

**Interspecies transmission and evolution of the emerging coronaviruses: perspectives from bat physiology and protein spatial structure**

Baicheng HUANG¹, Kegong TIAN (✉)¹,²

¹ National Research Center for Veterinary Medicine, Luoyang 471003, China
² College of Animal Science and Veterinary Medicine, Henan Agricultural University, Zhengzhou 450046, China

**Abstract**

Emergent coronaviruses (CoVs) such as SARS-CoV and MERS-CoV have posed great threats to public health worldwide over the past two decades. Currently, the emergence of SARS-CoV-2 as a pandemic causes greater public health concern. CoV diversity is due to the large size and replication mechanisms of the genomes together with having bats as their optimum natural hosts. The ecological behavior and unique immune characteristics of bats are optimal for the homologous recombination of CoVs. The relationship of spatial structural characteristics of the spike protein, a protein that is critical for recognition by host receptors, in different CoVs may provide evidence in explaining the coevolution of CoVs and their hosts. This information may help to enhance our understanding of CoV evolution and thus provide part of the basis of preparations for any future outbreaks.

**Keywords**

bat, coronavirus, evolution, host receptor, spike protein, transmission

---

**1 Introduction**

The human coronaviruses (CoVs) have been reported to represent only a minimal threat to public health[1] until the outbreak of severe acute respiratory syndrome (SARS) caused by SARS-CoV in 2002 and 2003[2]. The Middle East respiratory syndrome coronaviruses (MERS-CoV) then emerged in Middle Eastern countries in 2012[3] and cases continue to be reported. The recent COVID-19 outbreak, associated with the novel coronavirus SARS-CoV-2, in December 2019[4] shows a significantly larger scale of infection than previous CoV outbreaks and has therefore caused greater public health concern worldwide, and was designated a pandemic by WHO on March 11, 2020.

Bats are a major source of zoonotic viruses[5] and have been demonstrated to be reservoirs for several emergent viruses including SARS-related coronaviruses (SARS-CoVs[6], Ebola virus[7] and Marburg virus[8]. SARS-CoV and MERS-CoV are highly transmissible and most likely originated in bats[9]. SARS-CoV-2 has a high similarity of genome sequence to SARSr-CoVs[10], demonstrating a high potential to be of bat origin. SARS-CoV-2 uses cellular receptors for cell entry, whereas both SARS-CoV and SARS-CoV-2 use the angiotensin-converting enzyme 2 (ACE2)[6,11] and MERS-CoV uses dipептидил пептидаза 4 (DPP4, also known as CD26)[12].

In this review, we focus on the physiological characteristics of bats and the protein structural biology of the virus in humans. Specifically, we emphasize the adaptability of bats as the reservoirs of CoVs from the perspective of their immune characteristics and ecological behavior. We also emphasize the structural characteristics of these emerging viruses and their host receptors which may facilitate virus transmission. It is intended that this information will enhance our understanding of the evolutionary relationships between the new emerging CoVs and their hosts, and help in the development of countermeasures against any future outbreaks of novel CoVs.

---

**2 Epidemics and transmission of CoVs**

**2.1 CoV taxonomy and scope of infection**

CoVs are members of the subfamily Coronavirinae in the family Coronaviridae of the order Nidovirales (International Committee on Taxonomy of Viruses). All CoVs share similarities in the organization of the viral genome and can be separated into four genera based on...
phylogenetic clustering: Alphacoronavirus (α-CoV), Betacoronavirus (β-CoV), Gammacoronavirus (γ-CoV) and Deltacoronavirus (δ-CoV). CoVs are ecologically diverse. The hosts of α-CoV and β-CoV are thought to be restricted to mammals whereas γ-CoV and δ-CoV infect birds but some can infect mammals[13]. For example, pigs can be infected by porcine deltacoronavirus (PDCoV)[13,14].

A variety of diseases in farm and companion animals have been attributed to CoV infections such as infectious bronchitis virus (IBV) and canine respiratory CoV which cause respiratory symptoms in chickens[15] and dogs[16], respectively. Other CoV infections in animals result in gastrointestinal symptoms including transmissible gastroenteritis virus[17], porcine epidemic diarrhea virus[18], PDCoV[14], swine acute diarrhea syndrome CoV[19], bovine CoV[20,21], feline CoV[22] and canine CoV[23].

