVs are more than 90% identical to those of human/civet SARS-CoVs. Indeed, bats may not be the only reservoir of MERS-CoV, as one study has found a close relative of MERS-CoV, Hedgehog coronavirus 1, in the European hedgehog (*Erinaceus europaeus*). Interspecies transmission from zoonotic sources can occur by viral spillover events during direct or indirect contact between humans and reservoir hosts. As a result of expanding human activity and climate change, previously geographically restricted viruses and their reservoir or intermediate hosts are continuously driven to new ecological niche with potential close human contact. Emerging zoonotic viruses can also expand their receptor repertoire by relentless recombination events, evolving from animal-restricted pathogen to one that can utilize human receptors for efficient viral entry and replication. Continuous surveillance on animal species including bats and other mammals residing near bat habitats is essential for early identification of potential zoonotic outbreak and prediction of possible interspecies jumping. Emphasis should be placed around wet markets, farms and abattoirs to safeguard humans from novel zoonotic pathogens.

## Acknowledgments

This work is partly supported by the Theme-based Research Scheme (Project No. T11/707/15), University Grant Committee; Consultancy Service for Enhancing Laboratory Surveillance of Emerging Infectious Diseases and Research Capability on Antimicrobial Resistance for Department of Health, HKSAR Government; and University Development Fund, The University of Hong Kong.

## Further Reading

- Cavanagh D and Britton P (2008) Coronaviruses: General features. In: Mahy BWJ and Van Regenmortel MHV (eds.) *Encyclopedia of virology*, 3rd edn. Oxford: Academic Press.
- Coleman CM and Frieman MB (2014) Coronaviruses: Important emerging human pathogens. *Journal of Virology* 88: 5209–5212.
- De Wit E, Van Doremalen N, Falzarano D, and Munster VJ (2016) Sars and Mers: Recent insights into emerging coronaviruses. *Nature Reviews. Microbiology* 14: 523–534.
- Fehr AR and Perlman S (2015) Coronaviruses: An overview of their replication and pathogenesis. *Methods in Molecular Biology* 1282: 1–23.
- Gaunt ER, Hardie A, Claas ECJ, Simmonds P, and 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. *Journal of Clinical Microbiology* 48: 2940.