Four of the seven CoVs known to infect humans (NL63, 229E, OC43 and HKU1) induce only mild upper respiratory diseases and mild enteritis, with symptoms such as the common cold[24], whereas the remainder of the CoVs, including SARS-CoV, MERS-CoV and the recently emerged SARS-CoV-2, cause acute clinical symptoms characterized by severe pneumonia[25,26]. Human CoV 229E and NL63 belong to the genus α-CoV, and OC43, HKU1, SARS-CoV, MERS-CoV and SARS-CoV-2 are members of β-CoV.

2.2 Transmission of SARS-CoV

During the SARS epidemic the initial patients had a history of animal exposure before developing the disease. After SARS-CoV was isolated from two patients[27] the SARS-CoV antigen and antibody were found in palm civets (Paguma larvata) from screening in a live animal market[28], and then palm civet was considered to be the intermediate host of SARS-CoV. In 2005, there were reports that the novel human SARSr-CoVs were found in bats (Rhinolophus)[29,30] suggesting that bats are the natural hosts of SARS-CoV. In 2013, a novel bat CoV from Chinese horseshoe bats (F. Rhinolophidae) closely related to SARS-CoV was found in Yunnan, China, and the use of ACE2 for cell entry by the novel bat CoV was identified, providing the strongest evidence that Chinese horseshoe bats are natural reservoirs of SARS-CoV[6]. Subsequent serological methods found that a range of animal species had virus-specific antibodies[31] in addition to humans who live in close proximity to caves where bats carrying diverse SARSr-CoVs roost[32].

2.3 Transmission of MERS-CoV

After the first outbreak of MERS in 2012 the MERS-CoV was isolated from the sputum of a 60-year-old man who presented with acute pneumonia in Saudi Arabia[3]. Soon after, a virus detected in a species of bat (Taphozous perforatus) had full nucleotide identity to MERS-CoV from the human patient[33] suggesting that bats are the natural host of MERS-CoV. In 2014, two isolates of MERS-CoV from a dromedary and from a patient who died of MERS-CoV infection had identical genome sequences[34], indicating that the transmission of MERS-CoV infection was through close contact with the infected camel and that camels can be the intermediate host of MERS-CoV. In serological studies high positive rates of MERS-CoV neutralizing antibodies were detected in camel serum samples from east Africa (even in serum samples collected in 1983)[35] and Pakistan[36], suggesting persistent and widespread virus circulation in these animals.

2.4 Potential hosts and transmission of SARS-CoV-2

A previously unknown β-CoV (now known as SARS-CoV-2) was discovered from patients with pneumonia[4]. The genome sequence of SARS-CoV-2 has 79.6% identity to a SARS-CoV and 96.0% identity to a bat CoV, providing strong evidence that the virus is of bat origin and that bats are a natural host[10]. Currently, the available evidence indicates that pangolins (Manis javanica) may serve as an intermediate host for SARS-CoV-2 due to the high sequence identity between SARS-CoV-2 and a CoV isolated from pangolins[37,38], as well as use of the same cell entry receptor (ACE2) but the confirmation of an intermediate hosts for SARS-CoV-2 is not yet possible due to sequence diversity on the spike (S) protein[39].

3 Immunological and ecological characteristics of bats

3.1 Evolution of bats

Bats may be the most abundant, diverse and geographically dispersed vertebrate animals which evolved early and have changed relatively little[40]. Taxonomically, bats are members of the order Chiroptera (including the suborders Yinpterichiroptera and Yangochiroptera), having a sister relationship with the order Fereuungulata (including carnivorous animals such as cats, civets, dogs, pangolins and weasels) within the clade Scrotifera, with all of them in the superorder Laurasiatheria[41]. The civet and camel are members of the order Fereuungulata, which are considered to be intermediate hosts of SARS-CoV and MERS-CoV, respectively[28,34]. Through prediction of species evolution for the understanding of biogeographical history, Africa has been suggested to be the center of origin of modern-day bat families[42]. The corresponding ancient origins of zoonotic viruses such as rabies virus[43] maintained in bats indicate a long history of coevolution. For replication, viruses in bats may
have used cellular receptors and biochemical pathways that are conserved in mammals, and this may enhance the capacity for transmission of bat-associated viruses to other mammals\(^5\). Bats display unique immune characteristics and ecological behavior, and are widely considered to be the natural host of a range of zoonotic viruses such as rabies virus\(^{63,64}\), hantaviruses\(^{65}\), Nipah virus\(^{66}\), Hendra virus\(^{67}\), Ebola virus\(^{68}\), and the emergent CoVs such as SARS-CoV\(^2\) and MERS-CoV\(^3\).

3.2 Ecological behavior of bats facilitates virus transmission

The unique flight capability of bats is used for daily acquisition of food and seasonal long-distance migrations\(^5\) and this may promote virus transmission and recombination. Rabies virus variants have been identified in migrating subpopulations\(^49\) or might exist between the migrating and nonmigrating subpopulations\(^50\). In a study of CoV transmission related to bat migration, frequent recombination events were identified between different strains of the SARSr-CoVs in *Rhinolophus*, Chinese horseshoe bats, (SARSr-Rh-BatCoV) during migration\(^51\). The host shift in SARSr-CoVs was mostly related to geographical structure (CoVs from adjacent provinces were clustered) rather than to bat species\(^52\), indicating the impact of migration in SARSr-CoVs transmission by bats.

Bats are the only land mammals that use echolocation for navigation. Echolocation signals powered by the muscles of bats can generate droplets of oropharyngeal fluids that may promote the airborne transmission of viruses such as rabies virus\(^53\).

The daily torpor and seasonal hibernation of certain bat species with reduced metabolic activity and body temperature at night and during winter lead to significant energy savings\(^54\). Hibernating bats can serve as a viral reservoir and this is conducive to high-level recombination events in the rabies virus\(^55\). The CoVs identified in hibernating bats (*Myotis lucifugus*) appeared to be from two distinct clades and the virus infection occurred before hibernation\(^56\), suggesting that two closely related CoVs may circulate in bats during hibernation.

The long lifespan might be related to the inhibition of telomere shortening in the bat genus with greatest longevity\(^57\). Also, the possible antiviral mechanisms of bats may help maintain the viruses and promote the transmission to different animals\(^5\). Persistent viral infections in long-lived bats, together with the large population size and gregarious roosting behavior which may have a substantial impact on the basic reproductive number of infections, greatly increases the potential for intra- and interspecies transmission of viruses\(^5\).

3.3 Unique immune characteristics of bats

The unique immune characteristics of bats contribute to the dynamic equilibrium of virus coexistence. Vertebrates respond to viral infections by inducing interferons (IFNs) that trigger antiviral defense by inducing the interferon-stimulated gene (ISG). Type I IFNs, composed of multiple α subtypes and a single β subtype, are an important part of the immune response to viral infection\(^58\). Two signal transduction pathways are used for Type I IFN expression: (1) a toll-like receptor (TLR)-dependent pathway that depends on the reorganization of viral double-stranded RNA, single-stranded RNA and CpG DNA via TLR in cells for subsequent IFN-β expression, and (2) a TLR-independent pathway which depends on intracellular sensors such as retinoic acid-inducible gene I and melanoma differentiation-associated gene-5 for detection of viral components in the cytoplasm, and then transactivation of IFN-β mRNA\(^59\). Interferons and viruses maintain an equilibrium in Nature that allows regulation of viral replication\(^58\). Bat IFN-β shares the highest sequence identity with pig IFN-β at both the nucleotide and amino acid levels\(^60\). Also, stimulation of bat primary kidney cells with exogenous bat type I IFNs resulted in increased type I IFN mRNA expression\(^60\), thus, the high level expression of IFNs can facilitate the equilibrium between bats and their viruses.

The potential repressor (c-Rel) binding motif in bat TNFα promoter can inhibit the expression of the inflammatory gene TNFα\(^61\). In addition, the secretion of IL-1β is downregulated and the activation of NLRP3 inflammatory bodies is significantly dampened during viral infection\(^62\). Myxovirus resistance gene (*Mx*) is an important antiviral ISG, encoding two homologous proteins, Mx1 and Mx2. Bat *Mx1* gene is close to its human ortholog *MxA* phylogenetically, and the encoded protein Mx1 can significantly reduce the polymerase activity of viruses circulating in bats, including Ebola and influenza A-like viruses\(^63\). Compared to human ISGs in IFN-stimulated cells, ISG expression in some species of bats has the same early induction kinetics but distinct late-phase decline\(^64\). Notably, in unstimulated cells, bat ISGs were expressed more highly than their human equivalents\(^64\). This stand-by mechanism of interferon expression in bats can ensure that the high level of antiviral inflammatory response can be quickly induced in the early stages of infection for rapid control of virus replication, and subsequently the unique rapid decline results in reduced cytotoxicity in bats (this stage was not observed during human ISG response to infection), which can prevent the severe pathological consequences in humans, such as severe pneumonia, due to the prolonged inflammatory response after viral infection.

Flight is energetically a very costly form of locomotion and results in high levels of oxygen free radicals caused by
increased metabolic activity\cite{65}, making animals more prone to DNA damage\cite{66}. Excessive energy expenditure due to an increase in immune response responsible for DNA repair is energetically costly\cite{67}. Bats may therefore have evolved special mechanisms to suppress the activation of the immune response caused by DNA damage, thereby reducing the inflammatory response and against a free-radical effect on aging in certain species with small size and high metabolic rates\cite{68}. In bats the evolutionary suppression of inflammation and consequent susceptibility to virus infection is counteracted by constitutive expression of innate immune genes or novel genes to target viruses\cite{69}, and this creates opportunities for stable and dynamic circulation of the virus in bat populations.

These immune responses reveal a tight control of viral replication by the innate immune defense in bats, possibly contributing to the stable coexistence between bats and their viruses.

---

**4 CoV spike protein and its host receptors in virus adaptation**

**4.1 Cell entry and host-switch by spike protein**

The S protein of CoVs is highly glycosylated and can be divided into two functional domains: S1 that is responsible for receptor binding and S2 that responsible for cell fusion\cite{70}. CoV S protein is a typical class I virus fusion protein and is characterized by the formation of an α-helical coiled-coil structure\cite{71}. Cell entry by CoVs mainly relies on S protein which mediates the functions of receptor binding and subsequent fusion of the viral and cellular membranes, thereby releasing the viral nucleocapsid into the cytoplasm\cite{72}. CoVs use a variety of receptors to activate fusion, and S protein alterations are sufficient to extend or alter the host range such as the ability of cross-infection between murine CoV and feline infectious peritonitis virus (FIPV)\cite{73,74}, the interspecies S protein switching of IBV\cite{75}, and the possible tropism switch of feline enteric CoV and FIPV\cite{76,77}.

**4.2 SARS-CoV tropism by spike protein alteration**

ACE2, a type I integral membrane protein abundantly expressed in lung tissue, is a functional receptor for SARS-CoV\cite{78}. The S protein of SARS-CoV from the outbreak in 2002–2003 showed a higher binding affinity to human ACE2 than that of the S protein driven from the strains of humans and palm civets isolated in 2003–2004. Notably, the lower affinity can be complemented by specific residue mutation in human ACE2 or viral S proteins\cite{79}, indicating that SARS-CoV has structural selectivity with cellular receptors that may be caused by the accumulation of mutations in the S protein during evolution. This type of mechanism of S protein structural alteration (residue changes necessary for adaptation to the receptor)\cite{80,81} may explain the success or failure of SARS-CoV transmission\cite{79,82}.

**4.3 MERS-CoV tropism by spike protein alteration**

MERS-CoV uses DPP4 as the cellular receptor for entry, and the cellular restriction of MERS-CoV is determined by DPP4 expression rather than by downstream restriction factors\cite{12}. In a cell line expressing a species-derived variant of DPP4, a suboptimal variant of MERS-CoV adapted the cell by the accumulation of mutations in the S protein, resulting in the enhancement of cell entry by altering the surface charge distribution of the S protein\cite{83}. In addition, bat DPP4 proteins from seven species were all interacting with the receptor binding domain (RBD) of MERS-CoV\cite{69}. Substitution of key residues and their adjacent amino acids of bat DPP4 leads to the decreased binding affinity to the MERS-RBD\cite{84}, which may be caused by the shift in H-bond pairs (induced by the mutation-caused hydrophobicity), the limitation of the conformational flexibility\cite{84} and the diverse glycosylation patterns of the S protein\cite{85}. The diversity of bat DPP4 may confer the evolutionary stress for the cell adaptation of MERS-CoVs, and may have led to the generation of diversified strains during coevolution\cite{84}.

**4.4 High spatial similarity of S1-RBD between SARS-CoV-2 and SARS-CoV**

Until March 8, 2020, a total of the 447 items (438 viral proteins and 9 non-viral compounds such as peptides) could be found by searching for CoV in Research Collaboratory for Structural Bioinformatics Protein Data Bank (RCSB PDB, rcsb.org). Crystal structure data (PDB entry) are mainly focused on the SARS-CoV and MERS-CoV that caused the previous epidemics, and the SARS-CoV-2 in terms of the fast release in RCSB (Fig. 1). The S protein and the main proteases (Mpro and 3CLpro) are dominant (Table 1) in the available PDB data.

In computational analysis of protein spatial structure the root-mean-square deviation (RMSD) is a standard tool for comparing the similarity of two molecular structures\cite{86}, the lower values indicating higher similarity, and vice versa. According to the comparison based on the available spatial structural data for CoV S proteins, it was found that the similarity of RBD of S protein between SARS-CoV-2 and SARS-CoV (RMSD of 1.021 Å) is higher than that between SARS-CoV-2 and MERS-CoV (RMSD of 2.351 Å) (Fig. 2), and this is consistent with the sequence comparison result for the S protein from these CoVs\cite{10}, indicating that the evolutionary pattern of SARS-CoV-2 is highly SARS-CoV related. The high similarity of S proteins between SARS-CoV-2 and SARS-CoV, together with the accumulation of mutations in S protein, may promote the host adaption of these CoVs.
The available PDB data on CoVs in RCSB. The other CoVs include 229E, NL63, OC43, PEDV, PRCV, TGEV, PDCoV, IBV, FIPV, MHV, HKU1, HKU4, HKU5, HKU9, Bov-CoV, and rat CoV; the other NSP include NSP1, NSP3, NSP4, NSP7, NSP8, NSP9, NSP10, NSP12, NSP13, NSP14, NSP15 and NSP16; the other proteins include viral proteins E (envelope protein), HE, M (membrane protein), N (nucleocapsid protein), ORF7A, ORF9b, together with the antibodies of CoVs.

Comparison of the spatial structure of the receptor binding domains (RBD) of CoV S protein. The PDB data are listed as: SARS-CoV-2 S (PDB: 6VSB), SARS-CoV S (PDB: 5XLR) and MERS-CoV S (PDB: 5X59).
5 Summary

Over the past two decades the emergence of CoVs such as SARS-CoV, MERS-CoV and the new SARS-CoV-2 have caused great threats to public health. CoV diversity is fundamentally due to the large size and replication mechanisms of the genome and bats being optimal natural hosts. The ecological behavior and unique immune characteristics of bats promote the homologous recombination of CoVs, which may increase the accumulation of infectious CoVs in humans. The highly variable S protein is the major determinant of CoV tropism. From an evolutionary perspective the infectious CoVs show convergence with hosts at the molecular level under the stress of natural selection, and this highlights the need for virus surveillance and the development of antiviral strategies focusing mainly on S proteins in preparations for any future outbreaks.

Acknowledgements This work was funded by Luoyang Heluo Talent Plan.

Compliance with ethics guidelines Baicheng Huang and Kegong Tian declare that they have no conflicts of interest or financial conflicts to disclose.

This article is a review and does not contain any studies with human or animal subjects performed by either of the authors.

References


73. Kuo L, Godeke G J, Raamsman M J B, Masters P S, Rottier P J M. Retargeting of coronavirus by substitution of the spike glycoprotein ectodomain: crossing the host cell species barrier. *Journal of


